Theory of functions of a real variable.

Shlomo Sternberg

November 12, 2004

Contents

Chapter 1

The topology of metric spaces

1.1   Metric spaces

A **metric** for a set X is a function d from $X \times X$ to the non-negative real numbers (which we dente by $\mathbb{R}_{\geq 0}$),

$$d : X \times X \to \mathbb{R}_{\geq 0}$$

such that for all $x, y, z \in X$

1. $d(x, y) = d(y, x)$

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

3. $d(x, x) = 0$

4. If $d(x, y) = 0$ then $x = y$.

The inequality in 2) is known as the **triangle inequality** since if X is the plane and d the usual notion of distance, it says that the length of an edge of a triangle is at most the sum of the lengths of the two other edges. (In the plane, the inequality is strict unless the three points lie on a line.)

Condition 4) is in many ways inessential, and it is often convenient to drop it, especially for the purposes of some proofs. For example, we might want to consider the decimal expansions $.49999\ldots$ and $.50000\ldots$ as different, but as having zero distance from one another. Or we might want to "identify" these two decimal expansions as representing the same point.

A function d which satisfies only conditions 1) - 3) is called a **pseudo-metric**.

A **metric space** is a pair (X, d) where X is a set and d is a metric on X. Almost always, when d is understood, we engage in the abuse of language and speak of "the metric space X".

Similarly for the notion of a **pseudo-metric space**.

In like fashion, we call $d(x, y)$ the **distance** between x and y, the function d being understood.

If r is a positive number and $x \in X$, the (open) **ball of radius** r about x is defined to be the set of points at distance less than r from x and is denoted by $B_r(x)$. In symbols,

$$B_r(x) := \{y| \ d(x, y) < r\}.$$

If r and s are positive real numbers and if x and z are points of a pseudo-metric space X, it is possible that $B_r(x) \cap B_s(z) = \emptyset$. This will certainly be the case if $d(x, z) > r + s$ by virtue of the triangle inequality. Suppose that this intersection is not empty and that

$$w \in B_r(x) \cap B_s(z).$$

If $y \in X$ is such that $d(y, w) < \min[r - d(x, w), s - d(z, w)]$ then the triangle inequality implies that $y \in B_r(x) \cap B_s(z)$. Put another way, if we set $t := \min[r - d(x, w), s - d(z, w)]$ then

$$B_t(w) \subset B_r(x) \cap B_s(z).$$

Put still another way, this says that the intersection of two (open) balls is either empty or is a union of open balls. So if we call a set in X **open** if either it is empty, or is a union of open balls, we conclude that the intersection of any finite number of open sets is open, as is the union of any number of open sets. In technical language, we say that the open balls form a **base** for a **topology** on X.

A map $f : X \to Y$ from one pseudo-metric space to another is called **continuous** if the inverse image under f of any open set in Y is an open set in X. Since an open set is a union of balls, this amounts to the condition that the inverse image of an open ball in Y is a union of open balls in X, or, to use the familiar ϵ, δ language, that if $f(x) = y$ then for every $\epsilon > 0$ there exists a $\delta = \delta(x, \epsilon) > 0$ such that

$$f(B_\delta(x)) \subset B_\epsilon(y).$$

Notice that in this definition δ is allowed to depend both on x and on ϵ. The map is called **uniformly continuous** if we can choose the δ independently of x.

An even stronger condition on a map from one pseudo-metric space to another is the **Lipschitz condition**. A map $f : X \to Y$ from a pseudo-metric space (X, d_X) to a pseudo-metric space (Y, d_Y) is called a **Lipschitz map** with **Lipschitz constant** C if

$$d_Y(f(x_1), f(x_2)) \le Cd_X(x_1, x_2) \quad \forall x_1, x_2 \in X.$$

Clearly a Lipschitz map is uniformly continuous.

For example, suppose that A is a fixed subset of a pseudo-metric space X. Define the function $d(A, \cdot)$ from X to $\mathbb{R}$ by

$$d(A, x) := \inf\{d(x, w), \ w \in A\}.$$

The triangle inequality says that

$$d(x, w) \le d(x, y) + d(y, w)$$

for all w, in particular for $w \in A$, and hence taking lower bounds we conclude that

$$d(A, x) \le d(x, y) + d(A, y).$$

or

$$d(A, x) - d(A, y) \le d(x, y).$$

Reversing the roles of x and y then gives

$$|d(A, x) - d(A, y)| \le d(x, y).$$

Using the standard metric on the real numbers where the distance between a and b is $|a - b|$ this last inequality says that $d(A, \cdot)$ is a Lipschitz map from X to $\mathbb{R}$ with $C = 1$.

A **closed** set is defined to be a set whose complement is open. Since the inverse image of the complement of a set (under a map f) is the complement of the inverse image, we conclude that the inverse image of a closed set under a continuous map is again closed.

For example, the set consisting of a single point in $\mathbb{R}$ is closed. Since the map $d(A, \cdot)$ is continuous, we conclude that the set

$$\{x | d(A, x) = 0\}$$

consisting of all points at zero distance from A is a closed set. It clearly is a closed set which contains A. Suppose that S is some closed set containing A, and $y \notin S$. Then there is some $r > 0$ such that $B_r(y)$ is contained in the complement of S, which implies that $d(y, w) \ge r$ for all $w \in S$. Thus $\{x | d(A, x) = 0\} \subset S$. In short $\{x | d(A, x) = 0\}$ is a closed set containing A which is contained in all closed sets containing A. This is the definition of the **closure** of a set, which is denoted by $\overline{A}$. We have proved that

$$\overline{A} = \{x | d(A, x) = 0\}.$$

In particular, the closure of the one point set $\{x\}$ consists of all points u such that $d(u, x) = 0$.

Now the relation $d(x, y) = 0$ is an equivalence relation, call it R. (Transitivity being a consequence of the triangle inequality.) This then divides the space X into equivalence classes, where each equivalence class is of the form $\overline{\{x\}}$, the closure of a one point set. If $u \in \overline{\{x\}}$ and $v \in \overline{\{y\}}$ then

$$d(u, v) \le d(u, x) + d(x, y) + d(y, v) = d(x, y).$$

since $x \in \overline{\{u\}}$ and $y \in \overline{\{v\}}$ we obtain the reverse inequality, and so

$$d(u, v) = d(x, y).$$

In other words, we may define the distance function on the quotient space X/R, i.e. on the space of equivalence classes by

$$d(\overline{\{x\}}, \overline{\{y\}}) := d(u, v), \quad u \in \overline{\{x\}}, v \in \overline{\{y\}}$$

and this does not depend on the choice of u and v. Axioms 1)-3) for a metric space continue to hold, but now

$$d(\overline{\{x\}}, \overline{\{y\}}) = 0 \Rightarrow \overline{\{x\}} = \overline{\{y\}}.$$

In other words, X/R is a *metric* space. Clearly the projection map $x \mapsto \overline{\{x\}}$ is an isometry of X onto X/R. (An **isometry** is a map which preserves distances.) In particular it is continuous. It is also open.

In short, we have provided a canonical way of passing (via an isometry) from a pseudo-metric space to a metric space by identifying points which are at zero distance from one another.

A subset A of a pseudo-metric space X is called *dense* if its closure is the whole space. From the above construction, the image A/R of A in the quotient space X/R is again dense. We will use this fact in the next section in the following form:

If $f : Y \to X$ *is an isometry of* Y *such that* $f(Y)$ *is a dense set of* X, *then* f *descends to a map* F *of* Y *onto a dense set in the metric space* X/R.

1.2   Completeness and completion.

The usual notion of convergence and Cauchy sequence go over unchanged to metric spaces or pseudo-metric spaces Y. A sequence $\{y_n\}$ is said to **converge** to the point y if for every $\epsilon > 0$ there exists an $N = N(\epsilon)$ such that

$$d(y_n, y) < \epsilon \quad \forall\, n > N.$$

A sequence $\{y_n\}$ is said to be **Cauchy** if for any $\epsilon > 0$ there exists an $N = N(\epsilon)$ such that

$$d(y_n, y_m) < \epsilon \quad \forall\, m, n > N.$$

The triangle inequality implies that every convergent sequence is Cauchy. But not every Cauchy sequence is convergent. For example, we can have a sequence of rational numbers which converge to an irrational number, as in the approximation to the square root of 2. So if we look at the set of rational numbers as a metric space $\mathbb{Q}$ in its own right, not every Cauchy sequence of rational numbers converges in $\mathbb{Q}$. We must "complete" the rational numbers to obtain $\mathbb{R}$, the set of real numbers. We want to discuss this phenomenon in general.

So we say that a (pseudo-)metric space is **complete** if every Cauchy sequence converges. The key result of this section is that we can always "complete" a metric or pseudo-metric space. More precisely, we claim that

Any metric (or pseudo-metric) space can be mapped by a one to one isometry onto a dense subset of a complete metric (or pseudo-metric) space.

By the italicized statement of the preceding section, it is enough to prove this for a pseudo-metric spaces X. Let X_{seq} denote the set of Cauchy sequences in X, and define the distance between the Cauchy sequences $\{x_n\}$ and $\{y_n\}$ to be

$$d(\{x_n\}, \{y_n\}) := \lim_{n \to \infty} d(x_n, y_n).$$

It is easy to check that d defines a pseudo-metric on X_{seq}. Let $f : X \to X_{seq}$ be the map sending x to the sequence all of whose elements are x;

$$f(x) = (x, x, x, x, \cdots).$$

It is clear that f is one to one and is an isometry. The image is dense since by definition

$$\lim d(f(x_n), \{x_n\}) = 0.$$

Now since $f(X)$ is dense in X_{seq}, it suffices to show that any Cauchy sequence of points of the form $f(x_n)$ converges to a limit. But such a sequence converges to the element $\{x_n\}$. QED

1.3 Normed vector spaces and Banach spaces.

Of special interest are vector spaces which have a metric which is compatible with the vector space properties and which is complete: Let V be a vector space over the real or complex numbers. A **norm** is a real valued function

$$v \mapsto \|v\|$$

on V which satisfies

1. $\|v\| \geq 0$ and > 0 if $v \neq 0$,

2. $\|cv\| = |c|\|v\|$ for any real (or complex) number c, and

3. $\|v + w\| \leq \|v\| + \|w\| \ \forall \ v, w \in V$.

Then $d(v, w) := \|v - w\|$ is a metric on V, which satisfies $d(v+u, w+u) = d(v, w)$ for all $v, w, u \in V$. The ball of radius r about the origin is then the set of all v such that $\|v\| < r$. A vector space equipped with a norm is called a **normed vector space** and if it is complete relative to the metric it is called a **Banach space**.

Our construction shows that any vector space with a norm can be completed so that it becomes a Banach space.

1.4   Compactness.

A topological space X is said to be **compact** if it has one (and hence the other) of the following equivalent properties:

- Every open cover has a finite subcover. In more detail: if $\{U_\alpha\}$ is a collection of open sets with

$$X \subset \bigcup_\alpha U_\alpha$$

  then there are finitely many $\alpha_1, \ldots, \alpha_n$ such that

$$X \subset U_{\alpha_1} \cup \cdots \cup U_{\alpha_n}.$$

- If $\mathcal{F}$ is a family of closed sets such that

$$\bigcap_{F \in \mathcal{F}} = \emptyset$$

  then a finite intersection of the F's are empty:

$$F_1 \cap \cdots \cap F_n = \emptyset.$$

1.5   Total Boundedness.

A metric space X is said to be **totally bounded** if for every $\epsilon > 0$ there are finitely many open balls of radius ϵ which cover X.

Theorem 1.5.1 *The following assertions are equivalent for a metric space:*

1. *X is compact.*

2. *Every sequence in X has a convergent subsequence.*

3. *X is totally bounded and complete.*

Proof that 1. $\Rightarrow$ 2. Let $\{y_i\}$ be a sequence of points in X. We first show that there is a point x with the property for every $\epsilon > 0$, the open ball of radius ϵ centered at x contains the points y_i for infinitely many i. Suppose not. Then for any $z \in X$ there is an $\epsilon > 0$ such that the ball $B_\epsilon(z)$ contains only finitely many y_i. Since $z \in B_\epsilon(z)$, the set of such balls covers X. By compactness, finitely many of these balls cover X, and hence there are only finitely many i, a contradiction.

Now choose i_1 so that y_{i_1} is in the ball of radius $\frac{1}{2}$ centered at x. Then choose $i_2 > i_1$ so that y_{i_2} is in the ball of radius $\frac{1}{4}$ centered at x and keep going. We have constructed a subsequence so that the points y_{i_k} converge to x. Thus we have proved that 1. implies 2.

Proof that 2. ⇒ 3. If $\{x_j\}$ is a Cauchy sequence in X, it has a convergent subsequence by hypothesis, and the limit of this subsequence is (by the triangle inequality) the limit of the original sequence. Hence X is complete. We must show that it is totally bounded. Given $\epsilon > 0$, pick a point $y_1 \in X$ and let $B_\epsilon(y_1)$ be open ball of radius ϵ about y_1. If $B_\epsilon(y_1) = X$ there is nothing further to prove. If not, pick a point $y_2 \in X - B_\epsilon(y_1)$ and let $B_\epsilon(y_2)$ be the ball of radius ϵ about y_2. If $B_\epsilon(y_1) \cup B_\epsilon(y_2) = X$ there is nothing to prove. If not, pick a point $y_3 \in X - (B_\epsilon(y_1) \cup B_\epsilon(y_2))$ etc. This procedure can not continue indefinitely, for then we will have constructed a sequence of points which are all at a mutual distance $\geq \epsilon$ from one another, and this sequence has no Cauchy subsequence.

Proof that 3. ⇒ 2. Let $\{x_j\}$ be a sequence of points in X which we relabel as $\{x_{1,j}\}$. Let $B_{1,\frac{1}{2}}, \ldots, B_{n_1,\frac{1}{2}}$ be a finite number of balls of radius $\frac{1}{2}$ which cover X. Our hypothesis 3. asserts that such a finite cover exists. Infinitely many of the j must be such that the $x_{1,j}$ all lie in one of these balls. Relabel this subsequence as $\{x_{2,j}\}$. Cover X by finitely many balls of radius $\frac{1}{3}$. There must be infinitely many j such that all the $x_{2,j}$ lie in one of the balls. Relabel this subsequence as $\{x_{3,j}\}$. Continue. At the ith stage we have a subsequence $\{x_{i,j}\}$ of our original sequence (in fact of the preceding subsequence in the construction) all of whose points lie in a ball of radius $1/i$. Now consider the "diagonal" subsequence

$$x_{1,1}, \ x_{2,2}, x_{3,3}, \ldots.$$

All the points from $x_{i,i}$ on lie in a fixed ball of radius $1/i$ so this is a Cauchy sequence. Since X is assumed to be complete, this subsequence of our original sequence is convergent.

We have shown that 2. and 3. are equivalent. The hard part of the proof consists in showing that these two conditions imply 1. For this it is useful to introduce some terminology:

1.6 Separability.

A metric space X is called **separable** if it has a countable subset $\{x_j\}$ of points which are dense. For example **R** is separable because the rationals are countable and dense. Similarly, $\mathbb{R}^n$ is separable because the points all of whose coordinates are rational form a countable dense subset.

Proposition 1.6.1 *Any subset Y of a separable metric space X is separable (in the induced metric).*

Proof. Let $\{x_j\}$ be a countable dense sequence in X. Consider the set of pairs (j, n) such that

$$B_{1/2n}(x_j) \cap Y \neq \emptyset.$$

For each such (j, n) let $y_{j,n}$ be any point in this non-empty intersection. We claim that the countable set of points $y_{j,n}$ are dense in Y. Indeed, let y be any point of Y. Let n be any positive integer. We can find a point x_j such that $d(x_j, y) < 1/2n$ since the x_j are dense in X. But then $d(y, y_{j,n}) < 1/n$ by the triangle inequality. QED

Proposition 1.6.2 *Any totally bounded metric space X is separable.*

Proof. For each n let $\{x_{1,n}, \ldots, x_{i_n,n}\}$ be the centers of balls of radius $1/n$ (finite in number) which cover X. Put all of these together into one sequence which is clearly dense. QED

A **base** for the open sets in a topology on a space X is a collection $\mathcal{B}$ of open set such that every open set of X is the union of sets of $\mathcal{B}$

Proposition 1.6.3 *A family $\mathcal{B}$ is a base for the topology on X if and only if for every $x \in X$ and every open set U containing x there is a $V \in \mathcal{B}$ such that $x \in V$ and $V \subset U$.*

Proof. If $\mathcal{B}$ is a base, then U is a union of members of $\mathcal{B}$ one of which must therefore contain x. Conversely, let U be an open subset of X. For each $x \in U$ there is a $V_x \subset U$ belonging to $\mathcal{B}$. The union of these over all $x \in U$ is contained in U and contains all the points of U, hence equals U. So $\mathcal{B}$ is a base. QED

1.7   Second Countability.

A topological space X is said to be **second countable** or to satisfy the **second axiom of countability** if it has a base $\mathcal{B}$ which is (finite or) countable.

Proposition 1.7.1 *A metric space X is second countable if and only if it is separable.*

Proof. Suppose X is separable with a countable dense set $\{x_i\}$. The open balls of radius $1/n$ about the x_i form a countable base: Indeed, if U is an open set and $x \in U$ then take n sufficiently large so that $B_{2/n}(x) \subset U$. Choose j so that $d(x_j, x) < 1/n$. Then $V := B_{1/n}(x_j)$ satisfies $x \in V \subset U$ so by Proposition 1.6.3 the set of balls $B_{1/n}(x_j)$ form a base and they constitute a countable set. Conversely, let $\mathcal{B}$ be a countable base, and choose a point $x_j \in U_j$ for each $U_j \in \mathcal{B}$. If x is any point of X, the ball of radius $\epsilon > 0$ about x includes some U_j and hence contains x_j. So the x_j form a countable dense set. QED

Proposition 1.7.2 Lindelof's theorem. *Suppose that the topological space X is second countable. Then every open cover has a countable subcover.*

Let $\mathcal{U}$ be a cover, not necessarily countable, and let $\mathcal{B}$ be a countable base. Let $\mathcal{C} \subset \mathcal{B}$ consist of those open sets V belonging to $\mathcal{B}$ which are such that $V \subset U$ where $U \in \mathcal{U}$. By Proposition 1.6.3 these form a (countable) cover. For each $V \in \mathcal{C}$ choose a $U_V \in \mathcal{U}$ such that $V \subset U_V$. Then the $\{U_V\}_{V \in \mathcal{C}}$ form a countable subset of $\mathcal{U}$ which is a cover. QED

1.8   Conclusion of the proof of Theorem 1.5.1.

Suppose that condition 2. and 3. of the theorem hold for the metric space X. By Proposition 1.6.2, X is separable, and hence by Proposition 1.7.1, X is

second countable. Hence by Proposition 1.7.2, every cover $\mathcal{U}$ has a countable subcover. So we must prove that if U_1, U_2, U_3, ... is a sequence of open sets which cover X, then $X = U_1 \cup U_2 \cup \cdots \cup U_m$ for some finite integer m. Suppose not. For each m choose $x_m \in X$ with $x_m \notin U_1 \cup \cdots \cup U_m$. By condition 2. of Theorem 1.5.1, we may choose a subsequence of the $\{x_j\}$ which converge to some point x. Since $U_1 \cup \cdots \cup U_m$ is open, its complement is closed, and since $x_j \notin U_1 \cup \cdots \cup U_m$ for $j > m$ we conclude that $x \notin U_1 \cup \cdots \cup U_m$ for any m. This says that the $\{U_j\}$ do *not* cover X, a contradiction. QED

Putting the pieces together, we see that a closed bounded subset of $\mathbb{R}^m$ is compact. This is the famous Heine-Borel theorem. So Theorem 1.5.1 can be considered as a far reaching generalization of the Heine-Borel theorem.

1.9 Dini's lemma.

Let X be a metric space and let L denote the space of real valued continuous functions of compact support. So $f \in L$ means that f is continuous, and the closure of the set of all x for which $|f(x)| > 0$ is compact. Thus L is a real vector space, and $f \in L \Rightarrow |f| \in L$. Thus if $f \in L$ and $g \in L$ then $f + g \in L$ and also max $(f, g) = \frac{1}{2}(f + g + |f - g|) \in L$ and min $(f, g) = \frac{1}{2}(f + g - |f - g|) \in L$.

For a sequence of elements in L (or more generally in any space of real valued functions) we write $f_n \downarrow 0$ to mean that the sequence of functions is monotone decreasing, and at each x we have $f_n(x) \to 0$.

Theorem 1.9.1 Dini's lemma. *If $f_n \in L$ and $f_n \downarrow 0$ then $\|f_n\|_\infty \to 0$. In other words, monotone decreasing convergence to 0 implies uniform convergence to zero for elements of L.*

Proof. Given $\epsilon > 0$, let $C_n = \{x | f_n(x) \geq \epsilon\}$. Then the C_n are compact, $C_n \supset C_{n+1}$ and $\bigcap_k C_k = \emptyset$. Hence a finite intersection is already empty, which means that $C_n = \emptyset$ for some n. This means that $\|f_n\|_\infty \leq \epsilon$ for some n, and hence, since the sequence is monotone decreasing, for all subsequent n. QED

1.10 The Lebesgue outer measure of an interval is its length.

For any subset $A \subset \mathbb{R}$ we define its **Lebesgue outer measure** by

$$m^*(A) := \inf \sum \ell(I_n) : \ I_n \text{ are intervals with } A \subset \bigcup I_n. \qquad (1.1)$$

Here the length $\ell(I)$ of any interval $I = [a, b]$ is $b - a$ with the same definition for half open intervals $(a, b]$ or $[a, b)$, or open intervals. Of course if $a = -\infty$ and b is finite or $+\infty$, or if a is finite and $b = +\infty$ the length is infinite. So the infimum in (1.1) is taken over all covers of A by intervals. By the usual $\epsilon/2^n$ trick, i.e. by replacing each $I_j = [a_j, b_j]$ by $(a_j - \epsilon/2^{j+1}, b_j + \epsilon/2^{j+1})$ we may

assume that the infimum is taken over open intervals. (Equally well, we could use half open intervals of the form $[a, b)$, for example.).

It is clear that if $A \subset B$ then $m^*(A) \leq m^*(B)$ since any cover of B by intervals is a cover of A. Also, if Z is any set of measure zero, then $m^*(A \cup Z) = m^*(A)$. In particular, $m^*(Z) = 0$ if Z has measure zero. Also, if $A = [a, b]$ is an interval, then we can cover it by itself, so

$$m^*([a, b]) \leq b - a,$$

and hence the same is true for $(a, b]$, $[a, b)$, or (a, b). If the interval is infinite, it clearly can not be covered by a set of intervals whose total length is finite, since if we lined them up with end points touching they could not cover an infinite interval. We still must prove that

$$m^*(I) = \ell(I) \tag{1.2}$$

if I is a finite interval. We may assume that $I = [c, d]$ is a closed interval by what we have already said, and that the minimization in (1.1) is with respect to a cover by open intervals. So what we must show is that if

$$[c, d] \subset \bigcup_i (a_i, b_i)$$

then

$$d - c \leq \sum_i (b_i - a_i).$$

We first apply Heine-Borel to replace the countable cover by a finite cover. (This only decreases the right hand side of preceding inequality.) So let n be the number of elements in the cover. We want to prove that if

$$[c, d] \subset \bigcup_{i=1}^n (a_i, b_i) \ \text{ then } d - c \leq \sum_{i=1}^n (b_i - a_i).$$

We shall do this by induction on n. If $n = 1$ then $a_1 < c$ and $b_1 > d$ so clearly $b_1 - a_1 > d - c$.

Suppose that $n \geq 2$ and we know the result for all covers (of all intervals $[c, d]$) with at most $n - 1$ intervals in the cover. If some interval (a_i, b_i) is disjoint from $[c, d]$ we may eliminate it from the cover, and then we are in the case of $n - 1$ intervals. So every (a_i, b_i) has non-empty intersection with $[c, d]$. Among the the intervals (a_i, b_i) there will be one for which a_i takes on the minimum possible value. By relabeling, we may assume that this is (a_1, b_1). Since c is covered, we must have $a_1 < c$. If $b_1 > d$ then (a_1, b_1) covers $[c, d]$ and there is nothing further to do. So assume $b_1 \leq d$. We must have $b_1 > c$ since $(a_1, b_1) \cap [c, d] \neq \emptyset$. Since $b_1 \in [c, d]$, at least one of the intervals (a_i, b_i), $i > 1$ contains the point b_1. By relabeling, we may assume that it is (a_2, b_2). But now we have a cover of $[c, d]$ by $n - 1$ intervals:

$$[c, d] \subset (a_1, b_2) \cup \bigcup_{i=3}^n (a_i, b_i).$$

So by induction

$$d - c \le (b_2 - a_1) + \sum_{i=3}^{n}(b_i - a_i).$$

But $b_2 - a_1 \le (b_2 - a_2) + (b_1 - a_1)$ since $a_2 < b_1$. QED

1.11 Zorn's lemma and the axiom of choice.

In the first few sections we repeatedly used an argument which involved "choosing" this or that element of a set. That we can do so is an axiom known as

The axiom of choice. *If F is a function with domain D such that $F(x)$ is a non-empty set for every $x \in D$, then there exists a function f with domain D such that $f(x) \in F(x)$ for every $x \in D$.*

It has been proved by Gödel that if mathematics is consistent without the axiom of choice (a big "if"!) then mathematics remains consistent with the axiom of choice added.

In fact, it will be convenient for us to take a slightly less intuitive axiom as out starting point:

Zorn's lemma. *Every partially ordered set A has a maximal linearly ordered subset. If every linearly ordered subset of A has an upper bound, then A contains a maximum element.*

The second assertion is a consequence of the first. For let B be a maximum linearly ordered subset of A, and x an upper bound for B. Then x is a maximum element of A, for if $y \succ x$ then we could add y to B to obtain a larger linearly ordered set. Thus there is no element in A which is strictly larger than x which is what we mean when we say that x is a maximum element.

Zorn's lemma implies the axiom of choice.

Indeed, consider the set A of all functions g defined on subsets of D such that $g(x) \in F(x)$. We will let dom(g) denote the domain of definition of g. The set A is not empty, for if we pick a point $x_0 \in D$ and pick $y_0 \in F(x_0)$, then the function g whose domain consists of the single point x_0 and whose value $g(x_0) = y_0$ gives an element of A. Put a partial order on A by saying that $g \preceq h$ if dom(g) $\subset$ dom(h) and the restriction of h to dom g coincides with g. A linearly ordered subset means that we have an increasing family of domains X, with functions h defined consistently with respect to restriction. But this means that there is a function g defined on the union of these domains, $\bigcup X$ whose restriction to each X coincides with the corresponding h. This is clearly an upper bound. So A has a maximal element f. If the domain of f were not

all of D we could add a single point x_0 not in the domain of f and $y_0 \in F(x_0)$ contradicting the maximality of f. QED

1.12 The Baire category theorem.

Theorem 1.12.1 *In a complete metric space any countable intersection of dense open sets is dense.*

Proof. Let X be the space, let B be an open ball in X, and let $O_1,\ O_2 \dots$ be a sequence of dense open sets. We must show that

$$B \cap \left(\bigcap_n O_n \right) \neq \emptyset.$$

Since O_1 is dense, $B \cap O_1 \neq \emptyset$, and is open. Thus $B \cap O_1$ contains the closure $\overline{B_1}$ of some open ball B_1. We may choose B_1 (smaller if necessary) so that its radius is < 1. Since B_1 is open and O_2 is dense, $B_1 \cap O_2$ contains the closure $\overline{B_2}$ of some open ball B_2, of radius $< \frac{1}{2}$, and so on. Since X is complete, the intersection of the decreasing sequence of closed balls we have constructed contains some point x which belong both to B and to the intersection of all the O_i. QED

A **Baire space** is defined as a topological space in which every countable intersection of dense open sets is dense. Thus Baire's theorem asserts that every complete metric space is a Baire space. A set A in a topological space is called **nowhere dense** if its closure contains no open set. Put another way, a set A is nowhere dense if its complement A^c contains an open dense set. A set S is said to be of **first category** if it is a countable union of nowhere dense sets. Then Baire's category theorem can be reformulated as saying that the complement of any set of first category in a complete metric space (or in any Baire space) is dense. A property P of points of a Baire space is said to hold **quasi - surely** or **quasi-everywhere** if it holds on an intersection of countably many dense open sets. In other words, if the set where P does *not* hold is of first category.

1.13 Tychonoff's theorem.

Let I be a set, serving as an "index set". Suppose that for each $\alpha \in I$ we are given a non-empty topological space S_α. The Cartesian product

$$S := \prod_{\alpha \in I} S_\alpha$$

is defined as the collection of all functions x whose domain in I and such that $x(\alpha) \in S_\alpha$. This space is not empty by the axiom of choice. We frequently write x_α instead of $x(\alpha)$ and called x_α the "α coordinate of x".

The map

$$f_\alpha : \prod_{\alpha \in I} S_\alpha \to S_\alpha, \quad x \mapsto x_\alpha$$

is called the **projection** of S onto S_α. We put on S the weakest topology such that all of these projections are continuous. So the open sets of S are generated by the sets of the form

$$f_\alpha^{-1}(U_\alpha) \text{ where } U_\alpha \subset S_\alpha \text{ is open.}$$

Theorem 1.13.1 [Tychonoff.] *If all the S_α are compact, then so is $S = \prod_{\alpha \in I} S_\alpha$.*

Proof. Let $\mathcal{F}$ be a family of closed subsets of S with the property that the intersection of any finite collection of subsets from this family is not empty. We must show that the intersection of all the elements of $\mathcal{F}$ is not empty. Using Zorn, extend $\mathcal{F}$ to a maximal family $\mathcal{F}_0$ of (not necessarily closed) subsets of S with the property that the intersection of any finite collection of elements of $\mathcal{F}_0$ is not empty. For each α, the projection $f_\alpha(\mathcal{F}_0)$ has the property that there is a point $x_\alpha \in S_\alpha$ which is in the closure of all the sets belonging to $f_\alpha(\mathcal{F}_0)$. Let $x \in S$ be the point whose α-th coordinate is x_α. We will show that x is in the closure of every element of $\mathcal{F}_0$ which will complete the proof.

Let U be an open set containing x. By the definition of the product topology, there are finitely many α_i and open subsets $U_{\alpha_i} \subset S_{\alpha_i}$ such that

$$x \in \bigcap_{i=1}^{n} f_{\alpha_i}^{-1}(U_{\alpha_i}) \subset U.$$

So for each $i = 1, \ldots, n$, $x_{\alpha_i} \in U_{\alpha_i}$. This means that U_{α_i} intersects every set belonging to $f_{\alpha_i}(\mathcal{F}_0)$. So $f_{\alpha_i}^{-1}(U_{\alpha_i})$ intersects every set belonging to $\mathcal{F}_0$ and hence must belong to $\mathcal{F}_0$ by maximality. Therefore,

$$\bigcap_{i=1}^{n} f_{\alpha_i}^{-1}(U_{\alpha_i}) \in \mathcal{F}_0,$$

again by maximality. This says that U intersects every set of $\mathcal{F}_0$. In other words, any neighborhood of x intersects every set belonging to $\mathcal{F}_0$, which is just another way of saying x belongs to the closure of every set belonging to $\mathcal{F}_0$. QED

1.14 Urysohn's lemma.

A topological space S is called **normal** if it is Hausdorff, and if for any pair F_1, F_2 of closed sets with $F_1 \cap F_2 = \emptyset$ there are disjoint open sets U_1, U_2 with $F_1 \subset U_1$ and $F_2 \subset U_2$. For example, suppose that S is Hausdorff and compact. For each $p \in F_1$ and $q \in F_2$ there are neighborhoods O_q of p and W_q of q with $O_q \cap W_q = \emptyset$. This is the Hausdorff axiom. A finite number of the W_q cover F_2 since it is compact. Let the intersection of the corresponding O_q be called U_p and the union of the corresponding W_q be called V_p. Thus for each $p \in F_1$ we have found a neighborhood U_p of p and an open set V_p containing F_2 with

$U_p \cap V_p = \emptyset$. Once again, finitely many of the U_p cover F_1. So the union U of these and the intersection V of the corresponding V_p give disjoint open sets U containing F_1 and V containing F_2. So any compact Hausdorff space is normal.

Theorem 1.14.1 [Urysohn's lemma.] *If F_0 and F_1 are disjoint closed sets in a normal space S then there is a continuous real valued function $f : S \to \mathbb{R}$ such that $0 \le f \le 1$, $f = 0$ on F_0 and $f = 1$ on F_1.*

Proof.   Let

$$V_1 := F_1^c.$$

We can find an open set $V_{\frac{1}{2}}$ containing F_0 and whose closure is contained in V_1, since we can choose $V_{\frac{1}{2}}$ disjoint from an open set containing F_1. So we have

$$F_0 \subset V_{\frac{1}{2}}, \quad \overline{V_{\frac{1}{2}}} \subset V_1.$$

Applying our normality assumption to the sets F_0 and $V_{\frac{1}{2}}^c$ we can find an open set $V_{\frac{1}{4}}$ with $F_0 \subset V_{\frac{1}{4}}$ and $\overline{V_{\frac{1}{4}}} \subset V_{\frac{1}{2}}$. Similarly, we can find an open set $V_{\frac{3}{4}}$ with $\overline{V_{\frac{1}{2}}} \subset V_{\frac{3}{4}}$ and $\overline{V_{\frac{3}{4}}} \subset V_1$. So we have

$$F_0 \subset V_{\frac{1}{4}}, \ \overline{V_{\frac{1}{4}}} \subset V_{\frac{1}{2}}, \ \overline{V_{\frac{1}{2}}} \subset V_{\frac{3}{4}}, \ \overline{V_{\frac{3}{4}}} \subset V_1 = F_1^c.$$

Continuing in this way, for each $0 < r < 1$ where r is a dyadic rational, $r = m/2^k$ we produce an open set V_r with $F_0 \subset V_r$ and $\overline{V_r} \subset V_s$ if $r < s$, including $\overline{V_r} \subset V_1 = F_1^c$. Define f as follows: Set $f(x) = 1$ for $x \in F_1$. Otherwise, define

$$f(x) = \inf\{r | x \in V_r\}.$$

So $f = 0$ on F_0.
   If $0 < b \le 1$, then $f(x) < b$ means that $x \in V_r$ for some $r < b$. Thus

$$f^{-1}([0, b)) = \bigcup_{r < b} V_r.$$

This is a union of open sets, hence open. Similarly, $f(x) > a$ means that there is some $r > a$ such that $x \notin \overline{V_r}$. Thus

$$f^{-1}((a, 1]) = \bigcup_{r > a} (\overline{V_r})^c,$$

also a union of open sets, hence open. So we have shown that

$$f^{-1}([0, b)) \quad \text{and} \quad f^{-1}((a, 1])$$

are open. Hence $f^{-1}((a, b))$ is open. Since the intervals $[0, b), (a, 1]$ and (a, b) form a basis for the open sets on the interval $[0, 1]$, we see that the inverse image of any open set under f is open, which says that f is continuous. QED
   We will have several occasions to use this result.

1.15 The Stone-Weierstrass theorem.

This is an important generalization of Weierstrass's theorem which asserted that the polynomials are dense in the space of continuous functions on any compact interval, when we use the uniform topology. We shall have many uses for this theorem.

An algebra A of (real valued) functions on a set S is said to *separate points* if for any $p, q \in S$, $p \neq q$ there is an $f \in A$ with $f(p) \neq f(q)$.

Theorem 1.15.1 [Stone-Weierstrass.] *Let S be a compact space and A an algebra of continuous real valued functions on S which separates points. Then the closure of A in the uniform topology is either the algebra of all continuous functions on S, or is the algebra of all continuous functions on S which all vanish at a single point, call it x_∞.*

We will give two different proofs of this important theorem. For our first proof, we first state and prove some preliminary lemmas:

Lemma 1.15.1 *An algebra A of bounded real valued functions on a set S which is closed in the uniform topology is also closed under the lattice operations $\vee$ and $\wedge$.*

Proof. Since $f \vee g = \frac{1}{2}(f + g + |f - g|)$ and $f \wedge g = \frac{1}{2}(f + g - |f - g|)$ we must show that

$$f \in A \quad \Rightarrow |f| \in A.$$

Replacing f by $f/\|f\|_\infty$ we may assume that

$$|f| \leq 1.$$

The Taylor series expansion about the point $\frac{1}{2}$ for the function $t \mapsto (t + \epsilon^2)^{\frac{1}{2}}$ converges uniformly on $[0, 1]$. So there exists, for any $\epsilon > 0$ there is a polynomial P such that

$$|P(x^2) - (x^2 + \epsilon^2)^{\frac{1}{2}}| < \epsilon \quad \text{on} \ \ [-1, 1].$$

Let

$$Q := P - P(0).$$

We have $|P(0) - \epsilon| < \epsilon$ so

$$|P(0)| < 2\epsilon.$$

So $Q(0) = 0$ and

$$|Q(x^2) - (x^2 + \epsilon^2)^{\frac{1}{2}}| < 3\epsilon.$$

But

$$(x^2 + \epsilon^2)^{\frac{1}{2}} - |x| \leq \epsilon$$

for small ϵ. So

$$|Q(x^2) - |x| \,| < 4\epsilon \quad \text{on} \quad [0, 1].$$

As Q does not contain a constant term, and A is an algebra, $Q(f^2) \in A$ for any $f \in A$. Since we are assuming that $|f| \leq 1$ we have

$$Q(f^2) \in A, \quad \text{and} \quad \|Q(f^2) - |f| \|_\infty < 4\epsilon.$$

Since we are assuming that A is closed under $\| \cdot \|_\infty$ we conclude that $|f| \in A$ completing the proof of the lemma.

Lemma 1.15.2 *Let A be a set of real valued continuous functions on a compact space S such that*

$$f, g \in A \Rightarrow \quad f \wedge g \in A \quad and \quad f \vee g \in A.$$

Then the closure of A in the uniform topology contains every continuous function on S which can be approximated at every pair of points by a function belonging to A.

Proof.   Suppose that f is a continuous function on S which can be approximated at any pair of points by elements of A. So let $p, q \in S$ and $\epsilon > 0$, and let $f_{p,q,\epsilon} \in A$ be such that

$$|f(p) - f_{p,q,\epsilon}(p)| < \epsilon, \quad |f(q) - f_{p,q,\epsilon}(q)| < \epsilon.$$

Let

$$U_{p,q,\epsilon} := \{x | f_{p,q,\epsilon}(x) < f(x) + \epsilon\}, \quad V_{p,q,\epsilon} := \{x | f_{p,q,\epsilon}(x) > f(x) - \epsilon\}.$$

Fix q and ϵ. The sets $U_{p,q,\epsilon}$ cover S as p varies. Hence a finite number cover S since we are assuming that S is compact. We may take the minimum $f_{q,\epsilon}$ of the corresponding finite collection of $f_{p,q,\epsilon}$. The function $f_{q,\epsilon}$ has the property that

$$f_{q,\epsilon}(x) < f(x) + \epsilon$$

and

$$f_{q,\epsilon}(x) > f(x) - \epsilon$$

for

$$x \in \bigcap_p V_{p,q,\epsilon}$$

where the intersection is again over the same finite set of p's. We have now found a collection of functions $f_{q,\epsilon}$ such that

$$f_{q,\epsilon} < f + \epsilon$$

and $f_{q,\epsilon} > f - \epsilon$ on some neighborhood $V_{q,\epsilon}$ of q. We may choose a finite number of q so that the $V_{q,\epsilon}$ cover all of S. Taking the maximum of the corresponding $f_{q,\epsilon}$ gives a function $f_\epsilon \in A$ with $f - \epsilon < f < f + \epsilon$, i.e.

$$\|f - f_\epsilon\|_\infty < \epsilon.$$

Since we are assuming that A is closed in the uniform topology we conclude that $f \in A$, completing the proof of the lemma.

Proof of the Stone-Weierstrass theorem. Suppose first that for every $x \in S$ there is a $g \in A$ with $g(x) \neq 0$. Let $x \neq y$ and $h \in A$ with $h(y) \neq 0$. Then we may choose real numbers c and d so that $f = cg + dh$ is such that

$$0 \neq f(x) \neq f(y) \neq 0.$$

Then for any real numbers a and b we may find constants A and B such that

$$Af(x) + Bf^2(x) = a \quad \text{and} \quad Af(y) + Bf^2(y) = b.$$

We can therefore approximate (in fact hit exactly on the nose) any function at any two distinct points. We know that the closure of A is closed under $\vee$ and $\wedge$ by the first lemma. By the second lemma we conclude that the closure of A is the algebra of all real valued continuous functions.

The second alternative is that there is a point, call it p_∞ at which all $f \in A$ vanish. We wish to show that the closure of A contains all continuous functions vanishing at p_∞. Let B be the algebra obtained from A by adding the constants. Then B satisfies the hypotheses of the Stone-Weierstrass theorem and contains functions which do not vanish at p_∞. so we can apply the preceding result. If g is a continuous function vanishing at p_∞ we may, for any $\epsilon > 0$ find an $f \in A$ and a constant c so that

$$\|g - (f + c)\|_\infty < \frac{\epsilon}{2}.$$

Evaluating at p_∞ gives $|c| < \epsilon/2$. So

$$\|g - f\|_\infty < \epsilon.$$

QED

The reason for the apparently strange notation p_∞ has to do with the notion of the one point compactification of a locally compact space. A topological space S is called **locally compact** if every point has a closed compact neighborhood. We can make S compact by adding a single point. Indeed, let p_∞ be a point not belonging to S and set

$$S_\infty := S \cup p_\infty.$$

We put a topology on S_∞ by taking as the open sets all the open sets of S together with all sets of the form

$$O \cup p_\infty$$

where O is an open set of S whose complement is compact. The space S_∞ is compact, for if we have an open cover of S_∞, at least one of the open sets in this cover must be of the second type, hence its complement is compact, hence covered by finitely many of the remaining sets. If S itself is compact, then the empty set has compact complement, hence p_∞ has an open neighborhood

disjoint from S, and all we have done is add a disconnected point to S. The space S_∞ is called the **one-point compactification** of S. In applications of the Stone-Weierstrass theorem, we shall frequently have to do with an algebra of functions on a locally compact space consisting of functions which "vanish at infinity" in the sense that for any $\epsilon > 0$ there is a compact set C such that $|f| < \epsilon$ on the complement of C. We can think of these functions as being defined on S_∞ and all vanishing at p_∞.

We now turn to a second proof of this important theorem.

1.16   Machado's theorem.

Let $\mathcal{M}$ be a compact space and let $C_\mathbb{R}(\mathcal{M})$ denote the algebra of continuous real valued functions on $\mathcal{M}$. We let $\|\cdot\| = \|\cdot\|_\infty$ denote the uniform norm on $C_\mathbb{R}(\mathcal{M})$. More generally, for any closed set $F \subset \mathcal{M}$, we let

$$\|f\|_F = \sup_{x \in F} |f(x)|$$

so $\|\cdot\| = \|\cdot\|_\mathcal{M}$.

If $A \subset C_\mathbb{R}(\mathcal{M})$ is a collection of functions, we will say that a subset $E \subset \mathcal{M}$ is a **level set** (for A) if all the elements of A are constant on the set E. Also, for any $f \in C_\mathbb{R}(\mathcal{M})$ and any closed set $F \subset \mathcal{M}$, we let

$$d_f(F) := \inf_{g \in A} \|f - g\|_F.$$

So $d_f(F)$ measures how far f is from the elements of A on the set F. (I have suppressed the dependence on A in this notation.) We can look for "small" closed subsets which measure how far f is from A on all of $\mathcal{M}$; that is we look for closed sets with the property that

$$d_f(E) = d_f(\mathcal{M}). \tag{1.3}$$

Let $\mathcal{F}$ denote the collection of all non-empty closed subsets of $\mathcal{M}$ with this property. Clearly $\mathcal{M} \in \mathcal{F}$ so this collection is not empty. We order $\mathcal{F}$ by the reverse of inclusion: $F_1 \prec F_2$ if $F_1 \supset F_2$. Let $\mathcal{C}$ be a totally ordered subset of $\mathcal{F}$. Since $\mathcal{M}$ is compact, the intersection of any nested family of non-empty closed sets is again non-empty. We claim that the intersection of all the sets in $\mathcal{C}$ belongs to $\mathcal{F}$, i.e. satisfies (1.3). Indeed, since $d_f(F) = d_f(\mathcal{M})$ for any $F \in \mathcal{C}$ this means that for any $g \in A$, the sets

$$\{x \in F \,|\, |f(x) - g(x)| \geq d_f((M))\}$$

are non-empty. They are also closed and nested, and hence have a non-empty intersection. So on the set

$$E = \bigcap_{F \in \mathcal{C}} F$$

we have

$$\|f - g\|_E \geq d_f(\mathcal{M}).$$

So every chain has an upper bound, and hence by Zorn's lemma, there exists a maximum, i.e. there exists a non-empty closed subset E satisfying (1.3) which has the property that no proper subset of E satisfies (1.3). We shall call such a subset f-**minimal**.

Theorem 1.16.1 [Machado.] *Suppose that $A \subset C_\mathbb{R}(\mathcal{M})$ is a subalgebra which contains the constants and which is closed in the uniform topology. Then for every $f \in C_\mathbb{R}(\mathcal{M})$ there exists an A level set satisfying(1.3). In fact, every f-minimal set is an A level set.*

Proof.   Let E be an f-minimal set. Suppose it is not an A level set. This means that there is some $h \in A$ which is not constant on A. Replacing h by $ah + c$ where a and c are constant, we may arrange that

$$\min_{x \in E} h = 0 \quad \text{and} \quad \max_{x \in E} h = 1.$$

Let

$$E_0 := \{x \in E | 0 \le h(x) \le \frac{2}{3}\} \quad \text{and} \quad E_1 : \{x \in E | \frac{1}{3} \le x \le 1\}.$$

These are non-empty closed proper subsets of E, and hence the minimality of E implies that there exist $g_0, g_1 \in A$ such that

$$\|f - g_0\|_{E_0} \le d_f(\mathcal{M}) - \epsilon \quad \text{and} \quad \|f - g_1\|_{E_1} \le d_f(\mathcal{M}) - \epsilon$$

for some $\epsilon > 0$. Define

$$h_n := (1 - h^n)^{2^n} \quad \text{and} \quad k_n := h_n g_0 + (1 - h_n) g_1.$$

Both h_n and k_n belong to A and $0 \le h_n \le 1$ on E, with strict inequality on $E_0 \cap E_1$. At each $x \in E_0 \cap E_1$ i we have

$$
\begin{aligned}
|f(x) - k_n(x)| &= |h_n(x)f(x) - h_n(x)g_0(x) + (1 - h_n(x))f(x) - (1 - h_n(x))g_1(x)| \\
&\le h_n(x)\|f - g_0\|_{E_0 \cap E_1} + (1 - h_n)\|f - g_1\|_{E_0 \cap E_1} \\
&\le h_n(x)\|f - g_0\|_{E_0} + (1 - h_n(x)\|f - g_1\|_{E_1} \le d_f(\mathcal{M}) - \epsilon.
\end{aligned}
$$

We will now show that $h_n \to 1$ on $E_0 \setminus E_1$ and $h_n \to 0$ on $E_1 \setminus E_0$. Indeed, on $E_0 \setminus E_1$ we have

$$h^n < \left(\frac{1}{3}\right)^n$$

so

$$h_n = (1 - h^n)^{2^n} \ge 1 - 2^n h^n \ge 1 - \left(\frac{2}{3}\right)^n \to 1$$

since the binomial formula gives an alternating sum with decreasing terms. On the other hand,

$$h_n(1 + h^n)^{2^n} = 1 - h^{2 \cdot 2^n} \le 1$$

or

$$h_n \le \frac{1}{(1 + h^n)^{2^n}}.$$

Now the binomial formula implies that for any integer k and any positive number a we have $ka \le (1+a)^k$ or $(1+a)^{-k} \le 1/(ka)$. So we have

$$h_n \le \frac{1}{2^n h^n}.$$

On $E_0 \setminus E_1$ we have $h^n \ge \left(\frac{2}{3}\right)^n$ so there we have

$$h_n \le \left(\frac{3}{4}\right)^n \to 0.$$

Thus $k_n \to g_0$ uniformly on $E_0 \setminus E_1$ and $k_n \to g_1$ uniformly on $E_1 \setminus E_0$. We conclude that for n large enough

$$\|f - k_n\|_E < d_f(\mathcal{M})$$

contradicting our assumption that $d_f(E) = d_f(\mathcal{M})$. QED

Corollary 1.16.1 [The Stone-Weierstrass Theorem.] *If A is a uniformly closed subalgebra of $C_{\mathbf{R}}(\mathcal{M})$ which contains the constants and separates points, then $A = C_{\mathbb{R}}(\mathcal{M})$.*

Proof.   The only A-level sets are points. But since $\|f - f(a)\|_{\{a\}} = 0$, we conclude that $d_f(\mathcal{M}) = 0$, i.e. $f \in A$ for any $f \in C_{\mathbb{R}}(\mathcal{M})$. QED

1.17   The Hahn-Banach theorem.

This says:

Theorem 1.17.1 [Hahn-Banach]. *Let M be a subspace of a normed linear space B, and let F be a bounded linear function on M. Then F can be extended so as to be defined on all of B without increasing its norm.*

Proof by Zorn.   Suppose that we can prove

Proposition 1.17.1 *Let M be a subspace of a normed linear space B, and let F be a bounded linear function on on M. Let $y \in B, y \notin M$. Then F can be extended to $M + \{y\}$ without changing its norm.*

Then we could order the extensions of F by inclusion, one extension being $\succeq$ than another if it is defined on a larger space. The extension defined on the union of any family of subspaces ordered by inclusion is again an extension, and so is an upper bound. The proposition implies that a maximal extension must be defined on the whole space, otherwise we can extend it further. So we must prove the proposition.

I was careful in the statement not to specify whether our spaces are over the real or complex numbers. Let us first assume that we are dealing with a real vector space, and then deduce the complex case.

We want to choose a value

$$\alpha = F(y)$$

so that if we then define

$$F(x + \lambda y) := F(x) + \lambda F(y) = F(x) + \lambda \alpha, \quad \forall x \in M, \lambda \in \mathbb{R}$$

we do not increase the norm of F. If $F = 0$ we take $\alpha = 0$. If $F \neq 0$, we may replace F by $F/\|F\|$, extend and then multiply by $\|F\|$ so without loss of generality we may assume that $\|F\| = 1$. We want to choose the extension to have norm 1, which means that we want

$$|F(x) + \lambda \alpha| \leq \|x + \lambda y\| \quad \forall x \in M, \lambda \in \mathbb{R}.$$

If $\lambda = 0$ this is true by hypothesis. If $\lambda \neq 0$ divide this inequality by λ and replace $(1/\lambda)x$ by x. We want

$$|F(x) + \alpha| \leq \|x + y\| \quad \forall \, x \in M.$$

We can write this as two separate conditions:

$$F(x_2) + \alpha \leq \|x_2 + y\| \;\; \forall \, x_2 \in M \;\; \text{and} \;\; -F(x_1) - \alpha \leq \|x_1 + y\| \;\; \forall x_1 \in M.$$

Rewriting the second inequality this becomes

$$-F(x_1) - \|x_1 + y\| \leq \alpha \leq -F(x_2) + \|x_2 + y\|.$$

The question is whether such a choice is possible. In other words, is the supremum of the left hand side (over all $x_1 \in M$) less than or equal to the infimum of the right hand side (over all $x_2 \in M$)? If the answer to this question is yes, we may choose α to be any value between the sup of the left and the inf of the right hand sides of the preceding inequality. So our question is: Is

$$F(x_2) - F(x_1)| \leq \|x_2 + y\| + \|x_1 + y\| \quad \forall \, x_1, x_2 \in M?$$

But $x_1 - x_2 = (x_1 + y) - (x_2 + y)$ and so using the fact that $\|F\| = 1$ and the triangle inequality gives

$$|F(x_2) - F(x_1)| \leq \|x_2 - x_1\| \leq \|x_2 + y\| + \|x_1 + y\|.$$

This completes the proof of the proposition, and hence of the Hahn-Banach theorem over the real numbers.

We now deal with the complex case. If B is a complex normed vector space, then it is also a real vector space, and the real and imaginary parts of a complex linear function are real linear functions. In other words, we can write any complex linear function F as

$$F(x) = G(x) + iH(x)$$

where G and H are real linear functions. The fact that F is complex linear says that $F(ix) = iF(x)$ or

$$G(ix) = -H(x)$$

or

$$H(x) = -G(ix)$$

or

$$F(x) = G(x) - iG(ix).$$

The fact that $\|F\| = 1$ implies that $\|G\| \leq 1$. So we can adjoin the real one dimensional space spanned by y to M and extend the real linear function to it, keeping the norm ≤ 1. Next adjoin the real one dimensional space spanned by iy and extend G to it. We now have G extended to $M \oplus \mathbf{C}y$ with no increase in norm. Try to define

$$F(z) := G(z) - iG(iz)$$

on $M \oplus \mathbf{C}y$. This map of $M \oplus \mathbf{C}y \to \mathbf{C}$ is $\mathbf{R}$-linear, and coincides with F on M. We must check that it is complex linear and that its norm is ≤ 1: To check that it is complex linear it is enough to observe that

$$F(iz) = G(iz) - iG(-z) = i[G(z) - iG(iz)] = iF(z).$$

To check the norm, we may, for any z, choose θ so that $e^{i\theta}F(z)$ is real and is non-negative. Then

$$|F(z)| = |e^{i\theta}F(z)| = |F(e^{i\theta}z)| = G(e^{i\theta}z) \leq \|e^{i\theta}z\| = \|z\|$$

so $\|F\| \leq 1$. QED

Suppose that M is a closed subspace of B and that $y \notin M$. Let d denote the distance of y to M, so that

$$d := \inf_{x \in M} \|y - x\|.$$

Suppose we start with the zero function on M, and extend it first to $M \oplus y$ by

$$F(\lambda y - x) = \lambda d.$$

This is a linear function on $M + \{y\}$ and its norm is ≤ 1. Indeed

$$\|F\| = \sup_{\lambda, x} \frac{|\lambda d|}{\|\lambda y - x\|} = \sup_{x' \in M} \frac{d}{\|y - x'\|} = \frac{d}{d} = 1.$$

Let M^0 be the set of all continuous linear functions on B which vanish on M. Then, using the Hahn-Banach theorem we get

Proposition 1.17.2 *If $y \in B$ and $y \notin M$ where M is a closed linear subspace of B, then there is an element $F \in M^0$ with $\|F\| \leq 1$ and $F(y) \neq 0$. In fact we can arrange that $F(y) = d$ where d is the distance from y to M.*

We have an embedding

$$B \to B^{**} \quad x \mapsto x^{**} \quad \text{where } x^{**}(F) := F(x).$$

The first part of the preceding proposition can be formulated as

$$(M^0)^0 = M$$

if M is a closed subspace of B.

The map $x \mapsto x^{**}$ is clearly linear and

$$|x^{**}(F)| = |F(x)| \leq \|F\|\|x\|.$$

Taking the sup of $|x^{**}(F)|/\|F\|$ shows that

$$\|x^{**}\| \leq \|x\|$$

where the norm on the left is the norm on the space B^{**}. On the other hand, if we take $M = \{0\}$ in the preceding proposition, we can find an $F \in B^*$ with $\|F\| = 1$ and $F(x) = \|x\|$. For this F we have $|x^{**}(F)| = \|x\|$. So

$$\|x^{**}\| \geq \|x\|.$$

We have proved

Theorem 1.17.2 *The map $B \to B^{**}$ given above is a norm preserving injection.*

1.18   The Uniform Boundedness Principle.

Theorem 1.18.1 *Let $\mathbf{B}$ be a Banach space and $\{F_n\}$ be a sequence of elements in B^* such that for every fixed $x \in \mathbf{B}$ the sequence of numbers $\{|F_n(x)|\}$ is bounded. Then the sequence of norms $\{\|F_n\|\}$ is bounded.*

Proof.   The proof will be by a Baire category style argument. We will prove

Proposition 1.18.1 *There exists some ball $B = B(y, r), r > 0$ about a point y with $\|y\| \leq 1$ and a constant K such that $|F_n(z)| \leq K$ for all $z \in B$.*

Proof that the proposition implies the theorem. For any z with $\|z\| < 1$ we have

$$z - y = \frac{2}{r}\left(\frac{r}{2}(z - y)\right)$$

and

$$\|\frac{r}{2}(z - y)\| \leq r$$

since $\|z - y\| \leq 2$.

$$|F_n(z)| \leq |F_n(z - y)| + |F_n(y)| \leq \frac{2K}{r} + K.$$

So

$$\|F_n\| \leq \frac{2K}{r} + K$$

for all n proving the theorem from the proposition.

Proof of the proposition. If the proposition is false, we can find n_1 such that $|F_{n_1}(x)| > 1$ at some $x \in B(0,1)$ and hence in some ball of radius $\epsilon < \frac{1}{2}$ about x. Then we can find an n_2 with $|F_{n_2}(z)| > 2$ in some non-empty closed ball of radius $< \frac{1}{3}$ lying inside the first ball. Continuing inductively, we choose a subsequence n_m and a family of nested non-empty balls B_m with $|F_{n_m}(z)| > m$ throughout B_m and the radii of the balls tending to zero. Since B is complete, there is a point x common to all these balls, and $\{|F_n(x)|\}$ is unbounded, contrary to hypothesis. QED

We will have occasion to use this theorem in a "reversed form". Recall that we have the norm preserving injection $B \to B^{**}$ sending $x \mapsto x^{**}$ where $x^{**}(F) = F(x)$. Since B^* is a Banach space (even if B is incomplete) we have

Corollary 1.18.1 *If $\{x_n\}$ is a sequence of elements in a normed linear space such that the numerical sequence $\{|F(x_n)|\}$ is bounded for each fixed $F \in B^*$ then the sequence of norms $\{\|x_n\|\}$ is bounded.*

Chapter 2

Hilbert Spaces and Compact operators.

2.1 Hilbert space.

2.1.1 Scalar products.

V is a complex vector space. A rule assigning to every pair of vectors $f, g \in V$ a complex number (f, g) is called a **semi-scalar product** if

1. (f, g) is linear in f when g is held fixed.

2. $(g, f) = \overline{(f, g)}$. This implies that (f, g) is anti-linear in g when f is held fixed. In other words. $(f, ag + bh) = \bar{a}(f, g) + \bar{b}(f, h)$. It also implies that (f, f) is real.

3. $(f, f) \geq 0$ for all $f \in V$.

If 3. is replaced by the stronger condition

4. $(f, f) > 0$ for all non-zero $f \in V$

then we say that (,) is a **scalar product**.

Examples.

- $V = \mathbf{C}^n$, so an element $\mathbf{z}$ of V is a column vector of complex numbers:

$$\mathbf{z} = \begin{pmatrix} z_1 \\ \vdots \\ z_n \end{pmatrix}$$

and $(\mathbf{z}, \mathbf{w})$ is given by

$$(\mathbf{z}, \mathbf{w}) := \sum_1^n z_i \overline{w_i}.$$

33

- V consists of all continuous (complex valued) functions on the real line which are periodic of period 2π and

$$(f, g) := \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x)\overline{g(x)}dx.$$

We will denote this space by $\mathcal{C}(\mathbf{T})$. Here the letter $\mathbf{T}$ stands for the one dimensional torus, i.e. the circle. We are identifying functions which are periodic with period 2π with functions which are defined on the circle $\mathbf{R}/2\pi\mathbf{Z}$.

- V consists of all doubly infinite sequences of complex numbers

$$\mathbf{a} = \ldots, a_{-2}, a_{-1}, a_0, a_1, a_2, \ldots$$

which satisfy

$$\sum |a_i|^2 < \infty.$$

Here

$$(\mathbf{a}, \mathbf{b}) := \sum a_i \bar{b}_i.$$

All three are examples of scalar products.

2.1.2   The Cauchy-Schwartz inequality.

This says that if (,) is a semi-scalar product then

$$|(f, g)| \leq (f, f)^{\frac{1}{2}} (g, g)^{\frac{1}{2}}. \tag{2.1}$$

Proof. For any real number t condition 3. above says that $(f - tg, f - tg) \geq 0$. Expanding out gives

$$0 \leq (f - tg, f - tg) = (f, f) - t[(f, g) + (g, f)] + t^2(g, g).$$

Since $(g, f) = \overline{(f, g)}$, the coefficient of t in the above expression is twice the real part of (f, g). So the real quadratic form

$$Q(t) := (f, f) - 2\mathrm{Re}(f, g)t + t^2(g.g)$$

is nowhere negative. So it can not have distinct real roots, and hence by the $b^2 - 4ac$ rule we get

$$4(\mathrm{Re}(f, g))^2 - 4(f, f)(g, g) \leq 0$$

or

$$(\mathrm{Re}(f, g))^2 \leq (f, f)(g, g). \tag{2.2}$$

This is useful and almost but not quite what we want. But we may apply this inequality to $h = e^{i\theta}g$ for any θ. Then $(h, h) = (g, g)$. Choose θ so that

$$(f, g) = re^{i\theta}$$

where $r = |(f,g)|$. Then

$$(f,h) = (f, e^{i\theta}g) = e^{-i\theta}(f,g) = |(f,g)|$$

and the preceding inequality with g replaced by h gives

$$|(f,g)|^2 \le (f,f)(g,g)$$

and taking square roots gives (2.1).

2.1.3   The triangle inequality

For any semiscalar product define

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

so we can write the Cauchy-Schwartz inequality as

$$|(f,g)| \le \|f\|\|g\|.$$

The **triangle inequality** says that

$$\|f + g\| \le \|f\| + \|g\|. \tag{2.3}$$

Proof.

$$
\begin{aligned}
\|f + g\|^2 &= (f+g, f+g) \\
&= (f,f) + 2\mathrm{Re}(f,g) + (g,g) \\
&\le (f,f) + 2\|f\|\|g\| + (g,g) \quad \text{by (2.2)} \\
&= \|f\|^2 + 2\|f\|\|g\| + \|g\|^2 \\
&= (\|f\| + \|g\|)^2.
\end{aligned}
$$

Taking square roots gives the triangle inequality (2.3). Notice that

$$\|cf\| = |c|\|f\| \tag{2.4}$$

since $(cf, cf) = c\bar{c}(f,f) = |c|^2\|f\|^2$.

Suppose we try to define the distance between two elements of V by

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

Notice that then $d(f,f) = 0$, $d(f,g) = d(g,f)$ and for any three elements

$$d(f,h) \le d(f,g) + d(g,h)$$

by virtue of the triangle inequality. The only trouble with this definition is that we might have two distinct elements at zero distance, i.e. $0 = d(f,g) = \|f - g\|$. But this can not happen if $(\ ,\ )$ is a scalar product, i.e. satisfies condition 4.

A complex vector space V endowed with a scalar product is called a **pre-Hilbert** space.

Let V be a complex vector space and let $\|\cdot\|$ be a map which assigns to any $f \in V$ a non-negative real $\|f\|$ number such that $\|f\| > 0$ for all non-zero f. If $\|\cdot\|$ satisfies the triangle inequality (2.3) and equation (2.4) it is called a **norm**. A vector space endowed with a norm is called a normed space. The pre-Hilbert spaces can be characterized among all normed spaces by the parallelogram law as we will discuss below.

Later on, we will have to weaken condition (2.4) in our general study. But it is too complicated to give the general definition right now.

2.1.4   Hilbert and pre-Hilbert spaces.

The reason for the prefix "pre" is the following: The distance d defined above has all the desired properties we might expect of a distance. In particular, we can define the notions of "limit" and of a "Cauchy sequence" as is done for the real numbers: If f_n is a sequence of elements of V, and $f \in V$ we say that f is the limit of the f_n and write

$$\lim_{n \to \infty} f_n = f, \quad \text{or} \quad f_n \to f$$

if, for any positive number ϵ there is an $N = N(\epsilon)$ such that

$$d(f_n, f) < \epsilon \quad \text{for all } n \geq N.$$

If a sequence converges to some limit f, then this limit is unique, since any limits must be at zero distance and hence equal.

We say that a sequence of elements is **Cauchy** if for any $\delta > 0$ there is an $K = K(\delta)$ such that
$$d(f_m, f_n) < \delta \quad \forall, m, n \geq K.$$

If the sequence f_n has a limit, then it is Cauchy - just choose $K(\delta) = N(\frac{1}{2}\delta)$ and use the triangle inequality.

But it is quite possible that a Cauchy sequence has no limit. As an example of this type of phenomenon, think of the rational numbers with $|r - s|$ as the distance. The whole point of introducing the real numbers is to guarantee that every Cauchy sequence has a limit.

So we say that a pre-Hilbert space is a **Hilbert space** if it is "complete" in the above sense - if every Cauchy sequence has a limit.

Since the complex numbers are complete (because the real numbers are), it follows that $\mathbf{C}^n$ is complete, i.e. is a Hilbert space. Indeed, we can say that any finite dimensional pre-Hilbert space is a Hilbert space because it is isomorphic (as a pre-Hilbert space) to $\mathbf{C}^n$ for some n. (See below when we discuss orthonormal bases.)

The trouble is in the infinite dimensional case, such as the space of continuous periodic functions. This space is not complete. For example, let f_n be the function which is equal to one on $(-\pi + \frac{1}{n}, -\frac{1}{n})$, equal to zero on $(\frac{1}{n}, \pi - \frac{1}{n})$

and extended linearly $-\frac{1}{n}$ to $\frac{1}{n}$ and from $\pi - \frac{1}{n}$ to $\pi + \frac{1}{n}$ so as to be continuous and then extended so as to be periodic.(Thus on the interval $(\pi - \frac{1}{n}, \pi + \frac{1}{n})$ the function is given by $f_n(x) = \frac{n}{2}(x - (\pi - \frac{1}{n}))$.) If $m \le n$, the functions f_m and f_n agree outside two intervals of length $\frac{2}{m}$ and on these intervals $|f_m(x) - f_n(x)| \le 1$. So $\|f_m - f_n\|^2 \le \frac{1}{2\pi} \cdot 2/m$ showing that the sequence $\{f_n\}$ is Cauchy. But the limit would have to equal one on $(-\pi, 0)$ and equal zero on $(0, \pi)$ and so be discontinuous at the origin and at π. Thus the space of continuous periodic functions is not a Hilbert space, only a pre-Hilbert space.

But just as we complete the rational numbers (by throwing in "ideal" elements) to get the real numbers, we may similarly complete any pre-Hilbert space to get a unique Hilbert space. See the section *Completion* in the chapter on metric spaces for a general discussion of how to complete any metric space. In particular, the completion of any normed vector space is a complete normed vector space. A complete normed space is called a **Banach** space. The general construction implies that any normed vector space can be completed to a Banach space. From the parallelogram law discussed below, it will follow that the completion of a pre-Hilbert space is a Hilbert space.

The completion of the space of continuous periodic functions will be denoted by $L_2(\mathbf{T})$.

2.1.5 The Pythagorean theorem.

Let V be a pre-Hilbert space. We have

$$\|f + g\|^2 = \|f\|^2 + 2\mathrm{Re}(f, g) + \|g\|^2.$$

So

$$\|f + g\|^2 = \|f\|^2 + \|g\|^2 \quad \Leftrightarrow \quad \mathrm{Re}\,(f, g) = 0. \tag{2.5}$$

We make the definition

$$f \perp g \quad \Leftrightarrow \quad (f, g) = 0$$

and say that f is perpendicular to g or that f is orthogonal to g. Notice that this is a stronger condition than the condition for the Pythagorean theorem, the right hand condition in (2.5). For example $\|f + if\|^2 = 2\|f\|^2$ but $(f, if) = -i\|f\|^2 \ne 0$ if $\|f\| \ne 0$.

If u_i is some finite collection of mutually orthogonal vectors, then so are $z_i u_i$ where the z_i are any complex numbers. So if

$$u = \sum_i z_i u_i$$

then by the Pythagorean theorem

$$\|u\|^2 = \sum_i |z_i|^2 \|u_i\|^2.$$

In particular, if the $u_i \neq 0$, then $u = 0 \Rightarrow z_i = 0$ for all i. This shows that any set of mutually orthogonal (non-zero) vectors is linearly independent.

Notice that the set of functions

$$e^{in\theta}$$

is an **orthonormal** set in the space of continuous periodic functions in that not only are they mutually orthogonal, but each has norm one.

2.1.6   The theorem of Apollonius.

Adding the equations

$$\|f + g\|^2 \;\; = \;\; \|f\|^2 + 2\mathrm{Re}\,(f, g) + \|g\|^2 \tag{2.6}$$
$$\|f - g\|^2 \;\; = \;\; \|f\|^2 - 2\mathrm{Re}\,(f, g) + \|g\|^2 \tag{2.7}$$

gives

$$\|f + g\|^2 + \|f - g\|^2 = 2\left(\|f\|^2 + \|g\|^2\right). \tag{2.8}$$

This is known as the **parallelogram law**. It is the algebraic expression of the theorem of Apollonius which asserts that the sum of the areas of the squares on the sides of a parallelogram equals the sum of the areas of the squares on the diagonals.

If we subtract (2.7) from (2.6) we get

$$\mathrm{Re}(f, g) = \frac{1}{4}\left(\|f + g\|^2 - \|f - g\|^2\right). \tag{2.9}$$

Now $(if, g) = i(f, g)$ and $\mathrm{Re}\,\{i(f, g)\} = -\mathrm{Im}(f, g)$ so

$$\mathrm{Im}(f, g) = -\mathrm{Re}(if, g) = \mathrm{Re}(f, ig)$$

so

$$(f, g) = \frac{1}{4}\left(\|f + g\|^2 - \|f - g\|^2 + i\|f + ig\|^2 - i\|f - ig\|^2\right). \tag{2.10}$$

If we now complete a pre-Hilbert space, the right hand side of this equation is defined on the completion, and is a continuous function there. It therefore follows that the scalar product extends to the completion, and, by continuity, satisfies all the axioms for a scalar product, plus the completeness condition for the associated norm. In other words, the completion of a pre-Hilbert space is a Hilbert space.

2.1.7   The theorem of Jordan and von Neumann.

This is essentially a converse to the theorem of Apollonius. It says that if $\|\cdot\|$ is a norm on a (complex) vector space V which satisfies (2.8), then V is in fact a pre-Hilbert space with $\|f\|^2 = (f, f)$. If the theorem is true, then the scalar product must be given by (2.10). So we must prove that if we take (2.10) as the

definition, then all the axioms on a scalar product hold. The easiest axiom to verify is
$$(g, f) = \overline{(f, g)}.$$
Indeed, the real part of the right hand side of (2.10) is unchanged under the interchange of f and g (since $g - f = -(f - g)$ and $\| - h\| = \|h\|$ for any h is one of the properties of a norm). Also $g + if = i(f - ig)$ and $\|ih\| = \|h\|$ so the last two terms on the right of (2.10) get interchanged, proving that $(g, f) = \overline{(f, g)}$.

It is just as easy to prove that

$$(if, g) = i(f, g).$$

Indeed replacing f by if sends $\|f + ig\|^2$ into $\|if + ig\|^2 = \|f + g\|^2$ and sends $\|f + g\|^2$ into $\|if + g\|^2 = \|i(f - ig)\|^2 = \|f - ig\|^2 = i(-i\|f - ig\|^2)$ so has the effect of multiplying the sum of the first and fourth terms by i, and similarly for the sum of the second and third terms on the right hand side of (2.10).

Now (2.10) implies (2.9). Suppose we replace f, g in (2.8) by $f_1 + g, f_2$ and by $f_1 - g, f_2$ and subtract the second equation from the first. We get

$$\|f_1 + f_2 + g\|^2 - \|f_1 + f_2 - g\|^2 + \|f_1 - f_2 + g\|^2 - \|f_1 - f_2 - g\|^2$$

$$= 2 \left(\|f_1 + g\|^2 - \|f_1 - g\|^2 \right).$$

In view of (2.9) we can write this as

$$\mathrm{Re}\ (f_1 + f_2, g) + \mathrm{Re}\ (f_1 - f_2, g) = 2\mathrm{Re}\ (f_1, g). \tag{2.11}$$

Now the right hand side of (2.9) vanishes when $f = 0$ since $\|g\| = \| - g\|$. So if we take $f_1 = f_2 = f$ in (2.11) we get

$$\mathrm{Re}\ (2f, g) = 2\mathrm{Re}\ (f, g).$$

We can thus write (2.11) as

$$\mathrm{Re}\ (f_1 + f_2, g) + \mathrm{Re}\ (f_1 - f_2, g) = \mathrm{Re}\ (2f_1, g).$$

In this equation make the substitutions

$$f_1 \mapsto \frac{1}{2}(f_1 + f_2), \quad f_2 \mapsto \frac{1}{2}(f_1 - f_2).$$

This yields

$$\mathrm{Re}\ (f_1 + f_2, g) = \mathrm{Re}\ (f_1, g) + \mathrm{Re}\ (f_2, g).$$

Since it follows from (2.10) and (2.9) that

$$(f, g) = \mathrm{Re}\ (f, g) - i\mathrm{Re}\ (if, g)$$

we conclude that

$$(f_1 + f_2, g) = (f_1, g) + (f_2, g).$$

Taking $f_1 = -f_2$ shows that

$$(-f, g) = -(f, g).$$

Consider the collection $\mathcal{C}$ of complex numbers α which satisfy

$$(\alpha f, g) = \alpha(f, g)$$

(for all f, g). We know from $(f_1 + f_2, g) = (f_1, g) + (f_2, g)$ that

$$\alpha, \beta \in \mathcal{C} \Rightarrow \alpha + \beta \in \mathcal{C}.$$

So $\mathcal{C}$ contains all integers. If $0 \neq \beta \in \mathcal{C}$ then

$$(f, g) = (\beta \cdot (1/\beta)f, g) = \beta((1/\beta)f, g)$$

so $\beta^{-1} \in \mathcal{C}$. Thus $\mathcal{C}$ contains all (complex) rational numbers. The theorem will be proved if we can prove that $(\alpha f, g)$ is continuous in α. But the triangle inequality

$$\|f + g\| \leq \|f\| + \|g\|$$

applied to $f = f_2, g = f_1 - f_2$ implies that

$$\|f_1\| \leq \|f_1 - f_2\| + \|f_2\|$$

or

$$\|f_1\| - \|f_2\| \leq \|f_1 - f_2\|.$$

Interchanging the role of f_1 and f_2 gives

$$\big|\, \|f_1\| - \|f_2\| \,\big| \leq \|f_1 - f_2\|.$$

Therefore

$$\big|\, \|\alpha f \pm g\| - \|\beta f \pm g\| \,\big| \leq \|(\alpha - \beta)f\|.$$

Since $\|(\alpha - \beta)f\| \to 0$ as $\alpha \to \beta$ this shows that the right hand side of (2.10) when applied to αf and g is a continuous function of α. Thus $\mathcal{C} = \mathbf{C}$. We have proved

Theorem 2.1.1 [P. Jordan and J. von Neumann] *If V is a normed space whose norm satisfies (2.8) then V is a pre-Hilbert space.*

Notice that the condition (2.8) involves only two vectors at a time. So we conclude as an immediate consequence of this theorem that

Corollary 2.1.1 *A normed vector space is pre-Hilbert space if and only if every two dimensional subspace is a Hilbert space in the induced norm.*

Actually, a weaker version of this corollary, with two replaced by three had been proved by Fréchet, *Annals of Mathematics*, July 1935, who raised the problem of giving an abstract characterization of those norms on vector spaces which come from scalar products. In the immediately following paper Jordan and von Neumann proved the theorem above leading to the stronger corollary that two dimensions suffice.

2.1.8 Orthogonal projection.

We continue with the assumption that V is pre-Hilbert space. If A and B are two subsets of V, we write $A \perp B$ if $u \in A$ and $v \in B \Rightarrow u \perp v$, in other words if every element of A is perpendicular to every element of B. Similarly, we will write $v \perp A$ if the element v is perpendicular to all elements of A. Finally, we will write $A^\perp$ for the set of all v which satisfy $v \perp A$. Notice that $A^\perp$ is always a linear subspace of V, for any A.

Now let M be a (linear) subspace of V. Let v be some element of V, not necessarily belonging to M. We want to investigate the problem of finding a $w \in M$ such that $(v - w) \perp M$. Of course, if $v \in M$ then the only choice is to take $w = v$. So the interesting problem is when $v \notin M$. Suppose that such a w exists, and let x be any (other) point of M. Then by the Pythagorean theorem,

$$\|v - x\|^2 = \|(v - w) + (w - x)\|^2 = \|v - w\|^2 + \|w - x\|^2$$

since $(v - w) \perp M$ and $(w - x) \in M$. So

$$\|v - w\| \le \|v - x\|$$

and this inequality is strict if $x \ne w$. In words: if we can find a $w \in M$ such that $(v - w) \perp M$ then w is the unique solution of the problem of finding the point in M which is closest to v. Conversely, suppose we found a $w \in M$ which has this minimization property, and let x be any element of M. Then for any real number t we have

$$\|v - w\|^2 \le \|(v - w) + tx\|^2 = \|v - w\|^2 + 2t\mathrm{Re}\ (v - w, x) + t^2\|x\|^2.$$

Since the minimum of this quadratic polynomial in t occurring on the right is achieved at $t = 0$, we conclude (by differentiating with respect to t and setting $t = 0$, for example) that

$$\mathrm{Re}\ (v - w, x) = 0.$$

By our usual trick of replacing x by $e^{i\theta}x$ we conclude that

$$(v - w, x) = 0.$$

Since this holds for all $x \in M$, we conclude that $(v - w) \perp M$. So to find w we search for the minimum of $\|v - x\|$, $x \in M$.

Now $\|v - x\| \ge 0$ and is some finite number for any $x \in M$. So there will be some real number m such that $m \le \|v - x\|$ for $x \in M$, and such that no strictly larger real number will have this property. (m is known as the "greatest lower bound" of the values $\|v - x\|$, $x \in M$.) So we can find a sequence of vectors $x_n \in M$ such that

$$\|v - x_n\| \to m.$$

We claim that the x_n form a Cauchy sequence. Indeed,

$$\|x_i - x_j\|^2 = \|(v - x_j) - (v - x_i)\|^2$$

and by the parallelogram law this equals

$$2 \left(\|v - x_i\|^2 + \|v - x_j\|^2 \right) - \|2v - (x_i + x_j)\|^2.$$

Now the expression in parenthesis converges to $2m^2$. The last term on the right is

$$-\|2(v - \frac{1}{2}(x_i + x_j)\|^2.$$

Since $\frac{1}{2}(x_i + x_j) \in M$, we conclude that

$$\|2v - (x_i + x_j)\|^2 \geq 4m^2$$

so

$$\|x_i - x_j\|^2 \leq 4(m + \epsilon)^2 - 4m^2$$

for i and j large enough that $\|v - x_i\| \leq m + \epsilon$ and $\|v - x_j\| \leq m + \epsilon$. This proves that the sequence x_n is Cauchy.

Here is the crux of the matter: If M is complete, then we can conclude that the x_n converge to a limit w which is then the unique element in M such that $(v - w) \perp M$. It is at this point that completeness plays such an important role.

Put another way, we can say that if M is a subspace of V which is complete (under the scalar product (,) restricted to M) then we have the orthogonal direct sum decomposition

$$V = M \oplus M^{\perp},$$

which says that every element of V can be uniquely decomposed into the sum of an element of M and a vector perpendicular to M.

For example, if M is the one dimensional subspace consisting of all (complex) multiples of a non-zero vector y, then M is complete, since $\mathbf{C}$ is complete. So w exists. Since all elements of M are of the form ay, we can write $w = ay$ for some complex number a. Then $(v - ay, y) = 0$ or

$$(v, y) = a\|y\|^2$$

so

$$a = \frac{(v, y)}{\|y\|^2}.$$

We call a the **Fourier coefficient** of v with respect to y. Particularly useful is the case where $\|y\| = 1$ and we can write

$$a = (v, y). \tag{2.12}$$

Getting back to the general case, if $V = M \oplus M^{\perp}$ holds, so that to every v there corresponds a unique $w \in M$ satisfying $(v - w) \in M^{\perp}$ the map $v \mapsto w$ is called orthogonal projection of V onto M and will be denoted by π_M.

2.1.9 The Riesz representation theorem.

Let V and W be two complex vector spaces. A map

$$T : V \to W$$

is called **linear** if

$$T(\lambda x + \mu y) = \lambda T(x) + \mu T(Y) \quad \forall \; x, y \in V, \quad \lambda, \mu \in \mathbf{C}$$

and is called **anti-linear** if

$$T(\lambda x + \mu y) = \overline{\lambda} T(x) + \overline{\mu} T(Y) \quad \forall \; x, y \in V \quad \lambda, \mu \in \mathbf{C}.$$

If $\ell : V \to \mathbf{C}$ is a linear map, (also known as a linear function) then

$$\ker \ell := \{x \in V \mid \ell(x) = 0\}$$

has codimension one (unless $\ell \equiv 0$). Indeed, if

$$\ell(y) \neq 0$$

then

$$\ell(x) = 1 \quad \text{where } x = \frac{1}{\ell(y)} y$$

and for any $z \in V$,

$$z - \ell(z)x \in \ker \ell.$$

If V is a normed space and ℓ is continuous, then ker (ℓ) is a closed subspace. The space of continuous linear functions is denoted by V^*. It has its own norm defined by

$$\|\ell\| := \sup_{x \in V, \|x\| \neq 0} |\ell(x)| / \|x\|.$$

Suppose that H is a pre-hilbert space. Then we have an antilinear map

$$\phi : H \to H^*, \quad (\phi(g))(f) := (f, g).$$

The Cauchy-Schwartz inequality implies that

$$\|\phi(g)\| \leq \|g\|$$

and in fact

$$(g, g) = \|g\|^2$$

shows that

$$\|\phi(g)\| = \|g\|.$$

In particular the map ϕ is injective.

The Riesz representation theorem says that if H is a Hilbert space, then this map is surjective:

Theorem 2.1.2 *Every continuous linear function on H is given by scalar product by some element of H.*

The proof is a consequence of the theorem about projections applied to

$$N := \ker \ell :$$

If $\ell = 0$ there is nothing to prove. If $\ell \neq 0$ then N is a closed subspace of codimension one. Choose $v \notin N$. Then there is an $x \in N$ with $(v - x) \perp N$. Let

$$y := \frac{1}{\|v - x\|}(v - x).$$

Then

$$y \perp N$$

and

$$\|y\| = 1.$$

For any $f \in H$,

$$[f - (f, y)y] \perp y$$

so

$$f - (f, y)y \in N$$

or

$$\ell(f) = (f, y)\ell(y),$$

so if we set

$$g := \overline{\ell(y)}y$$

then

$$(f, g) = \ell(f)$$

for all $f \in H$. QED

2.1.10   What is $L_2(\mathbf{T})$?

We have defined the space $L^2(\mathbf{T})$ to be the completion of the space $\mathcal{C}(\mathbf{T})$ under the L_2 norm $\|f\|_2 = (f, f)^{\frac{1}{2}}$. In particular, every linear function on $\mathcal{C}(\mathbf{T})$ which is continuous with respect to the this L_2 norm extends to a unique continuous linear function on $L_2(\mathbf{T})$. By the Riesz representation theorem we know that every such continuous linear function is given by scalar product by an element of $L_2(\mathbf{T})$. Thus we may think of the elements of $L_2(\mathbf{T})$ as being the linear functions on $\mathcal{C}(\mathbf{T})$ which are continuous with respect to the L_2 norm. An element of $L_2(\mathbf{T})$ should not be thought of as a function, but rather as a linear function on the space of continuous functions relative to a special norm - the L_2 norm.

2.1.11 Projection onto a direct sum.

Suppose that the closed subspace M of a pre-Hilbert space is the orthogonal direct sum of a finite number of subspaces

$$M = \bigoplus_i M_i$$

meaning that the M_i are mutually perpendicular and every element x of M can be written as

$$x = \sum x_i, \quad x_i \in M_i.$$

(The orthogonality guarantees that such a decomposition is unique.) Suppose further that each M_i is such that the projection π_{M_i} exists. Then π_M exists and

$$\pi_M(v) = \sum \pi_{M_i}(v). \tag{2.13}$$

Proof. Clearly the right hand side belongs to M. We must show $v - \sum_i \pi_{M_i}(v)$ is orthogonal to every element of M. For this it is enough to show that it is orthogonal to each M_j since every element of M is a sum of elements of the M_j. So suppose $x_j \in M_j$. But $(\pi_{M_i} v, x_j) = 0$ if $i \neq j$. So

$$(v - \sum \pi_{M_i}(v), x_j) = (v - \pi_{M_j}(v), x_j) = 0$$

by the defining property of π_{M_j}.

2.1.12 Projection onto a finite dimensional subspace.

We now will put the equations (2.12) and (2.13) together: Suppose that M is a finite dimensional subspace with an orthonormal basis ϕ_i. This implies that M is an orthogonal direct sum of the one dimensional spaces spanned by the ϕ_i and hence π_M exists and is given by

$$\pi_M(v) = \sum a_i \phi_i \quad \text{where} \quad a_i = (v, \phi_i). \tag{2.14}$$

2.1.13 Bessel's inequality.

We now look at the infinite dimensional situation and suppose that we are given an orthonormal sequence $\{\phi_i\}_1^\infty$. Any $v \in V$ has its Fourier coefficients

$$a_i = (v, \phi_i)$$

relative to the members of this sequence. Bessel's inequality asserts that

$$\sum_1^\infty |a_i|^2 \leq \|v\|^2, \tag{2.15}$$

in particular the sum on the left converges.

Proof. Let

$$v_n := \sum_{i=1}^n a_i \phi_i,$$

so that v_n is the projection of v onto the subspace spanned by the first n of the ϕ_i. In any event, $(v - v_n) \perp v_n$ so by the Pythagorean Theorem

$$\|v\|^2 = \|v - v_n\|^2 + \|v_n\|^2 = \|v - v_n\|^2 + \sum_{i=1}^n |a_i|^2.$$

This implies that

$$\sum_{i=1}^n |a_i|^2 \leq \|v\|^2$$

and letting $n \to \infty$ shows that the series on the left of Bessel's inequality converges and that Bessel's inequality holds.

2.1.14   Parseval's equation.

Continuing the above argument, observe that

$$\|v - v_n\|^2 \to 0 \ \Leftrightarrow \ \sum |a_i|^2 = \|v\|^2.$$

But $\|v - v_n\|^2 \to 0$ if and only if $\|v - v_n\| \to 0$ which is the same as saying that $v_n \to v$. But v_n is the n-th partial sum of the series $\sum a_i \phi_i$, and in the language of series, we say that a series converges to a limit v and write $\sum a_i \phi_i = v$ if and only if the partial sums approach v. So

$$\sum a_i \phi_i = v \quad \Leftrightarrow \quad \sum_i |a_i|^2 = \|v\|^2. \tag{2.16}$$

In general, we will call the series $\sum_i a_i \phi_i$ the Fourier series of v (relative to the given orthonormal sequence) whether or not it converges to v. Thus Parseval's equality says that the Fourier series of v converges to v if and only if $\sum |a_i|^2 = \|v\|^2$.

2.1.15   Orthonormal bases.

We still suppose that V is merely a pre-Hilbert space. We say that an orthonormal sequence $\{\phi_i\}$ is a **basis** of V if every element of V is the sum of its Fourier series. For example, one of our tasks will be to show that the exponentials $\{e^{inx}\}_{n=-\infty}^{\infty}$ form a basis of $\mathcal{C}(\mathbf{T})$.

If the orthonormal sequence ϕ_i is a basis, then any v can be approximated as closely as we like by finite linear combinations of the ϕ_i, in fact by the partial sums of its Fourier series. We say that the finite linear combinations of the ϕ_i are *dense* in V. Conversely, suppose that the finite linear combinations of the

ϕ_i are dense in V. This means that for any v and any $\epsilon > 0$ we can find an n and a set of n complex numbers b_i such that

$$\|v - \sum b_i \phi_i\| \le \epsilon.$$

But we know that v_n is the closest vector to v among all the linear combinations of the first n of the ϕ_i. so we must have

$$\|v - v_n\| \le \epsilon.$$

But this says that the Fourier series of v converges to v, i.e. that the ϕ_i form a basis. For example, we know from Fejer's theorem that the exponentials e^{ikx} are dense in $\mathcal{C}(\mathbf{T})$. Hence we know that they form a basis of the pre-Hilbert space $\mathcal{C}(\mathbf{T})$. We will give some alternative proofs of this fact below.

In the case that V is actually a Hilbert space, and not merely a pre-Hilbert space, there is an alternative and very useful criterion for an orthonormal sequence to be a basis: Let M be the set of all limits of finite linear combinations of the ϕ_i. Any Cauchy sequence in M converges (in V) since V is a Hilbert space, and this limit belongs to M since it is itself a limit of finite linear combinations of the ϕ_i (by the diagonal argument for example). Thus $V = M \oplus M^{\perp}$, and the ϕ_i form a basis of M. So the ϕ_i form a basis of V if and only if $M^{\perp} = \{0\}$. But this is the same as saying that no non-zero vector is orthogonal to all the ϕ_i. So we have proved

Proposition 2.1.1 *In a Hilbert space, the orthonormal set $\{\phi_i\}$ is a basis if and only if no non-zero vector is orthogonal to all the ϕ_i.*

2.2 Self-adjoint transformations.

We continue to let V denote a pre-Hilbert space. Let T be a linear transformation of V into itself. This means that for every $v \in V$ the vector $Tv \in V$ is defined and that Tv depends linearly on v : $T(av + bw) = aTv + bTw$ for any two vectors v and w and any two complex numbers a and b. We recall from linear algebra that a non-zero vector v is called an eigenvector of T if Tv is a scalar times v, in other words if $Tv = \lambda v$ where the number λ is called the corresponding eigenvalue.

A linear transformation T on V is called **symmetric** if for any pair of elements v and w of V we have

$$(Tv, w) = (v, Tw).$$

Notice that if v is an eigenvector of a symmetric transformation T with eigenvalue λ, then

$$\lambda(v, v) = (\lambda v, v) = (Tv, v) = (v, Tv) = (v, \lambda v) = \overline{\lambda}(v, v),$$

so $\lambda = \overline{\lambda}$. In other words, all eigenvalues of a symmetric transformation are real.

We will let $\mathbf{S} = \mathbf{S}(V)$ denote the "unit sphere" of V, i.e. $\mathbf{S}$ denotes the set of all $\phi \in V$ such that $\|\phi\| = 1$. A linear transformation T is called **bounded** if $\|T\phi\|$ is bounded as ϕ ranges over all of $\mathbf{S}$. If T is bounded, we let

$$\|T\| := \max_{\phi \in \mathbf{S}} \|T\phi\|.$$

Then

$$\|Tv\| \le \|T\| \|v\|$$

for all $v \in V$. A linear transformation on a finite dimensional space is automatically bounded, but not so for an infinite dimensional space.

Also, for any linear transformation T, we will let $N(T)$ denote the kernel of T, so

$$N(T) = \{v \in V | Tv = 0\}$$

and $R(T)$ denote the range of T, so

$$R(T) := \{v | v = Tw \ \text{ for some } \ w \in V\}.$$

Both $N(T)$ and $R(T)$ are linear subspaces of V.

For bounded transformations, the phrase "self-adjoint" is synonymous with "symmetric". Later on we will need to study non-bounded (not everywhere defined) symmetric transformations, and then a rather subtle and important distinction will be made between self-adjoint transformations and those which are merely symmetric. But for the rest of this section we will only be considering bounded linear transformations, and so we will freely use the phrase "self-adjoint", and (usually) drop the adjective "bounded" since all our transformations will be assumed to be bounded.

We denote the set of all (bounded) self-adjoint transformations by $\mathcal{A}$, or by $\mathcal{A}(V)$ if we need to make V explicit.

2.2.1   Non-negative self-adjoint transformations.

If T is a self-adjoint transformation, then

$$(Tv, v) = (v, Tv) = \overline{(Tv, v)}$$

so (Tv, v) is always a real number. More generally, for any pair of elements v and w,

$$(Tv, w) = \overline{(Tw, v)}.$$

Since (Tv, w) depends linearly on v for fixed w, we see that the rule which assigns to every pair of elements v and w the number (Tv, w) satisfies the first two conditions in our definition of a semi-scalar product. Since (Tv, v) might be negative, condition 3. of the definition need not be satisfied. This leads to the following definition:

A self-adjoint transformation T is called **non-negative** if

$$(Tv, v) \ge 0 \quad \forall \ v \in V.$$

So if T is a non-negative self-adjoint transformation, then the rule which assigns to every pair of elements v and w the number (Tv, w) *is* a semi-scalar product to which we may apply the Cauchy-Schwartz inequality and conclude that

$$|(Tv, w)| \leq (Tv, v)^{\frac{1}{2}} (Tw, w)^{\frac{1}{2}}.$$

Now let us assume in addition that T is bounded with norm $\|T\|$. Let us take $w = Tv$ in the preceding inequality. We get

$$\|Tv\|^2 = |(Tv, Tv)| \leq (Tv, v)^{\frac{1}{2}} (TTv, Tv)^{\frac{1}{2}}.$$

Now apply the Cauchy-Schwartz inequality for the original scalar product to the last factor on the right:

$$(TTv, Tv)^{\frac{1}{2}} \leq \|TTv\|^{\frac{1}{2}} \|Tv\|^{\frac{1}{2}} \leq \|T\|^{\frac{1}{2}} \|Tv\|^{\frac{1}{2}} \|Tv\|^{\frac{1}{2}} = \|T\|^{\frac{1}{2}} \|Tv\|,$$

where we have used the defining property of $\|T\|$ in the form $\|TTv\| \leq \|T\|\|Tv\|$. Substituting this into the previous inequality we get

$$\|Tv\|^2 \leq (Tv, v)^{\frac{1}{2}} \|T\|^{\frac{1}{2}} \|Tv\|.$$

If $\|Tv\| \neq 0$ we may divide this inequality by $\|Tv\|$ to obtain

$$\|Tv\| \leq \|T\|^{\frac{1}{2}} (Tv, v)^{\frac{1}{2}}. \tag{2.17}$$

This inequality is clearly true if $\|Tv\| = 0$ and so holds in all cases.

We will make much use of this inequality. For example, it follows from (2.17) that

$$(Tv, v) = 0 \quad \Rightarrow \quad Tv = 0. \tag{2.18}$$

It also follows from (2.17) that if we have a sequence $\{v_n\}$ of vectors with $(Tv_n, v_n) \to 0$ then $\|Tv_v\| \to 0$ and so

$$(Tv_n, v_n) \to 0 \quad \Rightarrow \quad Tv_n \to 0. \tag{2.19}$$

Notice that if T is a bounded self adjoint transformation, not necessarily non-negative, then $rI - T$ is a non-negative self-adjoint transformation if $r \geq \|T\|$: Indeed,

$$((rI - T)v, v) = r(v, v) - (Tv, v) \geq (r - \|T\|)(v, v) \geq 0$$

since, by Cauchy-Schwartz,

$$(Tv, v) \leq |(Tv, v)| \leq \|Tv\|\|v\| \leq \|T\|\|v\|^2 = \|T\|(v, v).$$

So we may apply the preceding results to $rI - T$.

2.3   Compact self-adjoint transformations.

We say that the self-adjoint transformation T is **compact** if it has the following property: Given any sequence of elements $u_n \in \mathbf{S}$, we can choose a subsequence u_{n_i} such that the sequence Tu_{n_i} converges to a limit in V.

Some remarks about this complicated looking definition: In case V is finite dimensional, every linear transformation is bounded, hence the sequence Tu_n lies in a bounded region of our finite dimensional space, and hence by the completeness property of the real (and hence complex) numbers, we can always find such a convergent subsequence. So in finite dimensions every T is compact. More generally, the same argument shows that if $R(T)$ is finite dimensional and T is bounded then T is compact. So the definition is of interest essentially in the case when $R(T)$ is infinite dimensional.

Also notice that if T is compact, then T is bounded. Otherwise we could find a sequence u_n of elements of $\mathbf{S}$ such that $\|Tu_n\| \geq n$ and so no subsequence Tu_{n_i} can converge.

We now come to the key result which we will use over and over again:

Theorem 2.3.1 *Let T be a compact self-adjoint operator. Then $R(T)$ has an orthonormal basis $\{\phi_i\}$ consisting of eigenvectors of T and if $R(T)$ is infinite dimensional then the corresponding sequence $\{r_n\}$ of eigenvalues converges to 0.*

Proof. We know that T is bounded. If $T = 0$ there is nothing to prove. So assume that $T \neq 0$ and let

$$m_1 := \|T\| > 0.$$

By the definition of $\|T\|$ we can find a sequence of vectors $u_n \in \mathbf{S}$ such that $\|Tu_n\| \to \|T\|$. By the definition of compactness we can find a subsequence of this sequence so that $Tu_{n_i} \to w$ for some $w \in V$. On the other hand, the transformation T^2 is self-adjoint and bounded by $\|T\|^2$. Hence $\|T\|^2 I - T^2$ is non-negative, and

$$((\|T\|^2 I - T^2)u_n, u_n) = \|T\|^2 - \|Tu_n\|^2 \to 0.$$

So we know from (2.19) that

$$\|T\|^2 u_n - T^2 u_n \to 0.$$

Passing to the subsequence we have $T^2 u_{n_i} = T(Tu_{n_i}) \to Tw$ and so

$$\|T\|^2 u_{n_i} \to Tw$$

or

$$u_{n_i} \to \frac{1}{m_1^2} Tw.$$

Applying T to this we get

$$Tu_{n_i} \to \frac{1}{m_1^2} T^2 w$$

or

$$T^2 w = m_1^2 w.$$

Also $\|w\| = \|T\| = m_1 \neq 0$. So $w \neq 0$. So w is an eigenvector of T^2 with eigenvalue m_1^2. We have

$$0 = (T^2 - m_1^2)w = (T + m_1)(T - m_1)w.$$

If $(T - m_1)w = 0$, then w is an eigenvector of T with eigenvalue m_1 and we normalize by setting

$$\phi_1 := \frac{1}{\|w\|} w.$$

Then $\|\phi_1\| = 1$ and

$$T\phi_1 = m_1\phi_1.$$

If $(T - m_1)w \neq 0$ then $y := (T - m_1)w$ is an eigenvector of T with eigenvalue $-m_1$ and again we normalize by setting

$$\phi_1 := \frac{1}{\|y\|} y.$$

So we have found a unit vector $\phi_1 \in R(T)$ which is an eigenvector of T with eigenvalue $r_1 = \pm m_1$.

Now let

$$V_2 := \phi_1^\perp.$$

If $(w, \phi_1) = 0$, then

$$(Tw, \phi_1) = (w, T\phi_1) = r_1(w, \phi_1) = 0.$$

In other words,

$$T(V_2) \subset V_2$$

and we can consider the linear transformation T restricted to V_2 which is again compact. If we let m_2 denote the norm of the linear transformation T when restricted to V_2 then $m_2 \leq m_1$ and we can apply the preceding procedure to find a unit eigenvector ϕ_2 with eigenvalue $\pm m_2$.

We proceed inductively, letting

$$V_n := \{\phi_1, \ldots, \phi_{n-1}\}^\perp$$

and find an eigenvector ϕ_n of T restricted to V_n with eigenvalue $\pm m_n \neq 0$ if the restriction of T to V_n is not zero. So there are two alternatives:

- after some finite stage the restriction of T to V_n is zero. In this case $R(T)$ is finite dimensional with orthonormal basis $\phi_1, \ldots, \phi_{n-1}$. Or

- The process continues indefinitely so that at each stage the restriction of T to V_n is not zero and we get an infinite sequence of eigenvectors and eigenvalues r_i with $|r_i| \geq |r_{i+1}|$.

The first case is one of the alternatives in the theorem, so we need to look at the second alternative.

We first prove that $|r_n| \to 0$. If not, there is some $c > 0$ such that $|r_n| \geq c$ for all n (since the $|r_n|$ are decreasing). If $i \neq j$, then by the Pythagorean theorem we have

$$\|T\phi_i - T\phi_j\|^2 = \|r_i\phi_i - r_j\phi_j\|^2 = r_i^2\|\phi_i\|^2 + r_j^2\|\phi_j\|^2.$$

Since $\|\phi_i\| = \|\phi_j\| = 1$ this gives

$$\|T\phi_i - T\phi_j\|^2 = r_i^2 + r_j^2 \geq 2c^2.$$

Hence no subsequence of the $T\phi_i$ can converge, since all these vectors are at least a distance $c\sqrt{2}$ apart. This contradicts the compactness of T.

To complete the proof of the theorem we must show that the ϕ_i form a basis of $R(T)$. So if $w = Tv$ we must show that the Fourier series of w with respect to the ϕ_i converges to w. We begin with the Fourier coefficients of v relative to the ϕ_i which are given by

$$a_n = (v, \phi_n).$$

Then the Fourier coefficients of w are given by

$$b_i = (w, \phi_i) = (Tv, \phi_i) = (v, T\phi_i) = (v, r_i\phi_i) = r_i a_i.$$

So

$$w - \sum_{i=1}^n b_i\phi_i = Tv - \sum_{i=1}^n a_i r_i \phi_i = T(v - \sum_{i=1}^n a_i\phi_i).$$

Now $v - \sum_{i=1}^n a_i\phi_i$ is orthogonal to $\phi_1, \ldots, \phi_n$ and hence belongs to V_{n+1}. So

$$\|T(v - \sum_{i=1}^n a_i\phi_i)\| \leq |r_{n+1}| \|(v - \sum_{i=1}^n a_i\phi_i)\|.$$

By the Pythagorean theorem,

$$\|(v - \sum_{i=1}^n a_i\phi_i)\| \leq \|v\|.$$

Putting the two previous inequalities together we get

$$\|w - \sum_{i=1}^n b_i\phi_i\| = \|T(v - \sum_{i=1}^n a_i\phi_i)\| \leq |r_{n+1}| \|v\| \to 0.$$

This proves that the Fourier series of w converges to w concluding the proof of the theorem.

The "converse" of the above result is easy. Here is a version: Suppose that **H** is a Hilbert space with an orthonormal basis $\{\phi_i\}$ consisting of eigenvectors

of an operator T, so $T\phi_i = \lambda_i\phi_i$, and suppose that $\lambda_i \to 0$ as $i \to \infty$. Then T is compact. Indeed, for each j we can find an $N = N(j)$ such that

$$|\lambda_r| < \frac{1}{j} \quad \forall\ r > N(j).$$

We can then let $\mathbf{H}_j$ denote the closed subspace spanned by all the eigenvectors $\phi_r, r > N(j)$, so that
$$\mathbf{H} = \mathbf{H}_j^{\perp} \oplus \mathbf{H}_j$$
is an orthogonal decomposition and $\mathbf{H}_j^{\perp}$ is finite dimensional, in fact is spanned the first $N(j)$ eigenvectors of T.

Now let $\{u_i\}$ be a sequence of vectors with $\|u_i\| \leq 1$ say. We decompose each element as
$$u_i = u_i' \oplus u_i'', \quad u_i' \in \mathbf{H}_1^{\perp},\ u_i'' \in \mathbf{H}_j.$$

We can choose a subsequence so that u_{i_k}' converges, because they all belong to a finite dimensional space, and hence so does Tu_{i_k} since T is bounded. We can decompose every element of this subsequence into its $\mathbf{H}_2^{\perp}$ and $\mathbf{H}_2$ components, and choose a subsequence so that the first component converges. Proceeding in this way, and then using the Cantor diagonal trick of choosing the k-th term of the k-th selected subsequence, we have found a subsequence such that for any fixed j, the (now relabeled) subsequence, the $\mathbf{H}_j^{\perp}$ component of Tu_j converges. But the $\mathbf{H}_j$ component of Tu_j has norm less than $1/j$, and so the sequence converges by the triangle inequality.

2.4 Fourier's Fourier series.

We want to apply the theorem about compact self-adjoint operators that we proved in the preceding section to conclude that the functions e^{inx} form an orthonormal basis of the space $\mathcal{C}(\mathbf{T})$. In fact, a direct proof of this fact is elementary, using integration by parts. So we will pause to given this direct proof. Then we will go back and give a (more complicated) proof of the same fact using our theorem on compact operators. The reason for giving the more complicated proof is that it extends to far more general situations.

2.4.1 Proof by integration by parts.

We have let $\mathcal{C}(\mathbf{T})$ denote the space of continuous functions on the real line which are periodic with period 2π. We will let $\mathcal{C}^1(\mathbf{T})$ denote the space of periodic functions which have a continuous first derivative (necessarily periodic) and by $\mathcal{C}^2(\mathbf{T})$ the space of periodic functions with two continuous derivatives. If f and g both belong to $\mathcal{C}^1(\mathbf{T})$ then integration by parts gives

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} f'\overline{g}\,dx = -\frac{1}{2\pi}\int_{-\pi}^{\pi} f\overline{g}'\,dx$$

since the boundary terms, which normally arise in the integration by parts formula, cancel, due to the periodicity of f and g. If we take $g = e^{inx}/(in), n \neq 0$ the integral on the right hand side of this equation is the Fourier coefficient:

$$c_n = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x) e^{-inx} dx.$$

We thus obtain

$$c_n = \frac{1}{in} \frac{1}{2\pi} \int_{-\pi}^{\pi} f'(x) e^{-inx} dx$$

so, for $n \neq 0$,

$$|c_n| \leq \frac{A}{n} \quad \text{where } A := \frac{1}{2\pi} \int_{-\pi}^{\pi} |f'(x)| dx$$

is a constant independent of n (but depending on f).

If $f \in \mathcal{C}^2(\mathbf{T})$ we can take $g(x) = -e^{inx}/n^2$ and integrate by parts twice. We conclude that (for $n \neq 0$)

$$|c_n| \leq \frac{B}{n^2} \quad \text{where } B := \frac{1}{2\pi} \int_{-\pi}^{\pi} |f''(x)|^2$$

is again independent of n. But this proves that the Fourier series of f,

$$\sum c_n e^{inx}$$

converges uniformly and absolutely for and $f \in \mathcal{C}^2(\mathbf{T})$. The limit of this series is therefore some continuous periodic function. We must prove that this limit equals f. So we must prove that at each point f

$$\sum c_n e^{iny} \to f(y).$$

Replacing $f(x)$ by $f(x-y)$ it is enough to prove this formula for the case $y = 0$. So we must prove that for any $f \in \mathcal{C}^2(\mathbf{T})$ we have

$$\lim_{N,M \to \infty} \sum_{-N}^{M} c_n \to f(0).$$

Write $f(x) = (f(x) - f(0)) + f(0)$. The Fourier coefficients of any constant function c all vanish except for the c_0 term which equals c. So the above limit is trivially true when f is a constant. Hence, in proving the above formula, it is enough to prove it under the additional assumption that $f(0) = 0$, and we need to prove that in this case

$$\lim_{N,M \to \infty} (c_{-N} + c_{-N+1} + \cdots + c_M) \to 0.$$

The expression in parenthesis is

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} f(x) \overline{g_{N,M}(x)} dx$$

where

$$g_{N,M}(x) = e^{-iNx} + e^{-i(N-1)x} + \cdots + e^{iMx} = e^{-iNx}\left(1 + e^{ix} + \cdots + e^{i(M+N)x}\right) =$$

$$e^{-iNx}\frac{1 - e^{i(M+N+1)x}}{1 - e^{ix}} = \frac{e^{-iNx} - e^{i(M+1)x}}{1 - e^{ix}}, \quad x \neq 0$$

where we have used the formula for a geometric sum. By l'Hôpital's rule, this extends continuously to the value $M + N + 1$ for $x = 0$. Now $f(0) = 0$, and since f has two continuous derivatives, the function

$$h(x) := \frac{f(x)}{1 - e^{-ix}}$$

defined for $x \neq 0$ (or any multiple of 2π) extends, by l'Hôpital's rule, to a function defined at all values, and which is continuously differentiable and periodic. Hence the limit we are computing is

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} h(x)e^{iNx}dx - \frac{1}{2\pi}\int_{-\pi}^{\pi} h(x)e^{-i(M+1)x}dx$$

and we know that each of these terms tends to zero.

We have thus proved that the Fourier series of any twice differentiable periodic function converges uniformly and absolutely to that function. If we consider the space $\mathcal{C}^2(\mathbf{T})$ with our usual scalar product

$$(f, g) = \frac{1}{2\pi}\int_{-\pi}^{\pi} f\overline{g}dx$$

then the functions e^{inx} are dense in this space, since uniform convergence implies convergence in the $\|\ \|$ norm associated to $(\ ,\ )$. So, on general principles, Bessel's inequality and Parseval's equation hold.

It is not true in general that the Fourier series of a continuous function converges uniformly to that function (or converges at all in the sense of uniform convergence). However it is true that we *do* have convergence in the L_2 norm, i.e. the Hilbert space $\|\ \|$ norm on $\mathcal{C}(\mathbf{T})$. To prove this, we need only prove that the exponential functions e^{inx} are dense, and since they are dense in $\mathcal{C}^2(\mathbf{T})$, it is enough to prove that $\mathcal{C}^2(\mathbf{T})$ is dense in $\mathcal{C}(\mathbf{T})$. For this, let ϕ be a function defined on the line with at least two continuous bounded derivatives with $\phi(0) = 1$ and of total integral equal to one and which vanishes rapidly at infinity. A favorite is the Gauss normal function

$$\phi(x) := \frac{1}{\sqrt{2\pi}}e^{-x^2/2}$$

Equally well, we could take ϕ to be a function which actually vanishes outside of some neighborhood of the origin. Let

$$\phi_t(x) := \frac{1}{t}\phi\left(\frac{x}{t}\right).$$

As $t \to 0$ the function ϕ_t becomes more and more concentrated about the origin, but still has total integral one. Hence, for any bounded continuous function f, the function $\phi_t \star f$ defined by

$$(\phi_t \star f)(x) := \int_{-\infty}^{\infty} f(x-y)\phi_t(y)dy = \int_{-\infty}^{\infty} f(u)\phi_t(x-u)du.$$

satisfies $\phi_t \star f \to f$ uniformly on any finite interval. From the rightmost expression for $\phi_t \star f$ above we see that $\phi_t \star f$ has two continuous derivatives. From the first expression we see that $\phi_t \star f$ is periodic if f is. This proves that $\mathcal{C}^2(\mathbf{T})$ is dense in $\mathcal{C}(\mathbf{T})$. We have thus proved convergence in the L_2 norm.

2.4.2   Relation to the operator $\frac{d}{dx}$.

Each of the functions e^{inx} is an eigenvector of the operator

$$D = \frac{d}{dx}$$

in that

$$D\left(e^{inx}\right) = ine^{inx}.$$

So they are also eigenvalues of the operator D^2 with eigenvalues $-n^2$. Also, on the space of twice differentiable periodic functions the operator D^2 satisfies

$$(D^2 f, g) = \frac{1}{2\pi}\int_{-\pi}^{\pi} f''(x)\overline{g(x)}dx = f'(x)\overline{g(x)}\Big|_{-\pi}^{\pi} - \frac{1}{2\pi}\int_{-\pi}^{\pi} f'(x)\overline{g'(x)}dx$$

by integration by parts. Since f' and g are assumed to be periodic, the end point terms cancel, and integration by parts once more shows that

$$(D^2 f, g) = (f, D^2 g) = -(f', g').$$

But of course D and certainly D^2 is not defined on $\mathcal{C}(\mathbf{T})$ since some of the functions belonging to this space are not differentiable. Furthermore, the eigenvalues of D^2 are tending to infinity rather than to zero. So somehow the operator D^2 must be replaced with something like its inverse. In fact, we will work with the inverse of $D^2 - 1$, but first some preliminaries.

We will let $\mathcal{C}^2([-\pi, \pi])$ denote the functions defined on $[-\pi.\pi]$ and twice differentiable there, with continuous second derivatives up to the boundary. We denote by $\mathcal{C}([-\pi, \pi])$ the space of functions defined on $[-\pi, \pi]$ which are continuous up to the boundary. We can regard $\mathcal{C}(\mathbf{T})$ as the subspace of $\mathcal{C}([-\pi, \pi])$ consisting of those functions which satisfy the boundary conditions $f(\pi) = f(-\pi)$ (and then extended to the whole line by periodicity).

We regard $\mathcal{C}([-\pi, \pi])$ as a pre-Hilbert space with the same scalar product that we have been using:

$$(f, g) = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(x)\overline{g(x)}dx.$$

If we can show that every element of $\mathcal{C}([-\pi, \pi])$ is a sum of its Fourier series (in the pre-Hilbert space sense) then the same will be true for $\mathcal{C}(\mathbf{T})$. So we will work with $\mathcal{C}([-\pi, \pi])$.

We can consider the operator $D^2 - 1$ as a linear map

$$D^2 - 1 : \mathcal{C}^2([-\pi, \pi]) \to \mathcal{C}([-\pi, \pi]).$$

This map is surjective, meaning that given any continuous function g we can find a twice differentiable function f satisfying the differential equation

$$f'' - f = g.$$

In fact we can find a whole two dimensional family of solutions because we can add any solution of the homogeneous equation

$$h'' - h = 0$$

to f and still obtain a solution. We could write down an explicit solution for the equation $f'' - f = g$, but we will not need to. It is enough for us to know that the solution exists, which follows from the general theory of ordinary differential equations.

The general solution of the homogeneous equation is given by

$$h(x) = ae^x + be^{-x}.$$

Let

$$M \subset \mathcal{C}^2([-\pi, \pi])$$

be the subspace consisting of those functions which satisfy the "periodic boundary conditions"

$$f(\pi) = f(-\pi), \quad f'(\pi) = f'(-\pi).$$

Given any f we can always find a solution of the homogeneous equation such that $f - h \in M$. Indeed, we need to choose the complex numbers a and b such that if h is as given above, then

$$h(\pi) - h(-\pi) = f(\pi) - f(-\pi), \text{ and } h'(\pi) - h'(-\pi) = f'(\pi) - f'(-\pi).$$

Collecting coefficients and denoting the right hand side of these equations by c and d we get the linear equations

$$(e^\pi - e^{-\pi})(a - b) = c, (e^\pi - e^{-\pi})(a + b) = d$$

which has a unique solution.

So there exists a unique operator

$$T : \mathcal{C}([-\pi, \pi]) \to M$$

with the property that

$$(D^2 - I) \circ T = I.$$

We will prove that

$$T \text{ is self adjoint and compact.} \tag{2.20}$$

Once we will have proved this fact, then we know every element of M can be expanded in terms of a series consisting of eigenvectors of T with non-zero eigenvalues. But if

$$Tw = \lambda w$$

then

$$D^2 w = (D^2 - I)w + w = \frac{1}{\lambda}[(D^2 - I) \circ T]w + w = \left(\frac{1}{\lambda} + 1\right)w.$$

So w must be an eigenvector of D^2; it must satisfy

$$w'' = \mu w.$$

So if $\mu = 0$ then $w = $ a constant is a solution. If $\mu = r^2 > 0$ then w is a linear combination of e^{rx} and e^{-rx} and as we showed above, no non-zero such combination can belong to M. If $\mu = -r^2$ then the solution is a linear combination of e^{irx} and e^{-irx} and the above argument shows that r must be such that $e^{ir\pi} = e^{-ir\pi}$ so $r = n$ is an integer.

Thus (2.20) will show that the e^{inx} are a basis of M, and a little more work that we will do at the end will show that they are in fact also a basis of $\mathcal{C}([-\pi, \pi])$. But first let us work on (2.20).

It is easy to see that T is self adjoint. Indeed, let $f = Tu$ and $g = Tv$ so that f and g are in M and

$$(u, Tv) = ([D^2 - 1]f, g) = -(f', g') - (f, g) = (f, [D^2 - 1]g) = (Tu, v)$$

where we have used integration by parts and the boundary conditions defining M for the two middle equalities.

2.4.3   Gårding's inequality, special case.

We now turn to the compactness. We have already verified that for any $f \in M$ we have

$$([D^2 - 1]f, f) = -(f', f') - (f, f).$$

Taking absolute values we get

$$\|f'\|^2 + \|f\|^2 \le |([D^2 - 1]f, f)|. \tag{2.21}$$

(We actually get equality here, the more general version of this that we will develop later will be an inequality.)

Let $u = [D^2 - 1]f$ and use the Cauchy-Schwartz inequality

$$|([D^2 - 1]f, f)| = |(u, f)| \le \|u\|\|f\|$$

on the right hand side of (2.21) to conclude that

$$\|f\|^2 \le \|u\| \|f\|$$

or

$$\|f\| \le \|u\|.$$

Use (2.21) again to conclude that

$$\|f'\|^2 \le \|u\| \|f\| \le \|u\|^2$$

by the preceding inequality. We have $f = Tu$, and let us now suppose that u lies on the unit sphere i.e. that $\|u\| = 1$. Then we have proved that

$$\|f\| \le 1, \quad \text{and} \quad \|f'\| \le 1. \tag{2.22}$$

We wish to show that from any sequence of functions satisfying these two conditions we can extract a subsequence which converges. Here convergence means, of course, with respect to the norm given by

$$\|f\|^2 = \frac{1}{2\pi} \int_{-\pi}^{\pi} |f(x)|^2 dx.$$

In fact, we will prove something stronger: that given any sequence of functions satisfying (2.22) we can find a subsequence which converges in the uniform norm

$$\|f\|_\infty := \max_{x \in [-\pi, \pi]} |f(x)|.$$

Notice that

$$\|f\| = \left(\frac{1}{2\pi} \int_{-\pi}^{\pi} |f(x)|^2 dx \right)^{\frac{1}{2}} \le \left(\frac{1}{2\pi} \int_{-\pi}^{\pi} (\|f\|_\infty)^2 dx \right)^{\frac{1}{2}} = \|f\|_\infty$$

so convergence in the uniform norm implies convergence in the norm we have been using.

To prove our result, notice that for any $\pi \le a < b \le \pi$ we have

$$|f(b) - f(a)| = \left| \int_a^b f'(x)dx \right| \le \int_a^b |f'(x)|dx = 2\pi(|f'|, \mathbf{1}_{[a,b]})$$

where $\mathbf{1}_{[a,b]}$ is the function which is one on $[a,b]$ and zero elsewhere. Apply Cauchy-Schwartz to conclude that

$$|(|f'|, \mathbf{1}_{[a,b]})| \le \| \, |f'| \, \| \cdot \|\mathbf{1}_{[a,b]}\|.$$

But

$$\|\mathbf{1}_{[a,b]}\|^2 = \frac{1}{2\pi} |b - a|$$

and

$$\| \, |f'| \, \| = \|f'\| \le 1.$$

We conclude that

$$|f(b) - f(a)| \le (2\pi)^{\frac{1}{2}}|b - a|^{\frac{1}{2}}. \tag{2.23}$$

In this inequality, let us take b to be a point where $|f|$ takes on its maximum value, so that $|f(b)| = \|f\|_\infty$. Let a be a point where $|f|$ takes on its minimum value. (If necessary interchange the role of a and b to arrange that $a < b$ or observe that the condition $a < b$ was not needed in the above proof.) Then (2.23) implies that

$$\|f\|_\infty - \min|f| \le (2\pi)^{\frac{1}{2}}|b - a|^{\frac{1}{2}}.$$

But

$$1 \ge \|f\| = \left(\frac{1}{2\pi}\int_{-\pi}^{\pi}|f|^2(x)dx\right)^{\frac{1}{2}} \ge \left(\frac{1}{2\pi}\int_{-\pi}^{\pi}(\min|f|)^2 dx\right)^{\frac{1}{2}} = \min|f|$$

and $|b - a| \le 2\pi$ so

$$\|f\|_\infty \le 1 + 2\pi.$$

Thus the values of all the $f \in T[S]$ are all uniformly bounded - (they take values in a circle of radius $1 + 2\pi$) and they are equicontinuous in that (2.23) holds. This is enough to guarantee that out of every sequence of such f we can choose a uniformly convergent subsequence.

(We recall how the proof of this goes: Since all the values of all the f are bounded, at any point we can choose a subsequence so that the values of the f at that point converge, and, by passing to a succession of subsequences (and passing to a diagonal), we can arrange that this holds at any countable set of points. In particular, we may choose say the rational points in $[-\pi, \pi]$. Suppose that f_n is this subsequence. We claim that (2.23) then implies that the f_n form a Cauchy sequence in the uniform norm and hence converge in the uniform norm to some continuous function. Indeed, for any ϵ choose δ such that

$$(2\pi)^{\frac{1}{2}}\delta^{\frac{1}{2}} < \frac{1}{3}\epsilon,$$

choose a finite number of rational points which are within δ distance of any point of $[-\pi, \pi]$ and choose N sufficiently large that $|f_i - f_j| < \frac{1}{3}\epsilon$ at each of these points, r. when i and j are $\ge N$. Then at any $x \in [-\pi, \pi]$

$$|f_i(x) - f_j(x)| \le |f_i(x) - f_i(r)| + |f_j(x) - f_j(r)| + |f_i(r) - f_j(r)| \le \epsilon$$

since we can choose r such that that the first two and hence all of the three terms is $\le \frac{1}{3}\epsilon$.)

2.5   The Heisenberg uncertainty principle.

In this section we show how the arguments leading to the Cauchy-Schwartz inequality give one of the most important discoveries of twentieth century physics, the Heisenberg uncertainty principle.

Let V be a pre-Hilbert space, and S denote the unit sphere in V. If ϕ and ψ are two unit vectors (i.e. elements of S) their scalar product (ϕ, ψ) is a complex number with $0 \leq |(\phi, \psi)|^2 \leq 1$. In quantum mechanics, this number is taken as a probability. Although in the "real world" V is usually infinite dimensional, we will warm up by considering the case where V is finite dimensional.

Given a $\phi \in S$ and an orthonormal basis $\phi_1, \ldots, \phi_n$ of V, we have

$$1 = \|\phi\|^2 = |(\phi, \phi_1)|^2 + \cdots + |(\phi, \phi_n)|^2.$$

The says that the various probabilities $|(\phi, \phi_i)|^2$ add up to one. We recall some language from elementary probability theory: Suppose we have an experiment resulting in a finite number of measured numerical outcomes λ_i, each with probability p_i of occurring. Then the mean $\langle \lambda \rangle$ is defined by

$$\langle \lambda \rangle := \lambda_1 p_1 + \cdots + \lambda_n p_n$$

and its variance $(\Delta \lambda)^2$

$$(\Delta \lambda)^2 := (\lambda_1 - \langle \lambda \rangle)^2 p_1 + \cdots + (\lambda_n - \langle \lambda \rangle)^2 p_n$$

and an immediate computation shows that

$$(\Delta \lambda)^2 = \langle \lambda^2 \rangle - \langle \lambda \rangle^2.$$

The square root $\Delta \lambda$ of the variance is called the standard deviation. The variance (or the standard deviation) measures the "spread" of the possible values of λ. To understand its meaning we have Chebychev's inequality which estimates the probability that λ_k can deviate from $\langle \lambda \rangle$ by as much as $r \Delta \lambda$ for any positive number r. Chebychev's inequality says that this probability is $\leq 1/r^2$. In symbols

$$\text{Prob } (|\lambda_k - \langle \lambda \rangle| \geq r \Delta \lambda) \leq \frac{1}{r^2}.$$

Indeed, the probability on the left is the sum of all the p_k such that $|\lambda_k - \langle \lambda \rangle| \geq r\Delta$. Denoting this sum by $\sum_r$ we have

$$\sum_r p_k \leq \sum_r p_k \frac{(\lambda - \langle \lambda \rangle)^2}{r^2 (\Delta \lambda)^2} \leq$$

$$\leq \sum_{\text{all } k} p_k \frac{(\lambda - \langle \lambda \rangle)^2}{r^2 (\Delta \lambda)^2} = \frac{1}{r^2 (\Delta \lambda)^2} \sum_{\text{all } k} (\lambda - \langle \lambda \rangle)^2 p_k = \frac{1}{r^2}.$$

Replacing λ_i by $\lambda_i + c$ does not change the variance.

Now suppose that A is a self-adjoint operator on V, that the λ_i are the eigenvalues of A with eigenvectors ϕ_i constituting an orthonormal basis, and that the $p_i = |(\phi, \phi_i)|^2$ as above.

1. Show that $\langle \lambda \rangle = (A\phi, \phi)$ and that $(\Delta \lambda)^2 = (A^2 \phi, \phi) - (A\phi, \phi)^2$.

We will write the expression $(A\phi, \phi)$ as $\langle A \rangle_\phi$, In quantum mechanics a unit vector is called a **state** and a self-adjoint operator is called an **observable** and the expression $\langle A \rangle_\phi$ is called the **expectation** of the observable A in the state ϕ. Similarly we denote $\left((A^2\phi, \phi) - (A\phi, \phi)^2\right)^{1/2}$ by $\Delta_\phi A$. It is called the **uncertainty** of the observable A in the state ϕ. Notice that

$$(\Delta_\phi A)^2 = \langle (A - \langle A \rangle I)^2 \rangle_\phi$$

where I denotes the identity operator. Indeed

$$((A - (A\phi, \phi)I)^2 = A^2 - 2(A\phi, \phi)A + (A\phi, \phi)^2 I$$

so

$$\langle (A - \langle A \rangle_\phi)^2 \rangle_\phi = (A^2\phi, \phi) - 2\langle A \rangle_\phi^2 + \langle A \rangle_\phi^2 = \langle A^2 \rangle_\phi - \langle A \rangle_\phi^2.$$

When the state ϕ is fixed in the course of discussion, we will drop the subscript ϕ and write $\langle A \rangle$ and ΔA instead of $\langle A \rangle_\phi$ and $\Delta_\phi A$. For example, we would write the previous result as

$$\Delta A = \langle (A - \langle A \rangle I)^2 \rangle.$$

If A and B are operators we let $[A, B]$ denote the commutator:

$$[A, B] := AB - BA.$$

Notice that $[A, B] = -[B, A]$ and $[I, B] = 0$ for any B. So if A and B are self adjoint, so is $i[A, B]$ and replacing A by $A - \langle A \rangle I$ and B by $B - \langle B \rangle I$ does not change ΔA, ΔB or $i[A, B]$.

The **uncertainty principle** says that for any two observables A and B we have

$$(\Delta A)(\Delta B) \geq \frac{1}{2}|\langle i[A, B] \rangle|q.$$

Proof. Set $A_1 := A - \langle A \rangle$, $\ B_1 := B - \langle B \rangle$ so that

$$[A_1, B_1] = [A, B].$$

Let

$$\psi := A_1\phi + ixB_1\phi.$$

Then

$$(\psi, \psi) = (\Delta A)^2 - x\langle i[A, B] \rangle + (\Delta B)^2.$$

Since $(\psi, \psi) \geq 0$ for all x this implies that $(b^2 \leq 4ac)$ that

$$\langle i[A, B] \rangle^2 \leq 4(\Delta A)^2(\Delta B)^2,$$

and taking square roots gives the result.

The purpose of the next few sections is to provide a vast generalization of the results we obtained for the operator D^2. We will prove the corresponding results for any "elliptic" differential operator (definitions below).

I plan to study differential operators acting on vector bundles over manifolds. But it requires some effort to set things up, and I want to get to the key analytic ideas which are essentially repeated applications of integration by parts. So I will start with elliptic operators L acting on functions on the torus $\mathbf{T} = \mathbf{T}^n$, where there are no boundary terms when we integrate by parts. Then an immediate extension gives the result for elliptic operators on functions on manifolds, and also for boundary value problems such as the Dirichlet problem.

The treatment here rather slavishly follows the treatment by Bers and Schechter in *Partial Differential Equations* by Bers, John and Schechter AMS (1964).

2.6 The Sobolev Spaces.

Recall that $\mathbf{T}$ now stands for the n-dimensional torus. Let $\mathbf{P} = \mathbf{P}(\mathbf{T})$ denote the space of trigonometric polynomials. These are functions on the torus of the form
$$u(x) = \sum a_\ell e^{i\ell \cdot x}$$
where
$$\ell = (\ell_1, \ldots, \ell_n)$$
is an n-tuplet of integers and the sum is finite. For each integer t (positive, zero or negative) we introduce the scalar product
$$(u, v)_t := \sum_\ell (1 + \ell \cdot \ell)^t a_\ell \overline{b}_\ell. \tag{2.24}$$

For $t = 0$ this is the scalar product
$$(u, v)_0 = \frac{1}{(2\pi)^n} \int_{\mathbf{T}} u(x)\overline{v(x)}dx.$$

This differs by a factor of $(2\pi)^{-n}$ from the scalar product that is used by Bers and Schecter. We will denote the norm corresponding to the scalar product $(\ ,\ )_s$ by $\|\ \|_s$.

If
$$\Delta := -\left(\frac{\partial^2}{\partial(x^1)^2} + \cdots + \frac{\partial^2}{\partial(x^n)^2}\right)$$
the operator $(1 + \Delta)$ satisfies
$$(1 + \Delta)u = \sum (1 + \ell \cdot \ell)a_\ell e^{i\ell \cdot x}$$
and so
$$((1 + \Delta)^t u, v)_s = (u, (1 + \Delta)^t v)_s = (u, v)_{s+t}$$
and
$$\|(1 + \Delta)^t u\|_s = \|u\|_{s+2t}. \tag{2.25}$$

We then get the "generalized Cauchy-Schwartz inequality"

$$|(u,v)_s| \leq \|u\|_{s+t}\|v\|_{s-t} \tag{2.26}$$

for any t, as a consequence of the usual Cauchy-Schwartz inequality. Indeed,

$$
\begin{aligned}
\sum_\ell (1+\ell\cdot\ell)^s a_\ell \bar{b}_\ell &= \sum_\ell (1+\ell\cdot\ell)^{\frac{s+t}{2}} a_\ell (1+\ell\cdot\ell)^{\frac{s-t}{2}} \bar{b}_\ell \\
&= ((1+\Delta)^{\frac{s+t}{2}} u, (1+\Delta)^{\frac{s-t}{2}} v)_0 \\
&\leq \|(1+\Delta)^{\frac{s+t}{2}} u\|_0 \|(1+\Delta)^{\frac{s-t}{2}} v\|_0 \\
&= \|u\|_{s+t}\|v\|_{s-t}.
\end{aligned}
$$

The generalized Cauchy-Schwartz inequality reduces to the usual Cauchy-Schwartz inequality when $t = 0$.

Clearly we have

$$\|u\|_s \leq \|u\|_t \quad \text{if } s \leq t.$$

If D^p denotes a partial derivative,

$$D^p = \frac{\partial^{|p|}}{\partial(x^1)^{p_1}\cdots\partial(x^n)^{p_m}}$$

then

$$D^p u = \sum (i\ell)^p a_\ell e^{i\ell\cdot x}.$$

In these equations we are using the following notations:

- If $p = (p_1,\ldots,p_n)$ is a vector with non-negative integer entries we set

$$|p| := p_1 + \cdots + p_n.$$

- If $\xi = (\xi_1,\ldots,\xi_n)$ is a (row) vector we set

$$\xi^p := \xi_1^{p_1}\cdot\xi_2^{p_2}\cdots\xi_n^{p_n}$$

It is then clear that

$$\|D^p u\|_t \leq \|u\|_{t+|p|} \tag{2.27}$$

and similarly

$$\|u\|_t \leq (\text{constant depending on } t) \sum_{|p|\leq t} \|D^p u\|_0 \quad \text{if } t \geq 0. \tag{2.28}$$

In particular,

Proposition 2.6.1 *The norms*

$$u \mapsto \|u\|_t$$

$t \geq 0$ and

$$u \mapsto \sum_{|p| \leq t} \|D^p u\|_0$$

are equivalent.

We let $\mathbf{H}_t$ denote the completion of the space $\mathbf{P}$ with respect to the norm $\| \ \|_t$. Each $\mathbf{H}_t$ is a Hilbert space, and we have natural embeddings

$$\mathbf{H}_t \hookrightarrow \mathbf{H}_s \ \text{ if } s < t.$$

Equation (2.25) says that

$$(1 + \Delta)^t : \mathbf{H}_{s+2t} \to \mathbf{H}_s$$

and is an isometry.

From the generalized Schwartz inequality we also have a natural pairing of $\mathbf{H}_t$ with $\mathbf{H}_{-t}$ given by the extension of $(\ , \ )_0$, so

$$|(u, v)_0| \leq \|u\|_t \|v\|_{-t}. \tag{2.29}$$

In fact, this pairing allows us to identify $\mathbf{H}_{-t}$ with the space of continuous linear functions on $\mathbf{H}_t$. Indeed, if ϕ is a continuous linear function on $\mathbf{H}_t$ the Riesz representation theorem tells us that there is a $w \in \mathbf{H}_t$ such that $\phi(u) = (u, w)_t$. Set

$$v := (1 + \Delta)^t w.$$

Then

$$v \in \mathbf{H}_{-t}$$

and

$$(u, v)_0 = (u, (1 + \Delta)^t w)_0 = (u, w)_t = \phi(u).$$

We record this fact as

$$\mathbf{H}_{-t} = (\mathbf{H}_t)^* . \tag{2.30}$$

As an illustration of (2.30), observe that the series

$$\sum_\ell (1 + \ell \cdot \ell)^s$$

converges for

$$s < -\frac{n}{2}.$$

This means that if define v by taking

$$b_\ell \equiv 1$$

then $v \in \mathbf{H}_s$ for $s < -\frac{n}{2}$. If u is given by $u(x) = \sum_\ell a_\ell e^{i\ell \cdot x}$ is any trigonometric polynomial, then

$$(u, v)_0 = \sum a_\ell = u(0).$$

So the natural pairing (2.29) allows us to extend the linear function sending $u \mapsto u(0)$ to all of $\mathbf{H}_t$ if $t > \frac{n}{2}$. We can now give v its "true name": it is the Dirac "delta function" δ (on the torus) where

$$(u, \delta)_0 = u(0).$$

So $\delta \in H_{-t}$ for $t > \frac{n}{2}$, and the preceding equation is usually written symbolically as

$$\frac{1}{(2\pi)^n} \int_{\mathbf{T}} u(x)\delta(x)dx = u(0);$$

but the true mathematical interpretation is as given above.

We set

$$\mathbf{H}_\infty := \bigcap \mathbf{H}_t, \quad \mathbf{H}_{-\infty} := \bigcup \mathbf{H}_t.$$

The space $\mathbf{H}_0$ is just $L_2(\mathbf{T})$, and we can think of the space $\mathbf{H}_t$, $t > 0$ as consisting of those functions having "generalized L_2 derivatives up to order t". Certainly a function of class C^t belongs to $\mathbf{H}_t$. With a loss of degree of differentiability the converse is true:

Lemma 2.6.1 [Sobolev.]   *If $u \in \mathbf{H}_t$ and*

$$t \geq \left[\frac{n}{2}\right] + k + 1$$

then $u \in C^k(\mathbf{T})$ and

$$\sup_{x \in \mathbf{T}} |D^p u(x)| \leq \text{const.} \|u\|_t \quad \textit{for } |p| \leq k. \tag{2.31}$$

By applying the lemma to $D^p u$ it is enough to prove the lemma for $k = 0$. So we assume that $u \in \mathbf{H}_t$ with $t \geq [n/2] + 1$. Then

$$\left(\sum |a_\ell|\right)^2 \leq \left(\sum (1 + \ell \cdot \ell)^t |a_\ell|^2\right) \sum (1 + \ell \cdot \ell)^{-t} < \infty,$$

since the series $\sum (1 + \ell \cdot \ell)^{-t}$ converges for $t \geq [n/2] + 1$. So for this range of t, the Fourier series for u converges absolutely and uniformly. The right hand side of the above inequality gives the desired bound. QED

A **distribution** on $\mathbf{T}^n$ is a linear function T on $C^\infty(\mathbf{T}^n)$ with the continuity condition that

$$\langle T, \phi_k \rangle \to 0$$

whenever

$$D^p \phi_k \to 0$$

uniformly for each fixed p. If $u \in \mathbf{H}_{-t}$ we may define

$$\langle u, \phi \rangle := (\phi, \overline{u})_0$$

and since $C^\infty(\mathbf{T})$ is dense in $\mathbf{H}_t$ we may conclude

Lemma 2.6.2 $\mathbf{H}_{-t}$ *is the space of those distributions T which are continuous in the $\| \ \|_t$ norm, i.e. which satisfy*

$$\|\phi_k\|_t \to 0 \quad \Rightarrow \quad \langle T, \phi_k \rangle \to 0.$$

We then obtain

Theorem 2.6.1 [Laurent Schwartz.] $\mathbf{H}_{-\infty}$ *is the space of all distributions. In other words, any distribution belongs to $\mathbf{H}_{-t}$ for some t.*

Proof. Suppose that T is a distribution that does not belong to any $\mathbf{H}_{-t}$. This means that for any $k > 0$ we can find a C^∞ function ϕ_k with

$$\|\phi_k\|_k < \frac{1}{k}$$

and

$$|\langle T, \phi_k \rangle| \geq 1.$$

But by Lemma 2.6.1 we know that $\|\phi_k\|_k < \frac{1}{k}$ implies that $D^p \phi_k \to 0$ uniformly for any fixed p contradicting the continuity property of T. QED

Suppose that ϕ is a C^∞ function on $\mathbf{T}$. Multiplication by ϕ is clearly a bounded operator on $\mathbf{H}_0 = L_2(\mathbf{T})$, and so it is also a bounded operator on $\mathbf{H}_t$, $t > 0$ since we can expand $D^p(\phi u)$ by applications of Leibnitz's rule.

For $t = -s < 0$ we know by the generalized Cauchy Schwartz inequality that

$$\|\phi u\|_t = \sup |(v, \phi u)_0| / \|v\|_s = \sup |(u, \overline{\phi} v)| / \|v\|_s \leq \|u\|_t \|\overline{\phi} v\|_s / \|v\|_s.$$

So in all cases we have

$$\|\phi u\|_t \leq (\text{const. depending on } \phi \text{ and } t)\|u\|_t. \tag{2.32}$$

Let

$$L = \sum_{|p| \leq m} \alpha_p(x) D^p$$

be a differential operator of degree m with C^∞ coefficients. Then it follows from the above that

$$\|Lu\|_{t-m} \leq \text{constant}\|u\|_t \tag{2.33}$$

where the constant depends on L and t.

Lemma 2.6.3 [Rellich's lemma.] *If $s < t$ the embedding $\mathbf{H_t} \hookrightarrow \mathbf{H}_s$ is compact.*

Proof.   We must show that the image of the unit ball B of $\mathbf{H}_t$ in $\mathbf{H}_s$ can be covered by finitely many balls of radius ϵ. Choose N so large that

$$(1 + \ell \cdot \ell)^{(s-t)/2} < \frac{\epsilon}{2}$$

when $\ell \cdot \ell > N$. Let Z_t be the subspace of $\mathbf{H}_t$ consisting of all u such that $a_\ell = 0$ when $\ell \cdot \ell \le N$. This is a space of finite codimension, and hence the unit ball of $Z_t^\perp \subset \mathbf{H}_t$ can be covered by finitely many balls of radius $\frac{\epsilon}{2}$. The space $Z_t^\perp$ consists of all u such that $a_\ell = 0$ when $\ell \cdot \ell > N$. The image of $Z_t^\perp$ in $\mathbf{H_s}$ is the orthogonal complement of the image of Z_t. On the other hand, for $u \in B \cap Z$ we have

$$\|u\|_s^2 \le (1 + N)^{s-t}\|u\|_t^2 \le \left(\frac{\epsilon}{2}\right)^2.$$

So the image of $B \cap Z$ is contained in a ball of radius $\frac{\epsilon}{2}$. Every element of the image of B can be written as a(n orthogonal) sum of an element in the image of $B \cap Z_t^\perp$ and an element of $B \cap Z_t$ and so the image of B is covered by finitely many balls of radius ϵ. QED

2.7   Gårding's inequality.

Let x, a, and b be positive numbers. Then

$$x^a + x^{-b} \ge 1$$

because if $x \ge 1$ the first summand is ≥ 1 and if $x \le 1$ the second summand is ≥ 1. Setting $x = \epsilon^{1/a}A$ gives

$$1 \le \epsilon A^a + \epsilon^{-b/a} A^{-b}$$

if ϵ and A are positive. Suppose that $t_1 > s > t_2$ and we set $a = t_1 - s$, $b = s - t_2$ and $A = 1 + \ell \cdot \ell$. Then we get

$$(1 + \ell \cdot \ell)^s \le \epsilon(1 + \ell \cdot \ell)^{t_1} + \epsilon^{-(s-t_2)/(t_1-s)}(1 + \ell \cdot \ell)^{t_2}$$

and therefore

$$\|u\|_s \le \epsilon\|u\|_{t_1} + \epsilon^{-(s-t_2)/(t_1-s)}\|u\|_{t_2} \quad \text{if } t_1 > s > t_2, \quad \epsilon > 0 \qquad (2.34)$$

for all $u \in \mathbf{H}_{t_1}$. This elementary inequality will be the key to several arguments in this section where we will combine (2.34) with integration by parts.

A differential operator $L = \sum_{|p| \le m} \alpha_p(x)D^p$ with real coefficients and m even is called **elliptic** if there is a constant $c > 0$ such that

$$(-1)^{m/2} \sum_{|p|=m} a_p(x)\xi^p \ge c(\xi \cdot \xi)^{m/2}. \qquad (2.35)$$

In this inequality, the vector ξ is a "dummy variable". (Its true invariant significance is that it is a covector, i.e. an element of the cotangent space at x.) The

expression on the left of this inequality is called the **symbol** of the operator L. It is a homogeneous polynomial of degree m in the variable ξ whose coefficients are functions of x. The symbol of L is sometimes written as $\sigma(L)$ or $\sigma(L)(x, \xi)$. Another way of expressing condition (2.35) is to say that there is a positive constant c such that

$$\sigma(L)(x, \xi) \geq c \text{ for all } x \text{ and } \xi \text{ such that } \xi \cdot \xi = 1.$$

We will assume until further notice that the operator L is elliptic and that m is a positive even integer.

Theorem 2.7.1 [Gårding's inequality.] *For every $u \in C^\infty(\mathbf{T})$ we have*

$$(u, Lu)_0 \geq c_1 \|u\|_{m/2}^2 - c_2 \|u\|_0^2 \tag{2.36}$$

where c_1 and c_2 are constants depending on L.

Remark. If $u \in \mathbf{H}_{m/2}$, then both sides of the inequality make sense, and we can approximate u in the $\| \ \|_{m/2}$ norm by C^∞ functions. So once we prove the theorem, we conclude that it is also true for all elements of $\mathbf{H}_{m/2}$.

We will prove the theorem in stages:

1. When L is constant coefficient and homogeneous.

2. When L is homogeneous and approximately constant.

3. When the L can have lower order terms but the homogeneous part of L is approximately constant.

4. The general case.

Stage 1. $L = \sum_{|p|=m} \alpha_p D^p$ where the α_p are constants. Then

$$
\begin{aligned}
(u, Lu)_0 &= \left(\sum a_\ell e^{i\ell \cdot x}, \sum_\ell \left(\sum_{|p|=m} \alpha_p (i\ell)^p \right) a_\ell e^{i\ell \cdot x} \right)_0 \\
&\geq c \sum_\ell (\ell \cdot \ell)^{m/2} |a_\ell|^2 \quad \text{by (2.35)} \\
&= c \sum [1 + (\ell \cdot \ell)^{m/2}] |a_\ell|^2 - c\|u\|_0^2 \\
&\geq cC\|u\|_{m/2}^2 - c\|u\|_0
\end{aligned}
$$

where

$$C = \sup_{r \geq 0} \frac{1 + r^{m/2}}{(1+r)^{m/2}}.$$

This takes care of stage 1.

Stage 2. $L = L_0 + L_1$ where L_0 is as in stage 1 and $L_1 = \sum_{|p|=m} \beta_p(x) D^p$ and

$$\max_{p,x} |\beta_p(x)| < \eta,$$

where η sufficiently small. (How small will be determined very soon in the course of the discussion.) We have

$$(u, L_0 u)_0 \geq c' \|u\|^2_{m/2} - c \|u\|^2_0$$

from stage 1.

We integrate $(u, L_1 u)_0$ by parts $m/2$ times. There are no boundary terms since we are on the torus. In integrating by parts some of the derivatives will hit the coefficients. Let us collect all the these terms as I_2. The other terms we collect as I_1, so

$$I_1 = \sum \int b_{p'+p''} D^{p'} u \overline{D^{p''} u} dx$$

where $|p'| = |p''| = m/2$ and $b_r = \pm\beta_r$. We can estimate this sum by

$$|I_1| \leq \eta \cdot \text{const.} \|u\|^2_{m/2}$$

and so will require that $\eta \cdot (\text{const.}) < c'$.

The remaining terms give a sum of the form

$$I_2 = \sum \int b_{p'q} D^{p'} u \overline{D^q u} dx$$

where $p' \leq m/2, q' < m/2$ so we have

$$|I_2| \leq \text{const.} \|u\|_{\frac{m}{2}} \|u\|_{\frac{m}{2}-1}.$$

Now let us take

$$s = \frac{m}{2} - 1, \ t_1 = \frac{m}{2}, \ t_2 = 0$$

in (2.34) which yields, for any $\epsilon > 0$,

$$\|u\|_{\frac{m}{2}-1} \leq \epsilon \|u\|_{\frac{m}{2}} + \epsilon^{-m/2} \|u\|_0.$$

Substituting this into the above estimate for I_2 gives

$$|I_2| \leq \epsilon \cdot \text{const.} \|u\|^2_{m/2} + \epsilon^{-m/2} \text{const.} \|u\|_{m/2} \|u\|_0.$$

For any positive numbers a, b and ζ the inequality $(\zeta a - \zeta^{-1} b)^2 \geq 0$ implies that $2ab \leq \zeta^2 a^2 + \zeta^{-2} b^2$. Taking $\zeta^2 = \epsilon^{\frac{m}{2}+1}$ we can replace the second term on the right in the preceding estimate for $|I_2|$ by

$$\epsilon^{-m-1} \cdot \text{const.} \|u\|^2_0$$

at the cost of enlarging the constant in front of $\|u\|^2_{\frac{m}{2}}$. We have thus established that

$$|I_1| \leq \eta \cdot (\text{const.})_1 \|u\|^2_{m/2}$$

where the constant depends only on m, and

$$|I_2| \leq \epsilon(\text{const.})_2 \|u\|_{m/2}^2 + \epsilon^{-m-1}\text{const.}\|u\|_0^2$$

where the constants depend on L_1 but ϵ is at our disposal. So if $\eta(\text{const.})_1 < c'$ and we then choose ϵ so that $\epsilon(\text{const.})_2 < c' - \eta \cdot (\text{const.})_1$ we obtain Gårding's inequality for this case.

Stage 3. $L = L_0 + L_1 + L_2$ where L_0 and L_1 are as in stage 2, and L_2 is a lower order operator. Here we integrate by parts and argue as in stage 2.

Stage 4, the general case. Choose an open covering of T such that the variation of each of the highest order coefficients in each open set is less than the η of stage 1. (Recall that this choice of η depended only on m and the c that entered into the definition of ellipticity.) Thus, if v is a smooth function supported in one of the sets of our cover, the action of L on v is the same as the action of an operator as in case 3) on v, and so we may apply Gårding's inequality. Choose a finite subcover and a partition of unity $\{\phi_i\}$ subordinate to this cover. Write $\phi_i = \psi_i^2$ (where we choose the ϕ so that the ψ are smooth). So $\sum \psi_i^2 \equiv 1$. Now

$$(\psi_i u, L(\psi_i u))_0 \geq c''\|\psi_i u\|_{m/2}^2 - \text{const.}\|\psi_i u\|_0^2$$

where c'' is a positive constant depending only on c, η, and on the lower order terms in L. We have

$$(u, Lu)_0 = \int \left(\sum \psi_i^2 u\right)\overline{Lu}\,dx = \sum (\psi_i u, L\psi_i u)_0 + R$$

where R is an expression involving derivatives of the ψ_i and hence lower order derivatives of u. These can be estimated as in case 2) above, and so we get

$$(u, Lu)_0 \geq c'''\sum \|\psi_i u\|_{m/2}^2 - \text{const.}\|u\|_0^2 \tag{2.37}$$

since $\|\psi_i u\|_0 \leq \|u\|_0$. Now $\|u\|_{m/2}$ is equivalent, as a norm, to $\sum_{p \leq m/2} \|D^p u\|_0$ as we verified in the preceding section. Also

$$\sum \|D^p(\psi_i u)\|_0 = \sum \|\psi_i D^p u\|_0 + R'$$

where R' involves terms differentiating the ψ and so lower order derivatives of u. Hence

$$\sum \|\psi_i u\|_{m/2}^2 \geq \text{pos. const.}\|u\|_{m/2}^2 - \text{const.}\|u\|_0^2$$

by the integration by parts argument again. Hence by (2.37)

$$(u, Lu)_0 \geq c'''\sum \|\psi_i u\|_{m/2}^2 - \text{const.}\|u\|_0^2$$

$$\geq \text{pos. const.}\|u\|_{m/2}^2 - \text{const.}\|u\|_0^2$$

which is Gårding's inequality. QED

For the time being we will continue to study the case of the torus. But a look ahead is in order. In this last step of the argument, where we applied the partition of unity argument, we have really freed ourselves of the restriction of being on the torus. Once we make the appropriate definitions, we will then get Gårding's inequality for elliptic operators on manifolds. Furthermore, the consequences we are about to draw from Gårding's inequality will be equally valid in the more general setting.

2.8   Consequences of Gårding's inequality.

Proposition 2.8.1 *For every integer t there is a constant $c(t) = c(t, L)$ and a positive number $\Lambda = \Lambda(t, L)$ such that*

$$\|u\|_t \leq c(t)\|Lu + \lambda u\|_{t-m} \tag{2.38}$$

when

$$\lambda > \Lambda$$

for all smooth u, and hence for all $u \in \mathbf{H}_t$.

Proof.   Let s be some non-negative integer. We will first prove (2.38) for $t = s + \frac{m}{2}$. We have

$$\|u\|_t \|Lu + \lambda u\|_{t-m} = \|u\|_t \|Lu + \lambda u\|_{s-\frac{m}{2}}$$

$$= \|u\|_t \|(1 + \Delta)^s Lu + \lambda(1 + \Delta)^s u\|_{-s-\frac{m}{2}}$$

$$\geq (u, (1 + \Delta)^s Lu + \lambda(1 + \Delta)^s u)_0$$

by the generalized Cauchy - Schwartz inequality (2.26).

The operator $(1 + \Delta)^s L$ is elliptic of order $m + 2s$ so (2.25) and Gårding's inequality gives

$$(u, (1 + \Delta)^s Lu + \lambda(1 + \Delta)^s u)_0 \geq c_1 \|u\|_{s+\frac{m}{2}}^2 - c_2 \|u\|_0^2 + \lambda \|u\|_s^2.$$

Since $\|u\|_s \geq \|u\|_0$ we can combine the two previous inequalities to get

$$\|u\|_t \|Lu + \lambda u\|_{t-m} \geq c_1 \|u\|_t^2 + (\lambda - c_2)\|u\|_0^2.$$

If $\lambda > c_2$ we can drop the second term and divide by $\|u\|_t$ to obtain (2.38).

We now prove the proposition for the case $t = \frac{m}{2} - s$ by the same sort of argument: We have

$$\|u\|_t \|Lu + \lambda u\|_{-s-\frac{m}{2}} = \|(1 + \Delta)^{-s} u\|_{s+\frac{m}{2}} \|Lu + \lambda u\|_{-s-\frac{m}{2}}$$

$$\geq ((1 + \Delta)^{-s} u, L(1 + \Delta)^s (1 + \Delta)^{-s} u + \lambda u)_0.$$

Now use the fact that $L(1+\Delta)^s$ is elliptic and Gårding's inequality to continue the above inequalities as

$$\geq c_1 \|(1+\Delta)^{-s}u\|_{s+\frac{m}{2}}^2 - c_2 \|(1+\Delta)^{-s}u\|_0^2 + \lambda \|u\|_{-s}^2$$

$$= c_1 \|u\|_t^2 - c_2 \|u\|_{-2s}^2 + \lambda \|u\|_{-s}^2 \geq c_1 \|u\|_t^2$$

if $\lambda > c_2$. Again we may then divide by $\|u\|_t$ to get the result. QED

The operator $L + \lambda I$ is a bounded operator from $\mathbf{H}_t$ to $\mathbf{H}_{t-m}$ (for any t). Suppose we fix t and choose λ so large that (2.38) holds. Then (2.38) says that $(L+\lambda I)$ is invertible on its image, and bounded there with a bound independent of $\lambda > \Lambda$, and this image is a closed subspace of $\mathbf{H}_{t-m}$.

Let us show that this image is all of $\mathbf{H}_{t-m}$ for λ large enough. Suppose not, which means that there is some $w \in \mathbf{H}_{t-m}$ with

$$(w, Lu + \lambda u)_{t-m} = 0$$

for all $u \in \mathbf{H}_t$. We can write this last equation as

$$((1+\Delta)^{t-m}w, Lu + \lambda u)_0 = 0.$$

Integration by parts gives the adjoint differential operator L^* characterized by

$$(\phi, L\psi)_0 = (L^*\phi, \psi)_0$$

for all smooth functions ϕ and ψ, and by passing to the limit this holds for all elements of $\mathbf{H}_r$ for $r \geq m$. The operator L^* has the same leading term as L and hence is elliptic. So let us choose λ sufficiently large that (2.38) holds for L^* as well as for L. Now

$$0 = ((1+\Delta)^{t-m}w, Lu + \lambda u)_0 = (L^*(1+\Delta)^{t-m}w + \lambda(1+\Delta)^{t-m}w, u)_0$$

for all $u \in \mathbf{H}_t$ which is dense in $\mathbf{H}_0$ so

$$L^*(1+\Delta)^{t-m}w + \lambda(1+\Delta)^{t-m}w = 0$$

and hence (by (2.38)) $(1+\Delta)^{t-m}w = 0$ so $w = 0$. We have proved

Proposition 2.8.2 *For every t and for λ large enough (depending on t) the operator $L + \lambda I$ maps $\mathbf{H}_t$ bijectively onto $\mathbf{H}_{t-m}$ and $(L + \lambda I)^{-1}$ is bounded independently of λ.*

As an immediate application we get the important

Theorem 2.8.1 *If u is a distribution and $Lu \in \mathbf{H}_s$ then $u \in \mathbf{H}_{s+m}$.*

Proof. Write $f = Lu$. By Schwartz's theorem, we know that $u \in \mathbf{H}_k$ for some k. So $f + \lambda u \in \mathbf{H}_{\min(k,s)}$ for any λ. Choosing λ large enough, we conclude that $u = (L + \lambda I)^{-1}(f + \lambda u) \in \mathbf{H}_{\min(k+m,s+m)}$. If $k + m < s + m$ we can repeat

the argument to conclude that $u \in \mathbf{H}_{\min(k+2m,s+m)}$. we can keep going until we conclude that $u \in \mathbf{H}_{s+m}$. QED

Notice as an immediate corollary that any solution of the homogeneous equation $Lu = 0$ is C^∞.

We now obtain a second important consequence of Proposition 2.8.2. Choose λ so large that the operators

$$(L + \lambda I)^{-1} \quad \text{and} \quad (L^* + \lambda I)^{-1}$$

exist as operators from $\mathbf{H}_0 \to \mathbf{H}_m$. Follow these operators with the injection $\iota_m : \mathbf{H}_m \to \mathbf{H}_0$ and set

$$M := \iota_m \circ (L + \lambda I)^{-1}, \quad M^* := \iota_m \circ (L^* + \lambda I)^{-1}.$$

Since ι_m is compact (Rellich's lemma) and the composite of a compact operator with a bounded operator is compact, we conclude

Theorem 2.8.2 *The operators M and M^* are compact.*

Suppose that $L = L^*$. (This is usually expressed by saying that L is "formally self-adjoint". More on this terminology will come later.) This implies that $M = M^*$. In other words, M is a compact self adjoint operator, and we can apply Theorem 2.3.1 to conclude that eigenvectors of M form a basis of $R(M)$ and that the corresponding eigenvalues tend to zero. Prop 2.8.2 says that $R(M)$ is the same as $\iota_m(\mathbf{H}_m)$ which is dense in $\mathbf{H}_0 = L_2(\mathbf{T})$. We conclude that the eigenvectors of M form a basis of $L_2(\mathbf{T})$. If $Mu = ru$ then $u = (L + \lambda I)Mu = rLu + \lambda ru$ so u is an eigenvector of L with eigenvalue

$$\frac{1 - r\lambda}{r}.$$

We conclude that the eigenvectors of L are a basis of $\mathbf{H}_0$. We claim that only finitely many of these eigenvalues of L can be negative. Indeed, since we know that the eigenvalues r_n of M approach zero, the numerator in the above expression is positive, for large enough n, and hence if there were infinitely many negative eigenvalues μ_k, they would have to correspond to negative r_k and so these $\mu_k \to -\infty$. But taking $s_k = -\mu_k$ as the λ in (2.38) in Prop. 2.8.1 we conclude that $u = 0$, if $Lu = \mu_k u$ if k is large enough, contradicting the definition of an eigenvector. So all but a finite number of the r_n are positive, and these tend to zero. To summarize:

Theorem 2.8.3 *The eigenvectors of L are C^∞ functions which form a basis of $\mathbf{H}_0$. Only finitely many of the eigenvalues μ_k of L are negative and $\mu_n \to \infty$ as $n \to \infty$.*

It is easy to extend the results obtained above for the torus in two directions. One is to consider functions defined in a **domain** = bounded open set $\mathcal{G}$ of $\mathbf{R}^n$ and the other is to consider functions defined on a compact manifold. In both cases a few elementary tricks allow us to reduce to the torus case. We sketch what is involved for the manifold case.

2.9 Extension of the basic lemmas to manifolds.

Let $E \to M$ be a vector bundle over a manifold. We assume that M is equipped with a density which we shall denote by $|dx|$ and that E is equipped with a positive definite (smoothly varying) scalar product, so that we can define the L_2 norm of a smooth section s of E of compact support:

$$\|s\|_0^2 := \int_M |s|^2(x)|dx|.$$

Suppose for the rest of this section that M is compact. Let $\{U_i\}$ be a finite cover of M by coordinate neighborhoods over which E has a given trivialization, and ρ_i a partition of unity subordinate to this cover. Let ϕ_i be a diffeomorphism or U_i with an open subset of $\mathbf{T}^n$ where n is the dimension of M. Then if s is a smooth section of E, we can think of $(\rho_i s) \circ \phi_i^{-1}$ as an $\mathbf{R}^m$ or $\mathbf{C}^m$ valued function on $\mathbf{T}^n$, and consider the sum of the $\|\cdot\|_k$ norms applied to each component. We shall continue to denote this sum by $\|\rho_i f \circ \phi_i^{-1}\|_k$ and then define

$$\|f\|_k := \sum_i \|\rho_i f \circ \phi_i^{-1}\|_k$$

where the norms on the right are in the norms on the torus. These norms depend on the trivializations and on the partitions of unity. But any two norms are equivalent, and the $\| \ \|_0$ norm is equivalent to the "intrinsic" L_2 norm defined above. We define the Sobolev spaces $\mathbf{W}_k$ to be the completion of the space of smooth sections of E relative to the norm $\| \ \|_k$ for $k \geq 0$, and these spaces are well defined as topological vector spaces independently of the choices. Since Sobolev's lemma holds locally, it goes through unchanged. Similarly Rellich's lemma: if s_n is a sequence of elements of $\mathbf{W}_\ell$ which is bounded in the $\| \ \|_\ell$ norm for $\ell > k$, then each of the elements $\rho_i s_n \circ \phi_i^{-1}$ belong to $\mathbf{H}_\ell$ on the torus, and are bounded in the $\| \ \|_\ell$ norm, hence we can select a subsequence of $\rho_1 s_n \circ \phi_1^{-1}$ which converges in $\mathbf{H}_k$, then a subsubsequence such that $\rho_i s_n \circ \phi_i^{-1}$ for $i = 1, 2$ converge etc. arriving at a subsequence of s_n which converges in $\mathbf{W}_k$.

A differential operator L mapping sections of E into sections of E is an operator whose local expression (in terms of a trivialization and a coordinate chart) has the form

$$Ls = \sum_{|p| \leq m} \alpha_p(x) D^p s$$

Here the a_p are linear maps (or matrices if our trivializations are in terms of $\mathbf{R}^m$).

Under changes of coordinates and trivializations the change in the coefficients are rather complicated, but the **symbol** of the differential operator

$$\sigma(L)(\xi) := \sum_{|p|=m} a_p(x)\xi^p \quad \xi \in T^*M_x$$

is well defined.

If we put a Riemann metric on the manifold, we can talk about the length $|\xi|$ of any cotangent vector.

If L is a differential operator from E to itself (i.e. $F=E$) we shall call L **even elliptic** if m is even and there exists some constant C such that

$$\langle v, \sigma(L)(\xi)v \rangle \geq C|\xi|^m|v|^2$$

for all $x \in M$, $v \in E_x$, $\xi \in T^*M_x$ and $\langle \ , \ \rangle$ denotes the scalar product on E_x. Gårding's inequality holds. Indeed, locally, this is just a restatement of the (vector valued version) of Gårding's inequality that we have already proved for the torus. But Stage 4 in the proof extends unchanged (other than the replacement of scalar valued functions by vector valued functions) to the more general case.

2.10   Example: Hodge theory.

We assume knowledge of the basic facts about differentiable manifolds, in particular the existence of an operator $d : \Omega^k \to \Omega^{k+1}$ with its usual properties, where Ω^k denotes the space of exterior k-forms. Also, if M is orientable and carries a Riemann metric then the Riemann metric induces a scalar product on the exterior powers of T^*M and also picks out a volume form. So there is an induced scalar product $(\ , \ ) = (\ , \ )_k$ on Ω^k and a formal adjoint δ of d

$$\delta : \ \Omega^k \to \Omega^{k-1}$$

and satisfies

$$(d\psi, \phi) = (\phi, \delta\phi)$$

where ϕ is a $(k+1)$-form and ψ is a k-form. Then

$$\Delta := d\delta + \delta d$$

is a second order differential operator on Ω^k and satisfies

$$(\Delta\phi, \phi) = \|d\phi\|^2 + \|\delta\phi\|^2$$

where $\|\phi\|^2 = (\phi, \phi)$ is the intrinsic L_2 norm (so $\| \ \| = \| \ \|_0$ in terms of the notation of the preceding section). Furthermore, if

$$\phi = \sum_I \phi_I dx^I$$

is a local expression for the differential form ϕ, where

$$dx^I = dx_{i_1} \wedge \cdots \wedge dx_{i_k} \quad I = (i_1, \ldots, i_k)$$

then a local expression for Δ is

$$\Delta\phi = -\sum g^{ij} \frac{\partial \phi_I}{\partial x^i \partial x^j} + \cdots$$

where
$$g^{ij} = \langle dx^i, dx^j \rangle$$

and the $\cdots$ are lower order derivatives. In particular Δ is elliptic.

Let $\phi \in \Omega^k$ and suppose that
$$d\phi = 0.$$

Let $\mathcal{C}(\phi)$, the **cohomology class** of ϕ be the set of all $\psi \in \Omega^k$ which satisfy
$$\phi - \psi = d\alpha, \quad \alpha \in \Omega^{k-1}$$

and let
$$\overline{\mathcal{C}(\phi)}$$

denote the closure of $\mathcal{C}$ in the L_2 norm. It is a closed subspace of the Hilbert space obtained by completing Ω^k relative to its L_2 norm. Let us denote this space by L_2^k, so $\overline{\mathcal{C}(\phi)}$ is a closed subspace of L_2^k.

Proposition 2.10.1 *If $\phi \in \Omega^k$ and $d\phi = 0$, there exists a unique $\tau \in \overline{\mathcal{C}(\phi)}$ such that*
$$\|\tau\| \le \|\psi\| \quad \forall \, \psi \in \mathcal{C}(\phi).$$

Furthermore, τ is smooth,and
$$d\tau = 0 \quad and \quad \delta\tau = 0.$$

If choose a minimizing sequence for $\|\psi\|$ in $\mathcal{C}(\phi)$.

If we choose a minimizing sequence for $\|\psi\|$ in $\mathcal{C}(\phi)$ we know it is Cauchy, cf. the proof of the existence of orthogonal projections in a Hilbert space. So we know that τ exists and is unique. For any $\alpha \in \Omega^{k+1}$ we have
$$(\tau, \delta\alpha) = \lim(\psi, \delta\alpha) = \lim(d\psi, \alpha) = 0$$

as ψ ranges over a minimizing sequence. The equation $(\tau, \delta\alpha) = 0$ for all $\alpha \in \Omega^{k+1}$ says that τ is a weak solution of the equation $d\tau = 0$.

We claim that
$$(\tau, d\beta) = 0 \quad \forall \, \beta \in \Omega^{k-1}$$

which says that τ is a weak solution of $\delta\tau = 0$. Indeed, for any $t \in \mathbf{R}$,
$$\|\tau\|^2 \le \|\tau + td\beta\|^2 = \|\tau\|^2 + t^2\|d\beta\|^2 + 2t(\tau, d\beta)$$

so
$$-2t(\tau, d\beta) \le t^2\|d\beta\|^2.$$

If $(\tau, d\beta) \ne 0$, we can choose
$$t = -\epsilon \frac{(\tau, d\beta)}{|(\tau, d\beta)|}, \quad \epsilon > 0$$

so
$$|(\tau, d\beta)| \leq \epsilon |d\beta|^2.$$

As ϵ is arbitrary, this implies that $(\tau, d\beta) = 0$.

So $(\tau, \Delta\psi) = (\tau, [d\delta + \delta d]\psi) = 0$ for any $\psi \in \Omega^k$. Hence τ is a weak solution of $\Delta\tau = 0$ and so is smooth. The space $\mathcal{H}^k$ of weak, and hence smooth solutions of $\Delta\tau = 0$ is finite dimensional by the general theory. It is called the space of Harmonic forms. We have seen that there is a unique harmonic form in the cohomology class of any closed form, s the cohomology groups are finite dimensional. In fact, the general theory tells us that

$$L_2^k \bigoplus_\lambda E_\lambda^k$$

(Hilbert space direct sum) where E_λ^k is the eigenspace with eigenvalue λ of Δ. Each E_λ is finite dimensional and consists of smooth forms, and the $\lambda \to \infty$. The eigenspace E_0^k is just $\mathcal{H}^k$, the space of harmonic forms. Also, since

$$(\Delta\phi, \phi) = \|d\phi\|^2 + \|\delta\phi\|^2$$

we know that all the eigenvalues λ are non-negative.

Since $d\Delta = d(d\delta + \delta d) = d\delta d = \Delta d$, we see that

$$d : E_\lambda^k \to E_\lambda^{k+1}$$

and similarly

$$\delta : E_\lambda^k \to E_\lambda^{k-1}.$$

For $\lambda \neq 0$, if $\phi \in E_\lambda^k$ and $d\phi = 0$, then $\lambda\phi = \Delta\phi = d\delta\phi$ so $\phi = d(1/\lambda)\delta\phi$ so d restricted to the E_λ is exact, and similarly so is δ. Furthermore, on $\bigoplus_k E_\lambda^k$ we have

$$\lambda I = \Delta = (d + \delta)^2$$

so we have

$$E_\lambda^k = dE_\lambda^{k-1} \oplus \delta E_\lambda^{k+1}$$

and this decomposition is orthogonal since $(d\alpha, \delta\beta) = (d^2\alpha, \beta) = 0$.

As a first consequence we see that

$$L_2^k = \mathcal{H}^k \oplus \overline{d\Omega^{k-1}} \oplus \overline{\delta\Omega^{k-1}}$$

(the Hodge decomposition). If H denotes projection onto the first component, then Δ is invertible on the image of $I - H$ with a an inverse there which is compact. So if we let N denote this inverse on im $I - H$ and set $N = 0$ on $\mathcal{H}^k$ we get

$$\begin{aligned}
\Delta N &= I - H \\
Nd &= dN \\
\delta N &= N\delta \\
\Delta N &= N\Delta \\
NH &= 0
\end{aligned}$$

which are the fundamental assertions of Hodge theory, together with the assertion proved above that $H\phi$ is the unique minimizing element in its cohomology class.

We have seen that

$$d + \delta : \bigoplus_k E_\lambda^{2k} \to \bigoplus_k E_\lambda^{2k+1} \quad \text{is an isomorphism for } \lambda \neq 0 \qquad (2.39)$$

which of course implies that

$$\sum_k (-1)^k \dim E_\lambda^k = 0$$

This shows that the index of the operator $d + \delta$ acting on $\bigoplus L_2^k$ is the Euler characteristic of the manifold. (The index of any operator is the difference between the dimensions of the kernel and cokernel).

Let $P_{k,\lambda}$ denote the projection of L_2^k onto E_λ^k. So

$$e^{-t\Delta} = \sum e^{-\lambda t} P_{k,\lambda}$$

is the solution of the heat equation on L_2^k. As $t \to \infty$ this approaches the operator H projecting L_2^k onto $\mathcal{H}_k$. Letting Δ_k denote the operator Δ on L_2^k we see that

$$\operatorname{tr} e^{-t\Delta_k} = \sum e^{-\lambda_k}$$

where the sum is over all eigenvalues λ_k of Δ_k counted with multiplicity. It follows from (2.39) that the alternating sum over k of the corresponding sum over non-zero eigenvalues vanishes. Hence

$$\sum (-1)^k \operatorname{tr} e^{-t\Delta_k} = \chi(M)$$

is independent of t. The index theorem computes this trace for small values of t in terms of local geometric invariants.

The operator $d + \delta$ is an example of a Dirac operator whose general definition we will not give here. The corresponding assertion and local evaluation is the content of the celebrated Atiyah-Singer index theorem, one of the most important theorems discovered in the twentieth century.

2.11 The resolvent.

In order to connect what we have done here notation that will come later, it is convenient to let $A = -L$ so that now the operator

$$(zI - A)^{-1}$$

is compact as an operator on $\mathbf{H}_0$ for z sufficiently negative. (I have dropped the ι_m which should come in front of this expression.) The operator A now has only

finitely many positive eigenvalues, with the corresponding spaces of eigenvectors being finite dimensional. In fact, the eigenvectors $\lambda_n = \lambda_n(A)$ (counted with multiplicity) approach $-\infty$ as $n \to \infty$ and the operator $(zI - A)^{-1}$ exists and is a bounded (in fact compact) operator so long as $z \neq \lambda_n$ for any n. Indeed, we can write any $u \in \mathbf{H}_0$ as

$$u = \sum_n a_n \phi_n$$

where ϕ_n is an eigenvector of A with eigenvalue λ_n and the ϕ form an orthonormal basis of $\mathbf{H}_0$. Then

$$(zI - A)^{-1}u = \sum \frac{1}{z - \lambda_n} a_n \phi_n.$$

The operator $(zI - A)^{-1}$ is called the **resolvent** of A at the point z and denoted by

$$R(z, A)$$

or simply by $R(z)$ if A is fixed. So

$$R(z, A) := (zI - A)^{-1}$$

for those values of $z \in \mathbf{C}$ for which the right hand side is defined.

If z and are complex numbers with $\mathrm{Re}z > \mathrm{Re}a$, then the integral

$$\int_0^\infty e^{-zt} e^{at} dt$$

converges, and we can evaluate it as

$$\frac{1}{z - a} = \int_0^\infty e^{-zt} e^{at} dt.$$

If $\mathrm{Re}z$ is greater than the largest of the eigenvalues of A we can write

$$R(z, A) = \int_0^\infty e^{-zt} e^{tA} dt$$

where we may interpret this equation as a shorthand for doing the integral for the coefficient of each eigenvector, as above, or as an actual operator valued integral. We will spend a lot of time later on in this course generalizing this formula and deriving many consequences from it.

Chapter 3

The Fourier Transform.

3.1   Conventions, especially about 2π.

The space $\mathcal{S}$ consists of all functions on $mathbbR$ which are infinitely differentiable and vanish at infinity rapidly with all their derivatives in the sense that

$$\|f\|_{m,n} := \sup\{|x^m f^{(n)}(x)|\} < \infty.$$

The $\|\cdot\|_{m,n}$ give a family of semi-norms on $\mathcal{S}$ making $\mathcal{S}$ into a Frechet space - that is, a vector space space whose topology is determined by a countable family of semi-norms. More about this later in the course. We use the measure

$$\frac{1}{\sqrt{2\pi}}dx$$

on $\mathbf{R}$ and so define the Fourier transform of an element of $\mathcal{S}$ by

$$\hat{f}(\xi) := \frac{1}{\sqrt{2\pi}} \int_{\mathbf{R}} f(x)e^{-ix\xi}dx$$

and the convolution of two elements of $\mathcal{S}$ by

$$(f \star g)(x) := \frac{1}{\sqrt{2\pi}} \int_{\mathbf{R}} f(x-t)g(t)dt.$$

The Fourier transform is well defined on $\mathcal{S}$ and

$$\left[\left(\frac{d}{dx}\right)^m ((-ix)^n f)\right]\hat{} = (i\xi)^m \left(\frac{d}{d\xi}\right)^n \hat{f},$$

as follows by differentiation under the integral sign and by integration by parts. This shows that the Fourier transform maps $\mathcal{S}$ to $\mathcal{S}$.

3.2   Convolution goes to multiplication.

$$
\begin{aligned}
(f \star g)\hat{}(\xi) &= \frac{1}{2\pi} \int \int f(x-t)g(t)dx e^{-ix\xi}dx \\
&= \frac{1}{2\pi} \int \int f(u)g(t)e^{-i(u+t)\xi}dudt \\
&= \frac{1}{\sqrt{2\pi}} \int_{\mathbf{R}} f(u)e^{-iu\xi}du \frac{1}{\sqrt{2\pi}} \int_{\mathbf{R}} g(t)e^{-it\xi}dt
\end{aligned}
$$

so

$$
(f \star g)\hat{} = \hat{f}\hat{g}.
$$

3.3   Scaling.

For any $f \in \mathcal{S}$ and $a > 0$ define $S_a f$ by $(S_a)f(x) := f(ax)$. Then setting $u = ax$ so $dx = (1/a)du$ we have

$$
\begin{aligned}
(S_a f)\hat{}(\xi) &= \frac{1}{\sqrt{2\pi}} \int_{\mathbf{R}} f(ax)e^{-ix\xi}dx \\
&= \frac{1}{\sqrt{2\pi}} \int_{\mathbf{R}} (1/a)f(u)e^{-iu(\xi/a)}du
\end{aligned}
$$

so

$$
(S_a f)\hat{} = (1/a)S_{1/a}\hat{f}.
$$

3.4   Fourier transform of a Gaussian is a Gaussian.

The polar coordinate trick evaluates

$$
\frac{1}{\sqrt{2\pi}} \int_{\mathbf{R}} e^{-x^2/2}dx = 1.
$$

The integral

$$
\frac{1}{\sqrt{2\pi}} \int_{\mathbf{R}} e^{-x^2/2-x\eta}dx
$$

converges for all complex values of η, uniformly in any compact region. Hence it defines an analytic function of η that can be evaluated by taking η to be real and then using analytic continuation. For real η we complete the square and make a change of variables:

$$
\begin{aligned}
\frac{1}{\sqrt{2\pi}} \int_{\mathbf{R}} e^{-x^2/2-x\eta}dx &= \frac{1}{\sqrt{2\pi}} \int_{\mathbf{R}} e^{-(x+\eta)^2/2+\eta^2/2}dx \\
&= e^{\eta^2/2}\frac{1}{\sqrt{2\pi}} \int_{\mathbf{R}} e^{-(x+\eta)^2/2}dx \\
&= e^{\eta^2/2}.
\end{aligned}
$$

Setting $\eta = i\xi$ gives

$$\hat{n} = n \quad \text{if } n(x) := e^{-x^2/2}.$$

If we set $a = \epsilon$ in our scaling equation and define

$$\rho_\epsilon := S_\epsilon n$$

so

$$\rho_\epsilon(x) = e^{-\epsilon^2 x^2/2},$$

then

$$(\rho_\epsilon)\hat{}(x) = \frac{1}{\epsilon} e^{-x^2/2\epsilon^2}.$$

Notice that for any $g \in \mathcal{S}$ we have

$$\int_{\mathbf{R}} (1/a)(S_{1/a}g)(\xi)d\xi = \int_{\mathbf{R}} g(\xi)d\xi$$

so setting $a = \epsilon$ we conclude that

$$\frac{1}{\sqrt{2\pi}} \int_{\mathbf{R}} (\rho_\epsilon)\hat{}(\xi)d\xi = 1$$

for all ϵ.

Let

$$\psi := \psi_1 := (\rho_1)\hat{}$$

and

$$\psi_\epsilon := (\rho_\epsilon)\hat{}.$$

Then

$$\psi_\epsilon(\eta) = \frac{1}{\epsilon} \psi \left(\frac{\eta}{\epsilon} \right)$$

so

$$(\psi_\epsilon \star g)(\xi) - g(\xi) = \frac{1}{\sqrt{2\pi}} \int_{\mathbf{R}} [g(\xi - \eta) - g(\xi)] \frac{1}{\epsilon} \psi \left(\frac{\eta}{\epsilon} \right) d\eta =$$

$$= \frac{1}{\sqrt{2\pi}} \int_{\mathbf{R}} [g(\xi - \epsilon\zeta) - g(\xi)]\psi(\zeta)d\zeta.$$

Since $g \in \mathcal{S}$ it is uniformly continuous on $\mathbf{R}$, so that for any $\delta > 0$ we can find ϵ_0 so that the above integral is less than δ in absolute value for all $0 < \epsilon < \epsilon_0$. In short,

$$\|\psi_\epsilon \star g - g\|_\infty \to 0, \quad \text{as } \epsilon \to 0.$$

3.5   The multiplication formula.

This says that

$$\int_{\mathbf{R}} \hat{f}(x)g(x)dx = \int_{\mathbf{R}} f(x)\hat{g}(x)dx$$

for any $f, g \in \mathcal{S}$. Indeed the left hand side equals

$$\frac{1}{\sqrt{2\pi}} \int_{\mathbf{R}} \int_{\mathbf{R}} f(y)e^{-ixy}dy g(x)dx.$$

We can write this integral as a double integral and then interchange the order of integration which gives the right hand side.

3.6   The inversion formula.

This says that for any $f \in \mathcal{S}$

$$f(x) = \frac{1}{\sqrt{2\pi}} \int_{\mathbf{R}} \hat{f}(\xi)e^{ix\xi}d\xi.$$

To prove this, we first observe that for any $h \in \mathcal{S}$ the Fourier transform of $x \mapsto e^{i\eta x}h(x)$ is just $\xi \mapsto \hat{h}(\xi - \eta)$ as follows directly from the definition.

Taking $g(x) = e^{itx}e^{-\epsilon^2 x^2/2}$ in the multiplication formula gives

$$\frac{1}{\sqrt{2\pi}} \int_{\mathbf{R}} \hat{f}(t)e^{itx}e^{-\epsilon^2 t^2/2}dt = \frac{1}{\sqrt{2\pi}} \int_{\mathbf{R}} f(t)\psi_\epsilon(t-x)dt = (f \star \psi_\epsilon)(x).$$

We know that the right hand side approaches $f(x)$ as $\epsilon \to 0$. Also, $e^{-\epsilon^2 t^2/2} \to 1$ for each fixed t, and in fact uniformly on any bounded t interval. Furthermore, $0 < e^{-\epsilon^2 t^2/2} \leq 1$ for all t. So choosing the interval of integration large enough, we can take the left hand side as close as we like to $\frac{1}{\sqrt{2\pi}} \int_{\mathbf{R}} \hat{f}(x)e^{ixt}dt$ by then choosing ϵ sufficiently small. QED

3.7   Plancherel's theorem

Let

$$\tilde{f}(x) := \overline{f(-x)}.$$

Then the Fourier transform of $\tilde{f}$ is given by

$$\frac{1}{\sqrt{2\pi}} \int_{\mathbf{R}} \overline{f(-x)}e^{-ix\xi}dx = \frac{1}{\sqrt{2\pi}} \int_{\mathbf{R}} \overline{f(u)}e^{iu\xi}du = \overline{\hat{f}(\xi)}$$

so

$$(\tilde{f})\hat{} = \overline{\hat{f}}.$$

Thus

$$(f \star \tilde{f})\hat{} = |\hat{f}|^2.$$

The inversion formula applied to $f \star \tilde{f}$ and evaluated at 0 gives

$$(f \star \tilde{f})(0) = \frac{1}{\sqrt{2\pi}} \int_{\mathbf{R}} |\hat{f}|^2 dx.$$

The left hand side of this equation is

$$\frac{1}{\sqrt{2\pi}} \int_{\mathbf{R}} f(x) \tilde{f}(0 - x) dx = \frac{1}{\sqrt{2\pi}} \int_{\mathbf{R}} |f(x)|^2 dx.$$

Thus we have proved Plancherel's formula

$$\frac{1}{\sqrt{2\pi}} \int_{\mathbf{R}} |f(x)|^2 dx = \frac{1}{\sqrt{2\pi}} \int_{\mathbf{R}} |\hat{f}(x)|^2 dx.$$

Define $L_2(\mathbf{R})$ to be the completion of $\mathcal{S}$ with respect to the L_2 norm given by the left hand side of the above equation. Since $\mathcal{S}$ is dense in $L_2(\mathbf{R})$ we conclude that the Fourier transform extends to unitary isomorphism of $L_2(\mathbf{R})$ onto itself.

3.8 The Poisson summation formula.

This says that for any $g \in \mathcal{S}$ we have

$$\sum_k g(2\pi k) = \frac{1}{\sqrt{2\pi}} \sum_m \hat{g}(m).$$

To prove this let

$$h(x) := \sum_k g(x + 2\pi k)$$

so h is a smooth function, periodic of period 2π and

$$h(0) = \sum_k g(2\pi k).$$

We may expand h into a Fourier series

$$h(x) = \sum_m a_m e^{imx}$$

where

$$a_m = \frac{1}{2\pi} \int_0^{2\pi} h(x) e^{-imx} dx = \frac{1}{2\pi} \int_{\mathbf{R}} g(x) e^{-imx} dx = \frac{1}{\sqrt{2\pi}} \hat{g}(m).$$

Setting $x = 0$ in the Fourier expansion

$$h(x) = \frac{1}{\sqrt{2\pi}} \sum \hat{g}(m) e^{imx}$$

gives

$$h(0) = \frac{1}{\sqrt{2\pi}} \sum_m \hat{g}(m).$$

3.9   The Shannon sampling theorem.

Let $f \in \mathcal{S}$ be such that its Fourier transform is supported in the interval $[-\pi, \pi]$. Then a knowledge of $f(n)$ for all $n \in \mathbf{Z}$ determines f. More explicitly,

$$f(t) = \frac{1}{\pi} \sum_{n=-\infty}^{\infty} f(n) \frac{\sin \pi(n-t)}{n-t}. \tag{3.1}$$

Proof. Let g be the periodic function (of period 2π) which extends $\hat{f}$, the Fourier transform of f. So

$$g(\tau) = \hat{f}(\tau), \quad \tau \in [-\pi, \pi]$$

and is periodic.

Expand g into a Fourier series:

$$g = \sum_{n \in \mathbf{Z}} c_n e^{in\tau},$$

where

$$c_n = \frac{1}{2\pi} \int_{-\pi}^{\pi} g(\tau) e^{-in\tau} d\tau = \frac{1}{2\pi} \int_{-\infty}^{\infty} \hat{f}(\tau) e^{-in\tau} d\tau,$$

or

$$c_n = \frac{1}{(2\pi)^{\frac{1}{2}}} f(-n).$$

But

$$f(t) = \frac{1}{(2\pi)^{\frac{1}{2}}} \int_{-\infty}^{\infty} \hat{f}(\tau) e^{it\tau} d\tau = \frac{1}{(2\pi)^{\frac{1}{2}}} \int_{-\pi}^{\pi} g(\tau) e^{it\tau} d\tau =$$

$$\frac{1}{(2\pi)^{\frac{1}{2}}} \int_{-\pi}^{\pi} \sum \frac{1}{(2\pi)^{\frac{1}{2}}} f(-n) e^{i(n+t)\tau} d\tau.$$

Replacing n by $-n$ in the sum, and interchanging summation and integration, which is legitimate since the $f(n)$ decrease very fast, this becomes

$$f(t) = \frac{1}{2\pi} \sum_n f(n) \int_{-\pi}^{\pi} e^{i(t-n)\tau} d\tau.$$

But

$$\int_{-\pi}^{\pi} e^{i(t-n)\tau} d\tau = \frac{e^{i(t-n)\tau}}{i(t-n)} \Big|_{-\pi}^{\pi} = \frac{e^{i(t-n)\pi} - e^{i(t-n)\pi}}{i(t-n)} = 2 \frac{\sin \pi(n-t)}{n-t}. \quad \text{QED}$$

It is useful to reformulate this via rescaling so that the interval $[-\pi, \pi]$ is replaced by an arbitrary interval symmetric about the origin: In the engineering literature the **frequency** λ is defined by

$$\xi = 2\pi\lambda.$$

Suppose we want to apply (3.1) to $g = S_a f$. We know that the Fourier transform of g is $(1/a)S_{1/a}\hat{f}$ and

$$\text{supp } S_{1/a}\hat{f} = a\text{supp } \hat{f}.$$

So if

$$\text{supp } \hat{f} \subset [-2\pi\lambda_c, 2\pi\lambda_c]$$

we want to choose a so that $a2\pi\lambda_c \leq \pi$ or

$$a \leq \frac{1}{2\lambda_c}. \tag{3.2}$$

For a in this range (3.1) says that

$$f(ax) = \frac{1}{\pi}\sum f(na)\frac{\sin\pi(x-n)}{x-n},$$

or setting $t = ax$,

$$f(t) = \sum_{n=-\infty}^{\infty} f(na)\frac{\sin(\frac{\pi}{a}(t-na))}{\frac{\pi}{a}(t-na)}. \tag{3.3}$$

This holds in L_2 under the assumption that f satisfies supp $\hat{f} \subset [-2\pi\lambda_c, 2\pi\lambda_c]$. We say that f has **finite bandwidth** or is **bandlimited** with bandlimit λ_c. The critical value $a_c = 1/2\lambda_c$ is known as the **Nyquist sampling interval** and $(1/a) = 2\lambda_c$ is known as the **Nyquist sampling rate**. Thus the Shannon sampling theorem says that a band-limited signal can be recovered completely from a set of samples taken at a rate $\geq$ the Nyquist sampling rate.

3.10 The Heisenberg Uncertainty Principle.

Let $f \in \mathcal{S}(\mathbf{R})$ with

$$\int |f(x)|^2 dx = 1.$$

We can think of $x \mapsto |f(x)|^2$ as a probability density on the line. The mean of this probability density is

$$x_m := \int x|f(x)|^2 dx.$$

If we take the Fourier transform, then Plancherel says that

$$\int |\hat{f}(\xi)|^2 d\xi = 1$$

as well, so it defines a probability density with mean

$$\xi_m := \int \xi|\hat{f}(\xi)|^2 d\xi.$$

Suppose for the moment that these means both vanish. The **Heisenberg Uncertainty Principle** says that

$$\left(\int |xf(x)|^2 dx\right)\left(\int |\xi \hat{f}(\xi)|^2 d\xi\right) \geq \frac{1}{4}.$$

Proof. Write $-i\xi f(\xi)$ as the Fourier transform of f' and use Plancherel to write the second integral as $\int |f'(x)|^2 dx$. Then the Cauchy - Schwarz inequality says that the left hand side is $\geq$ the square of

$$\int |xf(x)f'(x)|dx \geq \left|\int \mathrm{Re}(xf(x)\overline{f'(x)})dx\right| =$$

$$\frac{1}{2}\left|\int x(f(x)\overline{f'(x)} + \overline{f(x)}f'(x)dx\right|$$

$$= \frac{1}{2}\left|\int x\frac{d}{dx}|f|^2 dx\right| = \frac{1}{2}\left|\int -|f|^2 dx\right| = \frac{1}{2}. \quad \text{QED}$$

If f has norm one but the mean of the probability density $|f|^2$ is not necessarily of zero (and similarly for for its Fourier transform) the Heisenberg uncertainty principle says that

$$\left(\int |(x - x_m)f(x)|^2 dx\right)\left(\int |(\xi - \xi_m)\hat{f}(\xi)|^2 d\xi\right) \geq \frac{1}{4}.$$

The general case is reduced to the special case by replacing $f(x)$ by

$$f(x + x_m)e^{i\xi_m x}.$$

3.11   Tempered distributions.

The space $\mathcal{S}$ was defined to be the collection of all smooth functions on $\mathbb{R}$ such that
$$\|f\|_{m,n} := \sup_x\{|x^m f^{(n)}(x)|\} < \infty.$$

The collection of these norms define a topology on $\mathcal{S}$ which is much finer that the L_2 topology: We declare that a sequence of functions $\{f_k\}$ approaches $g \in \mathcal{S}$ if and only if
$$\|f_k - g\|_{m,n} \to 0$$

for every m and n.

A linear function on $\mathcal{S}$ which is continuous with respect to this topology is called a **tempered distribution**.

The space of tempered distributions is denoted by $\mathcal{S}'$. For example, every element $f \in \mathcal{S}$ defines a linear function on $\mathcal{S}$ by

$$\phi \mapsto \langle \phi, f\rangle = \frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}} \phi(x)\overline{f(x)}dx.$$

But this last expression makes sense for any element $f \in L_2(\mathbb{R})$, or for any piecewise continuous function f which grows at infinity no faster than any polynomial. For example, if $f \equiv 1$, the linear function associated to f assigns to ϕ the value

$$\frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \phi(x)dx.$$

This is clearly continuous with respect to the topology of $\mathcal{S}$ but this function of ϕ does not make sense for a general element ϕ of $L_2(\mathbb{R})$.

Another example of an element of $\mathcal{S}'$ is the Dirac δ-function which assigns to $\phi \in \mathcal{S}$ its value at 0. This is an element of $\mathcal{S}'$ but makes no sense when evaluated on a general element of $L_2(\mathbb{R})$.

If $f \in \mathcal{S}$, then the Plancherel formula formula implies that its Fourier transform $\mathcal{F}(f) = \hat{f}$ satisfies

$$(\phi, f) = (\mathcal{F}(\phi), \mathcal{F}(f)).$$

But we can now use this equation to *define* the Fourier transform of an arbitrary element of $\mathcal{S}'$: If $\ell \in \mathcal{S}'$ we define $\mathcal{F}(\ell)$ to be the linear function

$$\mathcal{F}(\ell)(\psi) := \ell(\mathcal{F}^{-1}(\psi)).$$

3.11.1 Examples of Fourier transforms of elements of $\mathcal{S}'$.

- If ℓ corresponds to the function $f \equiv 1$, then

$$\mathcal{F}(\ell)(\psi) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} (\mathcal{F}^{-1}\psi)(\xi)d\xi = \mathcal{F}\left(\mathcal{F}^{-1}\psi\right)(0) = \psi(0).$$

  So the Fourier transform of the function which is identically one is the Dirac δ-function.

- If δ denotes the Dirac δ-function, then

$$(\mathcal{F}(\delta)(\psi) = \delta(\mathcal{F}^{-1}(\psi)) = \left((\mathcal{F}^{-1}(\psi)\right)(0) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \psi(x)dx.$$

  So the Fourier transform of the Dirac δ function is the function which is identically one.

- In fact, this last example follows from the preceding one: If $m = \mathcal{F}(\ell)$ then

$$(\mathcal{F}(m)(\phi) = m(\mathcal{F}^{-1}(\phi)) = \ell(\mathcal{F}^{-1}(\mathcal{F}^{-1}(\phi))).$$

  But

$$\mathcal{F}^{-2}(\phi)(x) = \phi(-x).$$

  So if $m = \mathcal{F}(\ell)$ then $\mathcal{F}(m) = \check{\ell}$ where

$$\check{\ell}(\phi) := \ell(\phi(-\bullet)).$$

- The Fourier transform of the function x: This assigns to every $\psi \in \mathcal{S}$ the value

$$\frac{1}{\sqrt{2\pi}} \int \psi(\xi) e^{ix\xi} x \, d\xi \, dx = \frac{1}{\sqrt{2\pi}} \int \psi(\xi) \frac{1}{i} \frac{d}{d\xi} \left(e^{ix\xi} \right) d\xi \, dx =$$

$$i \frac{1}{\sqrt{2\pi}} \int \frac{d\psi(\xi)}{dx} e^{ix\xi} d\xi \, dx = i \left(\mathcal{F} \left((\mathcal{F}^{-1} \left(\frac{d\psi(\xi)}{dx} \right)) \right) \right)(0) = i\delta \left(\frac{d\psi(\xi)}{dx} \right).$$

Now for an element of $\mathcal{S}$ we have

$$\int \frac{d\phi}{dx} \cdot \overline{f} \, dx = -\frac{1}{\sqrt{2\pi}} \int \phi \overline{\frac{df}{dx}} \, dx.$$

So we define the derivative of an $\ell \in \mathcal{S}'$ by

$$\frac{d\ell}{dx}(\phi) = \ell \left(-\frac{d\phi}{dx} \right).$$

So the Fourier transform of x is $-i\frac{d\delta}{dx}$.

Chapter 4

Measure theory.

4.1 Lebesgue outer measure.

We recall some results from the chapter on metric spaces: For any subset $A \subset \mathbf{R}$ we defined its **Lebesgue outer measure** by

$$m^*(A) := \inf \sum \ell(I_n) : \ I_n \text{ are intervals with } A \subset \bigcup I_n. \qquad (4.1)$$

Here the length $\ell(I)$ of any interval $I = [a, b]$ is $b - a$ with the same definition for half open intervals $(a, b]$ or $[a, b)$, or open intervals. Of course if $a = -\infty$ and b is finite or $+\infty$, or if a is finite and $b = +\infty$ the length is infinite. So the infimum in (4.1) is taken over all covers of A by intervals. By the usual $\epsilon/2^n$ trick, i.e. by replacing each $I_j = [a_j, b_j]$ by $(a_j - \epsilon/2^{j+1}, b_j + \epsilon/2^{j+1})$ we may assume that the infimum is taken over open intervals. (Equally well, we could use half open intervals of the form $[a, b)$, for example.).

It is clear that if $A \subset B$ then $m^*(A) \le m^*(B)$ since any cover of B by intervals is a cover of A. Also, if Z is any set of measure zero, then $m^*(A \cup Z) = m^*(A)$. In particular, $m^*(Z) = 0$ if Z has measure zero. Also, if $A = [a, b]$ is an interval, then we can cover it by itself, so

$$m^*([a, b]) \le b - a,$$

and hence the same is true for $(a, b]$, $[a, b)$, or (a, b). If the interval is infinite, it clearly can not be covered by a set of intervals whose total length is finite, since if we lined them up with end points touching they could not cover an infinite interval. We recall the proof that

$$m^*(I) = \ell(I) \qquad (4.2)$$

if I is a finite interval: We may assume that $I = [c, d]$ is a closed interval by what we have already said, and that the minimization in (4.1) is with respect to a cover by open intervals. So what we must show is that if

$$[c, d] \subset \bigcup_i (a_i, b_i)$$

then

$$d - c \leq \sum_i (b_i - a_i).$$

We first applied Heine-Borel to replace the countable cover by a finite cover. (This only decreases the right hand side of preceding inequality.) So let n be the number of elements in the cover. We needed to prove that if

$$[c, d] \subset \bigcup_{i=1}^{n}(a_i, b_i) \text{ then } d - c \leq \sum_{i=1}^{n}(b_i - a_i),$$

and we did this this by induction on n. If $n = 1$ then $a_1 < c$ and $b_1 > d$ so clearly $b_1 - a_1 > d - c$.

Suppose that $n \geq 2$ and we know the result for all covers (of all intervals $[c, d]$) with at most $n - 1$ intervals in the cover. If some interval (a_i, b_i) is disjoint from $[c, d]$ we may eliminate it from the cover, and then we are in the case of $n - 1$ intervals. So every (a_i, b_i) has non-empty intersection with $[c, d]$. Among the the intervals (a_i, b_i) there will be one for which a_i takes on the minimum possible value. By relabeling, we may assume that this is (a_1, b_1). Since c is covered, we must have $a_1 < c$. If $b_1 > d$ then (a_1, b_1) covers $[c, d]$ and there is nothing further to do. So assume $b_1 \leq d$. We must have $b_1 > c$ since $(a_1, b_1) \cap [c, d] \neq \emptyset$. Since $b_1 \in [c, d]$, at least one of the intervals (a_i, b_i), $i > 1$ contains the point b_1. By relabeling, we may assume that it is (a_2, b_2). But now we have a cover of $[c, d]$ by $n - 1$ intervals:

$$[c, d] \subset (a_1, b_2) \cup \bigcup_{i=3}^{n}(a_i, b_i).$$

So by induction

$$d - c \leq (b_2 - a_1) + \sum_{i=3}^{n}(b_i - a_i).$$

But $b_2 - a_1 \leq (b_2 - a_2) + (b_1 - a_1)$ since $a_2 < b_1$. QED

We repeat that the intervals used in (4.1) could be taken as open, closed or half open without changing the definition. If we take them all to be half open, of the form $I_i = [a_i, b_i)$, we can write each I_i as a disjoint union of finite or countably many intervals each of length $< \epsilon$. So it makes no difference to the definition if we also require the

$$\ell(I_i) < \epsilon \tag{4.3}$$

in (4.1). We will see that when we pass to other types of measures this will make a difference.

We have verified, or can easily verify the following properties:

1.

$$m^*(\emptyset) = 0.$$

2.
$$A \subset B \Rightarrow m^*(A) \leq m^*(B).$$

3.
$$m^*(\bigcup_i A_i) \leq \sum_i m^*(A_i).$$

4. If dist $(A, B) > 0$ then
$$m^*(A \cup B) = m^*(A) + m^*(B).$$

5.
$$m^*(A) = \inf\{m^*(U): \ U \supset A, \ \ U \text{ open}\}.$$

6. For an interval
$$m^*(I) = \ell(I).$$

The only items that we have not done already are items 4 and 5. But these are immediate: for 4 we may choose the intervals in (4.1) all to have length $< \epsilon$ where $2\epsilon < $ dist (A, B) so that there is no overlap. As for item 5, we know from 2 that $m^*(A) \leq m^*(U)$ for any set U, in particular for any open set U which contains A. We must prove the reverse inequality: if $m^*(A) = \infty$ this is trivial. Otherwise, we may take the intervals in (4.1) to be open and then the union on the right is an open set whose Lebesgue outer measure is less than $m^*(A) + \delta$ for any $\delta > 0$ if we choose a close enough approximation to the infimum.

I should also add that all the above works for $\mathbf{R}^n$ instead of $\mathbf{R}$ if we replace the word "interval" by "rectangle", meaning a rectangular parallelepiped, i.e a set which is a product of one dimensional intervals. We also replace length by volume (or area in two dimensions). What is needed is the following

Lemma 4.1.1 *Let $\mathcal{C}$ be a finite non-overlapping collection of closed rectangles all contained in the closed rectangle J. Then*

$$\text{vol } J \geq \sum_{I \in \mathcal{C}} \text{vol } I.$$

If $\mathcal{C}$ is any finite collection of rectangles such that

$$J \subset \bigcup_{I \in \mathcal{C}} I$$

then

$$\text{vol } J \leq \sum_{I \in \mathcal{C}} \text{vol } (I).$$

This lemma occurs on page 1 of Strook, *A concise introduction to the theory of integration* together with its proof. I will take this for granted. In the next few paragraphs I will talk as if we are in $\mathbf{R}$, but everything goes through unchanged if $\mathbf{R}$ is replaced by $\mathbf{R}^n$.

4.2   Lebesgue inner measure.

Item 5. in the preceding paragraph says that the Lebesgue outer measure of any set is obtained by approximating it from the outside by open sets. The **Lebesgue inner measure** is defined as

$$m_*(A) = \sup\{m^*(K): \ K \subset A, \ K \text{ compact }\}. \tag{4.4}$$

Clearly

$$m_*(A) \le m^*(A)$$

since $m^*(K) \le m^*(A)$ for any $K \subset A$. We also have

Proposition 4.2.1 *For any interval I we have*

$$m_*(I) = \ell(I). \tag{4.5}$$

Proof. If $\ell(I) = \infty$ the result is obvious. So we may assume that I is a finite interval which we may assume to be open, $I = (a, b)$. If $K \subset I$ is compact, then I is a cover of K and hence from the definition of outer measure $m^*(K) \le \ell(I)$. So $m_*(I) \le \ell(I)$. On the other hand, for any $\epsilon > 0$, $\epsilon < \frac{1}{2}(b - a)$ the interval $[a + \epsilon, b - \epsilon]$ is compact and $m^*([a - \epsilon, a + \epsilon]) = b - a - 2\epsilon \le m_*(I)$. Letting $\epsilon \to 0$ proves the proposition. QED

4.3   Lebesgue's definition of measurability.

A set A with $m^*(A) < \infty$ is said to **measurable in the sense of Lebesgue** if

$$m_*(A) = m^*(A). \tag{4.6}$$

If A is measurable in the sense of Lebesgue, we write

$$m(A) = m_*(A) = m^*(A). \tag{4.7}$$

If K is a compact set, then $m_*(K) = m^*(K)$ since K is a compact set contained in itself. Hence all compact sets are measurable in the sense of Lebesgue. If I is a bounded interval, then I is measurable in the sense of Lebesgue by Proposition 4.2.1.

If $m^*(A) = \infty$, we say that A is measurable in the sense of Lebesgue if all of the sets $A \cap [-n, n]$ are measurable.

Proposition 4.3.1 *If $A = \bigcup A_i$ is a (finite or) countable disjoint union of sets which are measurable in the sense of Lebesgue, then A is measurable in the sense of Lebesgue and*

$$m(A) = \sum_i \ m(A_i).$$

Proof. We may assume that $m(A) < \infty$ - otherwise apply the result to $A \cap [-n, n]$ and $A_i \cap [-n, n]$ for each n. We have

$$m^*(A) \leq \sum_n m^*(A_n) = \sum_n m(A_n).$$

Let $\epsilon > 0$, and for each n choose compact $K_n \subset A_n$ with

$$m^*(K_n) \geq m_*(A_n) - \frac{\epsilon}{2^n} = m(A_n) - \frac{\epsilon}{2^n}$$

since A_n is measurable in the sense of Lebesgue. The sets K_n are pairwise disjoint, hence, being compact, at positive distances from one another. Hence

$$m^*(K_1 \cup \cdots \cup K_n) = m^*(K_1) + \cdots + m^*(K_n)$$

and $K_1 \cup \cdots \cup K_n$ is compact and contained in A. Hence

$$m_*(A) \geq m^*(K_1) + \cdots + m^*(K_n),$$

and since this is true for all n we have

$$m_*(A) \geq \sum_n m(A_n) - \epsilon.$$

Since this is true for all $\epsilon > 0$ we get

$$m_*(A) \geq \sum m(A_n).$$

But then $m_*(A) \geq m^*(A)$ and so they are equal, so A is measurable in the sense of Lebesgue, and $m(A) = \sum m(A_i)$. QED

Proposition 4.3.2 *Open sets and closed sets are measurable in the sense of Lebesgue.*

Proof. Any open set O can be written as the countable union of open intervals I_i, and

$$J_n := I_n \setminus \bigcup_{i=1}^{n-1} I_i$$

is a disjoint union of intervals (some open, some closed, some half open) and O is the disjont union of the J_n. So every open set is a disjoint union of intervals hence measurable in the sense of Lebesgue.

If F is closed, and $m^*(F) = \infty$, then $F \cap [-n, n]$ is compact, and so F is measurable in the sense of Lebesgue. Suppose that

$$m^*(F) < \infty.$$

For any $\epsilon > 0$ consider the sets

$$
\begin{aligned}
G_{1,\epsilon} &:= [-1 + \frac{\epsilon}{2^2}, 1 - \frac{\epsilon}{2^2}] \cap F \\
G_{2,\epsilon} &:= ([-2 + \frac{\epsilon}{2^3}, -1] \cap F) \cup ([1, 2 - \frac{\epsilon}{2^3}] \cap F) \\
G_{3,\epsilon} &:= ([-3 + \frac{\epsilon}{2^4}, -2] \cap F) \cup ([2, 3 - \frac{\epsilon}{2^4}] \cap F) \\
&\ \ \vdots
\end{aligned}
$$

and set

$$
G_\epsilon := \bigcup_i G_{i,\epsilon}.
$$

The $G_{i,\epsilon}$ are all compact, and hence measurable in the sense of Lebesgue, and the union in the definition of G_ϵ is disjoint, so is measurable in the sense of Lebesgue. Furthermore, the sum of the lengths of the "gaps" between the intervals that went into the definition of the $G_{i,\epsilon}$ is ϵ. So

$$
m(G_\epsilon) + \epsilon = m^*(G_\epsilon) + \epsilon \geq m^*(F) \geq m^*(G_\epsilon) = m(G_\epsilon) = \sum_i m(G_{i,\epsilon}).
$$

In particular, the sum on the right converges, and hence by considering a finite number of terms, we will have a finite sum whose value is at least $m(G_\epsilon) - \epsilon$. The corresponding union of sets will be a compact set K_ϵ contained in F with

$$
m(K_\epsilon) \geq m^*(F) - 2\epsilon.
$$

Hence all closed sets are measurable in the sense of Lebesgue. QED

Theorem 4.3.1 *A is measurable in the sense of Lebesgue if and only if for every $\epsilon > 0$ there is an open set $U \supset A$ and a closed set $F \subset A$ such that*

$$
m(U \setminus F) < \epsilon.
$$

Proof.   Suppose that A is measurable in the sense of Lebesgue with $m(A) < \infty$. Then there is an open set $U \supset A$ with $m(U) < m^*(A) + \epsilon/2 = m(A) + \epsilon/2$, and there is a compact set $F \subset A$ with $m(F) \geq m_*(A) - \epsilon = m(A) - \epsilon/2$. Since $U \setminus F$ is open, it is measurable in the sense of Lebesgue, and so is F as it is compact. Also F and $U \setminus F$ are disjoint. Hence by Proposition 4.3.1,

$$
m(U \setminus F) = m(U) - m(F) < m(A) + \frac{\epsilon}{2} - \left(m(A) - \frac{\epsilon}{2} \right) = \epsilon.
$$

If A is measurable in the sense of Lebesgue, and $m(A) = \infty$, we can apply the above to $A \cap I$ where I is any compact interval. So we can find open sets $U_n \supset A \cap [-n - 2\delta_{n+1}, n + 2\delta_{n+1}]$ and closed sets $F_n \subset A \cap [-n, n]$ with $m(U_n \setminus F_n) < \epsilon/2^n$. Here the δ_n are sufficiently small positive numbers. We

may enlarge each F_n if necessary so that $F_n \cap [-n+1, n-1] \supset F_{n-1}$. We may also decrease the U_n if necessary so that

$$U_n \cap (-n+1-\delta_n, n-1+\delta_n) \subset U_{n-1}.$$

Indeed, if we set $C_n := [-n+1-\delta_n, n-1+\delta_n] \cap U_{n-1}^c$ then C_n is a closed set with $C_n \cap A = \emptyset$. Then $U_n \cap C_n^c$ is still an open set containing $[-n-2\delta_{n+1}, n+2\delta_{n+1}] \cap A$ and

$$(U_n \cap C_n^c) \cap (-n+1-\delta_n, n-1+\delta_n) \subset C_n^c \cap (-n+1-\delta_n, n-1+\delta_n) \subset U_{n-1}.$$

Take $U := \bigcup U_n$ so U is open. Take

$$F := \bigcup (F_n \cap ([-n, -n+1] \cup [n-1, n])).$$

Then F is closed, $U \supset A \supset F$ and

$$U \setminus F \subset \bigcup (U_n / F_n) \cap ([-n, -n+1] \cup [n-1, n]) \subset \bigcup (U_n \setminus F_n)$$

In the other direction, suppose that for each ϵ, there exist $U \supset A \supset F$ with $m(U \setminus F) < \epsilon$. Suppose that $m^*(A) < \infty$. Then $m(F) < \infty$ and $m(U) \leq m(U \setminus F) + m(F) < \epsilon + m(F) < \infty$. Then

$$m^*(A) \leq m(U) < m(F) + \epsilon = m_*(F) + \epsilon \leq m_*(A) + \epsilon.$$

Since this is true for every $\epsilon > 0$ we conclude that $m_*(A) \geq m^*(A)$ so they are equal and A is measurable in the sense of Lebesgue.

If $m^*(A) = \infty$, we have $U \cap (-n-\epsilon, n+\epsilon) \supset A \cap [-n, n] \supset F \cap [-n, n]$ and

$$m((U \cap (-n-\epsilon, n+\epsilon) \setminus (F \cap [-n, n]) < 2\epsilon + \epsilon = 3\epsilon$$

so we can proceed as before to conclude that $m_*(A \cap [-n, n]) = m^*(A \cap [-n, n])$. QED

Several facts emerge immediately from this theorem:

Proposition 4.3.3 *If A is measurable in the sense of Lebesgue, so is its complement $A^c = \mathbf{R} \setminus A$.*

Indeed, if $F \subset A \subset U$ with F closed and U open, then $F^c \supset A^c \supset U^c$ with F^c open and U^c closed. Furthermore, $F^c \setminus U^c = U \setminus F$ so if A satisfies the condition of the theorem so does A^c.

Proposition 4.3.4 *If A and B are measurable in the sense of Lebesgue so is $A \cap B$.*

For $\epsilon > 0$ choose $U_A \supset A \supset F_A$ and $U_B \supset B \supset F_B$ with $m(U_A \setminus F_A) < \epsilon/2$ and $m(U_B \setminus F_B) < \epsilon/2$. Then $(U_A \cap U_B) \supset (A \cap B) \supset (F_A \cap F_B)$ and $(U_A \cap U_B) \setminus (F_A \cap F_B) \subset (U_A \setminus F_A) \cup (U_B \setminus F_B)$.QED

Putting the previous two propositions together gives

Proposition 4.3.5 *If A and B are measurable in the sense of Lebesgue then so is $A \cup B$.*

Indeed, $A \cup B = (A^c \cap B^c)^c$.

Since $A \setminus B = A \cap B^c$ we also get

Proposition 4.3.6 *If A and B are measurable in the sense of Lebesgue then so is $A \setminus B$.*

4.4   Caratheodory's definition of measurability.

A set $E \subset \mathbf{R}$ is said to be **measurable according to Caratheodory** if for any set $A \subset \mathbf{R}$ we have

$$m^*(A) = m^*(A \cap E) + m^*(A \cap E^c) \tag{4.8}$$

where we recall that E^c denotes the complement of E. In other words, $A \cap E^c = A \setminus E$. This definition has many advantages, as we shall see. Our first task is to show that it is equivalent to Lebesgue's:

Theorem 4.4.1 *A set E is measurable in the sense of Caratheodory if and only if it is measurable in the sense of Lebesgue.*

Proof.   We always have

$$m^*(A) \le m^*(A \cap E) + m^*(A \setminus E)$$

so condition (4.8) is equivalent to

$$m^*(A \cap E) + m^*(A \setminus E) \le m^*(A) \tag{4.9}$$

for all A.

Suppose E is measurable in the sense of Lebesgue. Let $\epsilon > 0$. Choose $U \supset E \supset F$ with U open, F closed and $m(U/F) < \epsilon$ which we can do by Theorem 4.3.1. Let V be an open set containing A. Then $A \setminus E \subset V \setminus F$ and $A \cap E \subset (V \cap U)$ so

$$
\begin{aligned}
m^*(A \setminus E) + m^*(A \cap E) &\le m(V \setminus F) + m(V \cap U) \\
&\le m(V \setminus U) + m(U \setminus F) + m(V \cap U) \\
&\le m(V) + \epsilon.
\end{aligned}
$$

(We can pass from the second line to the third since both $V \setminus U$ and $V \cap U$ are measurable in the sense of Lebesgue and we can apply Proposition 4.3.1.) Taking the infimum over all open V containing A, the last term becomes $m^*(A)$, and as ϵ is arbitrary, we have established (4.9) showing that E is measurable in the sense of Caratheodory.

In the other direction, suppose that E is measurable in the sense of Caratheodory. First suppose that
$$m^*(E) < \infty.$$

Then for any $\epsilon > 0$ there exists an open set $U \supset E$ with $m(U) < m^*(E) + \epsilon$. We may apply condition (4.8) to $A = U$ to get

$$m(U) = m^*(U \cap E) + m^*(U \setminus E) = m^*(E) + m^*(U \setminus E)$$

so

$$m^*(U \setminus E) < \epsilon.$$

This means that there is an open set $V \supset (U \setminus E)$ with $m(V) < \epsilon$. But we know that $U \setminus V$ is measurable in the sense of Lebesgue, since U and V are, and

$$m(U) \leq m(V) + m(U \setminus V)$$

so

$$m(U \setminus V) > m(U) - \epsilon.$$

So there is a closed set $F \subset U \setminus V$ with $m(F) > m(U) - \epsilon$. But since $V \supset U \setminus E$, we have $U \setminus V \subset E$. So $F \subset E$. So $F \subset E \subset U$ and

$$m(U \setminus F) = m(U) - m(F) < \epsilon.$$

Hence E is measurable in the sense of Lebesgue.

If $m(E) = \infty$, we must show that $E \cap [-n, n]$ is measurable in the sense of Caratheodory, for then it is measurable in the sense of Lebesgue from what we already know. We know that the interval $[-n, n]$ itself, being measurable in the sense of Lebesgue, is measurable in the sense of Caratheodory. So we will have completed the proof of the theorem if we show that the intersection of E with $[-n, n]$ is measurable in the sense of Caratheodory.

More generally, we will show that the union or intersection of two sets which are measurable in the sense of Caratheodory is again measurable in the sense of Caratheodory. Notice that the definition (4.8) is symmetric in E and E^c so if E is measurable in the sense of Caratheodory so is E^c. So it suffices to prove the next lemma to complete the proof.

Lemma 4.4.1 *If E_1 and E_2 are measurable in the sense of Caratheodory so is $E_1 \cup E_2$.*

For any set A we have

$$m^*(A) = m^*(A \cap E_1) + m^*(A \cap E_1^c)$$

by (4.8) applied to E_1. Applying (4.8) to $A \cap E_1^c$ and E_2 gives

$$m^*(A \cap E_1^c) = m^*(A \cap E_1^c \cap E_2) + m^*(A \cap E_1^c \cap E_2^c).$$

Substituting this back into the preceding equation gives

$$m^*(A) = m^*(A \cap E_1) + m^*(A \cap E_1^c \cap E_2) + m^*(A \cap E_1^c \cap E_2^c). \qquad (4.10)$$

Since $E_1^c \cap E_2^c = (E_1 \cup E_2)^c$ we can write this as

$$m^*(A) = m^*(A \cap E_1) + m^*(A \cap E_1^c \cap E_2) + m^*(A \cap (E_1 \cup E_2)^c).$$

Now $A \cap (E_1 \cup E_2) = (A \cap E_1) \cup (A \cap (E_1^c \cap E_2)$ so

$$m^*(A \cap E_1) + m^*(A \cap E_1^c \cap E_2) \geq m^*(A \cap (E_1 \cup E_2)).$$

Substituting this for the two terms on the right of the previous displayed equation gives

$$m^*(A) \geq m^*(A \cap (E_1 \cup E_2)) + m^*(A \cap (E_1 \cup E_2)^c)$$

which is just (4.9) for the set $E_1 \cup E_2$. This proves the lemma and the theorem.

We let $\mathcal{M}$ denote the class of measurable subsets of $\mathbf{R}$ - "measurability" in the sense of Lebesgue or Caratheodory these being equivalent. Notice by induction starting with two terms as in the lemma, that any finite union of sets in $\mathcal{M}$ is again in $\mathcal{M}$

4.5   Countable additivity.

The first main theorem in the subject is the following description of $\mathcal{M}$ and the function m on it:

Theorem 4.5.1 *$\mathcal{M}$ and the function $m : \mathcal{M} \to \mathbf{R}$ have the following properties:*

- $\mathbf{R} \in \mathcal{M}$.

- $E \in \mathcal{M} \Rightarrow E^c \in \mathcal{M}$.

- *If $E_n \in \mathcal{M}$ for $n = 1, 2, 3, \ldots$ then $\bigcup_n E_n \in \mathcal{M}$.*

- *If $F_n \in \mathcal{M}$ and the F_n are pairwise disjoint, then $F := \bigcup_n F_n \in \mathcal{M}$ and*

$$m(F) = \sum_{n=1}^{\infty} m(F_n).$$

Proof.   We already know the first two items on the list, and we know that a finite union of sets in $\mathcal{M}$ is again in $\mathcal{M}$. We also know the last assertion which is Proposition 4.3.1. But it will be instructive and useful for us to have a proof starting directly from Caratheodory's definition of measurablity:

If $F_1 \in \mathcal{M}$, $F_2 \in \mathcal{M}$ and $F_1 \cap F_2 = \emptyset$ then taking

$$A = F_1 \cup F_2, \quad E_1 = F_1, \quad E_2 = F_2$$

in (4.10) gives

$$m(F_1 \cup F_2) = m(F_1) + m(F_2).$$

Induction then shows that if $F_1, \ldots, F_n$ are pairwise disjoint elements of $\mathcal{M}$ then their union belongs to $\mathcal{M}$ and

$$m(F_1 \cup F_2 \cup \cdots \cup F_n) = m(F_1) + m(F_2) + \cdots + m(F_n).$$

More generally, if we let A be arbitrary and take $E_1 = F_1,\ E_2 = F_2$ in (4.10) we get

$$m^*(A) = m^*(A \cap F_1) + m^*(A \cap F_2) + m^*(A \cap (F_1 \cup F_2)^c).$$

If $F_3 \in \mathcal{M}$ is disjoint from F_1 and F_2 we may apply (4.8) with A replaced by $A \cap (F_1 \cup F_2)^c$ and E by F_3 to get

$$m^*(A \cap (F_1 \cup F_2)^c)) = m^*(A \cap F_3) + m^*(A \cap (F_1 \cup F_2 \cup F_3)^c),$$

since

$$(F_1 \cup F_2)^c \cap F_3^c = F_1^c \cap F_2^c \cap F_3^c = (F_1 \cup F_2 \cup F_3)^c.$$

Substituting this back into the preceding equation gives

$$m^*(A) = m^*(A \cap F_1) + m^*(A \cap F_2) + m^*(A \cap F_3) + m^*(A \cap (F_1 \cup F_2 \cup F_3)^c).$$

Proceeding inductively, we conclude that if $F_1, \ldots, F_n$ are pairwise disjoint elements of $\mathcal{M}$ then

$$m^*(A) = \sum_1^n m^*(A \cap F_i) + m^*(A \cap (F_1 \cup \cdots \cup F_n)^c). \tag{4.11}$$

Now suppose that we have a countable family $\{F_i\}$ of pairwise disjoint sets belonging to $\mathcal{M}$. Since

$$\left(\bigcup_{i=1}^n F_i \right)^c \supset \left(\bigcup_{i=1}^\infty F_i \right)^c$$

we conclude from (4.11) that

$$m^*(A) \geq \sum_1^n m^*(A \cap F_i) + m^* \left(A \cap \left(\bigcup_{i=1}^\infty F_i \right)^c \right)$$

and hence passing to the limit

$$m^*(A) \geq \sum_1^\infty m^*(A \cap F_i) + m^* \left(A \cap \left(\bigcup_{i=1}^\infty F_i \right)^c \right).$$

Now given any collection of sets B_k we can find intervals $\{I_{k,j}\}$ with

$$B_k \subset \bigcup_j I_{k,j}$$

and

$$m^*(B_k) \leq \sum_j \ell(I_{k,j}) + \frac{\epsilon}{2^k}.$$

So

$$\bigcup_k B_k \subset \bigcup_{k,j} I_{k,j}$$

and hence

$$m^* \left(\bigcup B_k \right) \leq \sum m^*(B_k),$$

the inequality being trivially true if the sum on the right is infinite. So

$$\sum_{i=1}^\infty m^*(A \cap F_k) \geq m^* \left(A \cap \left(\bigcup_{i=1}^\infty F_i \right) \right).$$

Thus

$$m^*(A) \geq \sum_1^\infty m^*(A \cap F_i) + m^* \left(A \cap \left(\bigcup_{i=1}^\infty F_i \right)^c \right) \geq$$

$$\geq m^* \left(A \cap \left(\bigcup_{i=1}^\infty F_i \right) \right) + m^* \left(A \cap \left(\bigcup_{i=1}^\infty F_i \right)^c \right).$$

The extreme right of this inequality is the left hand side of (4.9) applied to

$$E = \bigcup_i F_i,$$

and so $E \in \mathcal{M}$ and the preceding string of inequalities must be equalities since the middle is trapped between both sides which must be equal. Hence we have proved that if F_n is a disjoint countable family of sets belonging to $\mathcal{M}$ then their union belongs to $\mathcal{M}$ and

$$m^*(A) = \sum_i m^*(A \cap F_i) + m^* \left(A \cap \left(\bigcup_{i=1}^\infty F_i \right)^c \right). \tag{4.12}$$

If we take $A = \bigcup F_i$ we conclude that

$$m(F) = \sum_{n=1}^\infty m(F_n) \tag{4.13}$$

if the F_j are disjoint and

$$F = \bigcup F_j.$$

So we have reproved the last assertion of the theorem using Caratheodory's definition. For the third assertion, we need only observe that a countable union of sets in $\mathcal{M}$ can be always written as a countable disjoint union of sets in $\mathcal{M}$. Indeed, set

$$F_1 := E_1, \ F_2 := E_2 \setminus E_1 = E_1 \cap E_2^c$$

$$F_3 := E_3 \setminus (E_1 \cup E_2)$$

etc. The right hand sides all belong to $\mathcal{M}$ since $\mathcal{M}$ is closed under taking complements and finite unions and hence intersections, and

$$\bigcup_j F_j = \bigcup E_j.$$

We have completed the proof of the theorem.

A number of easy consequences follow: The symmetric difference between two sets is the set of points belonging to one or the other but not both:

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

Proposition 4.5.1 *If $A \in \mathcal{M}$ and $m(A \Delta B) = 0$ then $B \in \mathcal{M}$ and $m(A) = m(B)$.*

Proof. By assumption $A \setminus B$ has measure zero (and hence is measurable) since it is contained in the set $A \Delta B$ which is assumed to have measure zero. Similarly for $B \setminus A$. Also $(A \cap B) \in \mathcal{M}$ since

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

Thus

$$B = (A \cap B) \cup (B \setminus A) \in \mathcal{M}.$$

Since $B \setminus A$ and $A \cap B$ are disjoint, we have

$$m(B) = m(A \cap B) + m(B \setminus A) = m(A \cap B) = m(A \cap B) + m(A \setminus B) = m(A).$$

QED

Proposition 4.5.2 *Suppose that $A_n \in \mathcal{M}$ and $A_n \subset A_{n+1}$ for $n = 1, 2, \ldots$. Then*

$$m\left(\bigcup A_n\right) = \lim_{n \to \infty} m(A_n).$$

Indeed, setting $B_n := A_n \setminus A_{n-1}$ (with $B_1 = A_1$) the B_i are pairwise disjoint and have the same union as the A_i so

$$m\left(\bigcup A_n\right) = \sum_{i=1}^{\infty} m(B_i) = \lim_{n \to \infty} \sum_{i=1}^{n} m(B_n) = \lim_{n \to \infty} m\left(\bigcup_{i=1}^{n} B_i\right) = \lim_{n \to \infty} m(A_n).$$

QED

Proposition 4.5.3 *If $C_n \supset C_{n+1}$ is a decreasing family of sets in $\mathcal{M}$ and $m(C_1) < \infty$ then*

$$m\left(\bigcap C_n\right) = \lim_{n \to \infty} m(C_n).$$

Indeed, set $A_1 := \emptyset$, $A_2 := C_1 \setminus C_2$, $A_3 := C_1 \setminus C_3$ etc. The A's are increasing so

$$m\left(\bigcup (C_1 \setminus C_i)\right) = \lim_{n \to \infty} m(C_1 \setminus C_n) = m(C_1) - \lim_{n \to \infty} m(C_n)$$

by the preceding proposition. Since $m(C_1) < \infty$ we have

$$m(C_1 \setminus C_n) = m(C_1) - m(C_n).$$

Also

$$\bigcup_n (C_1 \setminus C_n) = C_1 \setminus \left(\bigcap_n C_n\right).$$

So

$$m\left(\bigcup_n (C_1 \setminus C_n)\right) = m(C_1) - m\left(\bigcap C_n\right) = m(C_1) - \lim_{n \to \infty} m(C_n).$$

Subtracting $m(C_1)$ from both sides of the last equation gives the equality in the proposition. QED

4.6 σ-fields, measures, and outer measures.

We will now take the items in Theorem 4.5.1 as axioms: Let X be a set. (Usually X will be a topological space or even a metric space). A collection $\mathcal{F}$ of subsets of X is called a σ field if:

- $X \in \mathcal{F}$,

- If $E \in \mathcal{F}$ then $E^c = X \setminus E \in \mathcal{F}$, and

- If $\{E_n\}$ is a sequence of elements in $\mathcal{F}$ then $\bigcup_n E_n \in \mathcal{F}$,

The intersection of any family of σ-fields is again a σ-field, and hence given any collection $\mathcal{C}$ of subsets of X, there is a smallest σ-field $\mathcal{F}$ which contains it. Then $\mathcal{F}$ is called the σ-field **generated** by $\mathcal{C}$.

If X is a metric space, the σ-field generated by the collection of open sets is called the **Borel** σ-field, usually denoted by $\mathcal{B}$ or $\mathcal{B}(X)$ and a set belonging to $\mathcal{B}$ is called a **Borel set**.

Given a σ-field $\mathcal{F}$ a (non-negative) **measure** is a function

$$m : \mathcal{F} \to [0, \infty]$$

such that

- $m(\emptyset) = 0$ and

- **Countable additivity:** If F_n is a disjoint collection of sets in $\mathcal{F}$ then

$$m\left(\bigcup_n F_n\right) = \sum_n m(F_n).$$

In the countable additivity condition it is understood that both sides might be infinite.

An **outer measure** on a set X is a map m^* to $[0, \infty]$ defined on the collection of *all* subsets of X which satisfies

- $m(\emptyset) = 0$,

- **Monotonicity:** If $A \subset B$ then $m^*(A) \leq m^*(B)$, and

- **Countable subadditivity:** $m^* \left(\bigcup_n A_n \right) \leq \sum_n m^*(A_n)$.

Given an outer measure, m^*, we defined a set E to be **measurable** (relative to m^*) if
$$m^*(A) = m^*(A \cap E) + m^*(A \cap E^c)$$
for all sets A. Then Caratheodory's theorem that we proved in the preceding section asserts that the collection of measurable sets is a σ-field, and the restriction of m^* to the collection of measurable sets is a measure which we shall usually denote by m.

There is an unfortunate disagreement in terminology, in that many of the professionals, especially in geometric measure theory, use the term "measure" for what we have been calling "outer measure". However we will follow the above conventions which used to be the old fashioned standard.

An obvious task, given Caratheodory's theorem, is to look for ways of constructing outer measures.

4.7 Constructing outer measures, Method I.

Let $\mathcal{C}$ be a collection of sets which cover X. For any subset A of X let

$$\mathrm{ccc}(A)$$

denote the set of (finite or) **c**ountable **c**overs of A by sets belonging to $\mathcal{C}$. In other words, an element of $\mathrm{ccc}(A)$ is a finite or countable collection of elements of $\mathcal{C}$ whose union contains A.

Suppose we are given a function

$$\ell : \mathcal{C} \to [0, \infty].$$

Theorem 4.7.1 *There exists a unique outer measure m^* on X such that*

- $m^*(A) \leq \ell(A)$ *for all $A \in \mathcal{C}$ and*

- *If n^* is any outer measure satisfying the preceding condition then $n^*(A) \leq m^*(A)$ for all subsets A of X.*

This unique outer measure is given by

$$m^*(A) = \inf_{\mathcal{D} \in \mathrm{ccc}(A)} \sum_{D \in \mathcal{D}} \ell(D). \tag{4.14}$$

In other words, for each countable cover of A by elements of $\mathcal{C}$ we compute the sum above, and then minimize over all such covers of A.

If we had two outer measures satisfying both conditions then each would have to be $\leq$ the other, so the uniqueness is obvious.

To check that the m^* defined by (4.14) is an outer measure, observe that for the empty set we may take the empty cover, and the convention about an empty sum is that it is zero, so $m^*(\emptyset) = 0$. If $A \subset B$ then any cover of B is a cover of A, so that $m^*(A) \leq m^*(B)$. To check countable subadditivity we use the usual $\epsilon/2^n$ trick: If $m^*(A_n) = \infty$ for any A_n the subadditivity condition is obviously satisfied. Otherwise, we can find a $\mathcal{D}_n \in \mathrm{ccc}(A_n)$ with

$$\sum_{D \in \mathcal{D}_n} \ell(D) \leq m^*(A_n) + \frac{\epsilon}{2^n}.$$

Then we can collect all the D together into a countable cover of A so

$$m^*(A) \leq \sum_n m^*(A_n) + \epsilon,$$

and since this is true for all $\epsilon > 0$ we conclude that m^* is countably subadditive. So we have verified that m^* defined by (4.14) is an outer measure. We must check that it satisfies the two conditions in the theorem. If $A \in \mathcal{C}$ then the single element collection $\{A\} \in \mathrm{ccc}(A)$, so $m^*(A) \leq \ell(A)$, so the first condition is obvious. As to the second condition, suppose n^* is an outer measure with $n^*(D) \leq \ell(D)$ for all $D \in \mathcal{C}$. Then for any set A and any countable cover $\mathcal{D}$ of A by elements of $\mathcal{C}$ we have

$$\sum_{D \in \mathcal{D}} \ell(D) \geq \sum_{D \in \mathcal{D}} n^*(D) \geq n^*\left(\bigcup_{D \in \mathcal{D}} D\right) \geq n^*(A),$$

where in the second inequality we used the countable subadditivity of n^* and in the last inequality we used the monotonicity of n^*. Minimizing over all $\mathcal{D} \in \mathrm{ccc}(A)$ shows that $m^*(A) \geq n^*(A)$. QED

This argument is basically a repeat performance of the construction of Lebesgue measure we did above. However there is some trouble:

4.7.1   A pathological example.

Suppose we take $X = \mathbf{R}$, and let $\mathcal{C}$ consist of all *half open* intervals of the form $[a, b)$. However, instead of taking ℓ to be the length of the interval, we take it to be the square root of the length:

$$\ell([a, b)) := (b - a)^{\frac{1}{2}}.$$

I claim that any half open interval (say $[0,1)$) of length one has $m^*([a,b)) = 1$. (Since ℓ is translation invariant, it does not matter which interval we choose.) Indeed, $m^*([0,1)) \leq 1$ by the first condition in the theorem, since $\ell([0,1)) = 1$. On the other hand, if

$$[0,1) \subset \bigcup_i [a_i, b_i)$$

then we know from the Heine-Borel argument that

$$\sum (b_i - a_i) \geq 1,$$

so squaring gives

$$\left(\sum (b_i - a_i)^{\frac{1}{2}}\right)^2 = \sum_i (b_i - a_i) + \sum_{i \neq j} (b_i - a_i)^{\frac{1}{2}} (b_j - a_j)^{\frac{1}{2}} \geq 1.$$

So $m^*([0,1)) = 1$.

On the other hand, consider an interval $[a,b)$ of length 2. Since it covers itself, $m^*([a,b)) \leq \sqrt{2}$.

Consider the closed interval $I = [0,1]$. Then

$$I \cap [-1,1) = [0,1) \quad \text{and} \quad I^c \cap [-1,1) = [-1,0)$$

so

$$m^*(I \cap [-1,1)) + m^*(I^c \cap [-1,1) = 2 > \sqrt{2} \geq m^*([-1,1)).$$

In other words, the closed unit interval is *not measurable* relative to the outer measure m^* determined by the theorem. We would like Borel sets to be measurable, and the above computation shows that the measure produced by Method I as above does not have this desirable property. In fact, if we consider two half open intervals I_1 and I_2 of length one separated by a small distance of size ϵ, say, then their union $I_1 \cup I_2$ is covered by an interval of length $2 + \epsilon$, and hence

$$m^*(I_1 \cup I_2) \leq \sqrt{2 + \epsilon} < m^*(I_1) + m^*(I_2).$$

In other words, m^* is not additive even on intervals separated by a finite distance. It turns out that this is the crucial property that is missing:

4.7.2 Metric outer measures.

Let X be a metric space. An outer measure on X is called a **metric outer measure** if

$$m^*(A \cup B) = m^*(A) + m^*(B) \quad \text{whenever } d(A,B) > 0. \qquad (4.15)$$

The condition $d(A,B) > 0$ means that there is an $\epsilon > 0$ (depending on A and B) so that $d(x.y) > \epsilon$ for all $x \in A$, $y \in B$. The main result here is due to Caratheodory:

Theorem 4.7.2 *If m^* is a metric outer measure on a metric space X, then all Borel sets of X are m^* measurable.*

Proof. Since the σ-field of Borel sets is generated by the closed sets, it is enough to prove that every closed set F is measurable in the sense of Caratheodory, i.e. that for any set A

$$m^*(A) \geq m^*(A \cap F) + m^*(A \setminus F).$$

Let

$$A_j := \{x \in A | d(x, F) \geq \frac{1}{j}\}.$$

We have $d(A_j, A \cap F) \geq 1/j$ so, since m^* is a metric outer measure, we have

$$m^*(A \cap F) + m^*(A_j) = m^*((A \cap F) \cup A_j) \leq m^*(A) \qquad (4.16)$$

since $(A \cap F) \cup A_j \subset A$. Now

$$A \setminus F = \bigcup A_j$$

since F is closed, and hence every point of A not belonging to F must be at a positive distance from F. We would like to be able to pass to the limit in (4.16). If the limit on the left is infinite, there is nothing to prove. So we may assume it is finite.

Now if $x \in A \setminus (F \cup A_{j+1})$ there is a $z \in F$ with $d(x, z) < 1/(j + 1)$ while if $y \in A_j$ we have $d(y, z) \geq 1/j$ so

$$d(x, y) \geq d(y, z) - d(x, z) \geq \frac{1}{j} - \frac{1}{j+1} > 0.$$

Let $B_1 := A_1$ and $B_2 := A_2 \setminus A_1$, $B_3 = A_3 \setminus A_2$ etc. Thus if $i \geq j + 2$, then $B_j \subset A_j$ and

$$B_i \subset A \setminus (F \cup A_{i-1}) \subset A \setminus (F \cup A_{j+1})$$

and so $d(B_i, B_j) > 0$. So m^* is additive on finite unions of even or odd B's:

$$m^* \left(\bigcup_{k=1}^{n} B_{2k-1} \right) = \sum_{k=1}^{n} m^*(B_{2k-1}), \quad m^* \left(\bigcup_{k=1}^{n} B_{2k} \right) = \sum_{k=1}^{n} m^*(B_{2k}).$$

Both of these are $\leq m^*(A_{2n})$ since the union of the sets involved are contained in A_{2n}. Since $m^*(A_{2n})$ is increasing, and assumed bounded, both of the above

series converge. Thus

$$
\begin{aligned}
m^*(A/F) \;&=\; m^*\left(\bigcup A_i\right) \\
&=\; m^*\left(A_j \cup \bigcup_{k \ge j+1} B_j\right) \\
&\le\; m^*(A_j) + \sum_{k=j+1}^{\infty} m^*(B_j) \\
&\le\; \lim_{n \to \infty} m^*(A_n) + \sum_{k=j+1}^{\infty} m^*(B_j).
\end{aligned}
$$

But the sum on the right can be made as small as possible by choosing j large, since the series converges. Hence

$$
m^*(A/F) \le \lim_{n \to \infty} m^*(A_n)
$$

QED.

4.8 Constructing outer measures, Method II.

Let $\mathcal{C} \subset \mathcal{E}$ be two covers, and suppose that ℓ is defined on $\mathcal{E}$, and hence, by restriction, on $\mathcal{C}$. In the definition (4.14) of the outer measure $m^*_{\ell,\mathcal{C}}$ associated to ℓ and $\mathcal{C}$, we are minimizing over a smaller collection of covers than in computing the metric outer measure $m^*_{\ell,\mathcal{E}}$ using all the sets of $\mathcal{E}$. Hence

$$
m^*_{\ell,\mathcal{C}}(A) \ge m^*_{\ell,\mathcal{E}}(A)
$$

for any set A.

We want to apply this remark to the case where X is a metric space, and we have a cover $\mathcal{C}$ with the property that for every $x \in X$ and every $\epsilon > 0$ there is a $C \in \mathcal{C}$ with $x \in C$ and $\mathrm{diam}(C) < \epsilon$. In other words, we are assuming that the

$$
\mathcal{C}_\epsilon := \{C \in \mathcal{C}| \ \mathrm{diam}\ (C) < \epsilon\}
$$

are covers of X for every $\epsilon > 0$. Then for every set A the

$$
m_{\ell,\mathcal{C}_\epsilon}(A)
$$

are increasing, so we can consider the function on sets given by

$$
m^*_{II}(A) := \sup_{\epsilon \to 0} m_{\ell,\mathcal{C}_\epsilon}(A).
$$

The axioms for an outer measure are preserved by this limit operation, so m^*_{II} is an outer measure. If A and B are such that $d(A,B) > 2\epsilon$, then any set of $\mathcal{C}_\epsilon$ which intersects A does not intersect B and vice versa, so throwing away extraneous sets in a cover of $A \cup B$ which does not intersect either, we see that $m^*_{II}(A \cup B) = m^*_{II}(A) + m^*_{II}(B)$. The method II construction always yields a metric outer measure.

4.8.1   An example.

Let X be the set of all (one sided) infinite sequences of 0's and 1's. So a point of X is an expression of the form

$$a_1 a_2 a_3 \cdots$$

where each a_i is 0 or 1. For any finite sequence α of 0's or 1's, let $[\alpha]$ denote the set of all sequences which begin with α. We also let $|\alpha|$ denote the length of α, that is, the number of bits in α. For each

$$0 < r < 1$$

we define a metric d_r on X by: If

$$x = \alpha x', \qquad y = \alpha y'$$

where the first bit in x' is different from the first bit in y' then

$$d_r(x, y) := r^{|\alpha|}.$$

In other words, the distance between two sequence is r^k where k is the length of the longest initial segment where they agree. Clearly $d_r(x, y) \geq 0$ and $= 0$ if and only if $x = y$, and $d_r(y, x) = d_r(x, y)$. Also, for three $x, y,$ and z we claim that

$$d_r(x, z) \leq \max\{d_r(x, y), d_r(y, z)\}.$$

Indeed, if two of the three points are equal this is obvious. Otherwise, let j denote the length of the longest common prefix of x and y, and let k denote the length of the longest common prefix of y and z. Let $m = \min(j, k)$. Then the first m bits of x agree with the first m bits of z and so $d_r(x, z) \leq r^m = \max(r^j, r^k)$. QED

A metric with this property (which is much stronger than the triangle inequality) is called an **ultrametric**.

Notice that

$$\text{diam } [\alpha] = r^\alpha. \tag{4.17}$$

The metrics for different r are different, and we will make use of this fact shortly. But

Proposition 4.8.1 *The spaces (X, d_r) are all homeomorphic under the identity map.*

It is enough to show that the identity map is a continuous map from (X, d_r) to (X, d_s) since it is one to one and we can interchange the role of r and s. So, given $\epsilon > 0$, we must find a $\delta > 0$ such that if $d_r(x, y) < \delta$ then $d_s(x, y) < \epsilon$. So choose k so that $s^k < \epsilon$. Then letting $r^k = \delta$ will do.

So although the metrics are different, the topologies they define are the same.

There is something special about the value $r = \frac{1}{2}$: Let $\mathcal{C}$ be the collection of all sets of the form $[\alpha]$ and let ℓ be defined on $\mathcal{C}$ by

$$\ell([\alpha]) = (\frac{1}{2})^{|\alpha|}.$$

We can construct the method II outer measure associated with this function, which will satisfy

$$m_{II}^*([\alpha]) \geq m_I^*([\alpha])$$

where m_I^* denotes the method I outer measure associated with ℓ. What is special about the value $\frac{1}{2}$ is that if $k = |\alpha|$ then

$$\ell([\alpha]) = \left(\frac{1}{2}\right)^k = \left(\frac{1}{2}\right)^{k+1} + \left(\frac{1}{2}\right)^{k+1} = \ell([\alpha 0]) + \ell([\alpha 1]).$$

So if we also use the metric $d_{\frac{1}{2}}$, we see, by repeating the above, that every $[\alpha]$ can be written as the disjoint union $C_1 \cup \cdots \cup C_n$ of sets in $\mathcal{C}_\epsilon$ with $\ell([\alpha])) = \sum \ell(C_i)$. Thus $m_{\ell,\mathcal{C}_\epsilon}^*([\alpha]) \leq \ell(\alpha)$ and so $m_{\ell,\mathcal{C}_\epsilon}^*([\alpha])(A) \leq m_I^*(A)$ or $m_{II}^* = m_I^*$. It also follows from the above computation that

$$m^*([\alpha]) = \ell([\alpha]).$$

There is also something special about the value $s = \frac{1}{3}$: Recall that one of the definitions of the Cantor set $\mathbf{C}$ is that it consists of all points $x \in [0, 1]$ which have a base 3 expansion involving only the symbols 0 and 2. Let

$$h : X \to \mathbf{C}$$

where h sends the bit 1 into the symbol 2, e.g.

$$h(011001\ldots) = .022002\ldots.$$

In other words, for any sequence z

$$h(0z) = \frac{h(z)}{3}, \qquad h(1z) = \frac{h(z) + 2}{3}. \tag{4.18}$$

I claim that:

$$\frac{1}{3}d_{\frac{1}{3}}(x, y) \leq |h(x) - h(y)| \leq d_{\frac{1}{3}}(x, y) \tag{4.19}$$

Proof. If x and y start with different bits, say $x = 0x'$ and $y = 1y'$ then $d_{\frac{1}{3}}(x, y) = 1$ while $h(x)$ lies in the interval $[0, \frac{1}{3}]$ and $h(y)$ lies in the interval $[\frac{2}{3}, 1]$ on the real line. So $h(x)$ and $h(y)$ are at least a distance $\frac{1}{3}$ and at most a distance 1 apart, which is what (4.19) says. So we proceed by induction. Suppose we know that (4.19) is true when $x = \alpha x'$ and $y = \alpha y'$ with x', y' starting with different digits, and $|\alpha| \leq n$. (The above case was where $|\alpha| = 0$.)

So if $|\alpha| = n + 1$ then either and $\alpha = 0\beta$ or $\alpha = 1\beta$ and the argument for either case is similar: We know that (4.19) holds for $\beta x'$ and $\beta y'$ and

$$d_{\frac{1}{3}}(x, y) = \frac{1}{3} d_{\frac{1}{3}}(\beta x', \beta y')$$

while $|h(x) - h(y)| = \frac{1}{3}|h(\beta x') - h(\beta y')|$ by (4.18). Hence (4.19) holds by induction. QED

In other words, the map h is a Lipschitz map with Lipschitz inverse from $(X, d_{\frac{1}{3}})$ to the Cantor set **C**.

In a short while, after making the appropriate definitions, these two computations, one with the measure associated to $\ell([\alpha]) = \left(\frac{1}{2}\right)^{|\alpha|}$ and the other associated with $d_{\frac{1}{3}}$ will show that the "Hausdorff dimension" of the Cantor set is $\log 2 / \log 3$.

4.9   Hausdorff measure.

Let X be a metric space. Recall that if A is any subset of X, the **diameter** of A is defined as

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

Take $\mathcal{C}$ to be the collection of all subsets of X, and for any positive real number s define

$$\ell_s(A) = \mathrm{diam}(A)^s$$

(with $0^s = 0$). Take $\mathcal{C}$ to consist of all subsets of X. The method II outer measure is called the s-**dimensional Hausdorff outer measure**, and its restriction to the associated σ-field of (Caratheodory) measurable sets is called the s-**dimensional Hausdorff measure**. We will let $m_{s,\epsilon}^*$ denote the method I outer measure associated to ℓ_s and ϵ, and let $\mathcal{H}_s^*$ denote the Hausdorff outer measure of dimension s, so that

$$\mathcal{H}_s^*(A) = \lim_{\epsilon \to 0} m_{s,\epsilon}^*(A).$$

For example, we claim that for $X = \mathbf{R}$, $\mathcal{H}^1$ is exactly Lebesgue outer measure, which we will denote here by L^*. Indeed, if A has diameter r, then A is contained in a closed interval of length r. Hence $L^*(A) \le r$. The Method I construction theorem says that $m_{1,\epsilon}^*$ is the largest outer measure satisfying $m^*(A) \le \mathrm{diam}\, A$ for sets of diameter less than ϵ. Hence $m_{1,\epsilon}^*(A) \ge L^*(A)$ for all sets A and all ϵ, and so

$$\mathcal{H}_1^* \ge L^*.$$

On the other hand, any bounded half open interval $[a, b)$ can be broken up into a finite union of half open intervals of length $< \epsilon$, whose sum of diameters is $b - a$. So $m_{1,\epsilon}^*([a, b) \le b - a$. But the method I construction theorem says that L^* is the largest outer measure satisfying

$$m^*([a, b)) \le b - a.$$

Hence $\mathcal{H}_1^* \leq L^*$ So they are equal.

In two or more dimensions, the Hausdorff measure $\mathcal{H}_k$ on $\mathbf{R}^k$ differs from Lebesgue measure by a constant. This is essentially because they assign different values to the ball of diameter one. In two dimensions for example, the Hausdorff measure $\mathcal{H}_2$ assigns the value one to the disk of diameter one, while its Lebesgue measure is $\pi/4$. For this reason, some authors prefer to put this "correction factor" into the definition of the Hausdorff measure, which would involve the Gamma function for non-integral s. I am following the convention that finds it simpler to drop this factor.

Theorem 4.9.1 *Let $F \subset X$ be a Borel set. Let $0 < s < t$. Then*

$$\mathcal{H}_s(F) < \infty \ \Rightarrow \mathcal{H}_t(F) = 0$$

and

$$\mathcal{H}_t(F) > 0 \ \ \Rightarrow \mathcal{H}_s(F) = \infty.$$

Indeed, if diam $A \leq \epsilon$, then

$$m_{t,\epsilon}^*(A) \leq (\operatorname{diam} A)^t \leq \epsilon^{t-s}(\operatorname{diam} A)^s$$

so by the method I construction theorem we have

$$m_{t,\epsilon}^*(B) \leq \epsilon^{t-s} m_{s,\epsilon}^*(B)$$

for all B. If we take $B = F$ in this equality, then the assumption $\mathcal{H}_s(F) < \infty$ implies that the limit of the right hand side tends to 0 as $\epsilon \to 0$, so $\mathcal{H}_t(F) = 0$. The second assertion in the theorem is the contrapositive of the first.

4.10   Hausdorff dimension.

This last theorem implies that for any Borel set F, there is a unique value s_0 (which might be 0 or ∞) such that $\mathcal{H}_t(F) = \infty$ for all $t < s_0$ and $\mathcal{H}_s(F) = 0$ for all for all $s > s_0$. This value is called the **Hausdorff dimension** of F. It is one of many competing (and non-equivalent) definitions of dimension. Notice that it is a metric invariant, and in fact is the same for two spaces different by a Lipschitz homeomorphism with Lipschitz inverse. But it is not a topological invariant. In fact, we shall show that the space X of all sequences of zeros and one studied above has Hausdorff dimension 1 relative to the metric $d_{\frac{1}{2}}$ while it has Hausdorff dimension $\log 2/\log 3$ if we use the metric $d_{\frac{1}{3}}$. Since we have shown that $(X, d_{\frac{1}{3}})$ is Lipschitz equivalent to the Cantor set $\mathbf{C}$, this will also prove that $\mathbf{C}$ has Hausdorff dimension $\log 2/\log 3$.

We first discuss the $d_{\frac{1}{2}}$ case and use the following lemma

Lemma 4.10.1 *If $\operatorname{diam}(A) > 0$, then there is an α such that $A \subset [\alpha]$ and $\operatorname{diam}([\alpha]) = \operatorname{diam} A$.*

Proof. Given any set A, it has a "longest common prefix". Indeed, consider the set of lengths of common prefixes of elements of A. This is finite set of non-negative integers since A has at least two distinct elements. Let n be the largest of these, and let α be a common prefix of this length. Then it is clearly the longest common prefix of A. Hence $A \subset [\alpha]$ and $\mathrm{diam}([\alpha]) = \mathrm{diam}\,A$.QED

Let $\mathcal{C}$ denote the collection of all sets of the form $[\alpha]$ and let ℓ be the function on $\mathcal{C}$ given by

$$\ell([\alpha]) = (\frac{1}{2})^{|\alpha|},$$

and let ℓ^* be the associated method I outer measure, and m the associated measure; all these as we introduced above. We have

$$\ell^*(A) \le \ell^*([\alpha]) = \mathrm{diam}([\alpha]) = \mathrm{diam}(A).$$

By the method I construction theorem, $m_{1,\epsilon}^*$ is the largest outer measure with the property that $n^*(A) \le \mathrm{diam}\,A$ for sets of diameter $< \epsilon$. Hence $\ell^* \le m_{1,\epsilon}^*$, and since this is true for all $\epsilon > 0$, we conclude that

$$\ell^* \le \mathcal{H}_1^*.$$

On the other hand, for any α and any $\epsilon > 0$, there is an n such that $2^{-n} < \epsilon$ and $n \ge |\alpha|$. The set $[\alpha]$ is the disjoint union of all sets $[\beta] \subset [\alpha]$ with $|\beta| \ge n$, and there are $2^{n-|\alpha|}$ of these subsets, each having diameter 2^{-n}. So

$$m_{1,\epsilon}^*([\alpha]) \le 2^{-|\alpha|}.$$

However ℓ^* is the largest outer measure satisfying this inequality for all $[\alpha]$. Hence $m_{1,\epsilon}^* \le \ell^*$ for all ϵ so $\mathcal{H}_1^* \le \ell^*$. In other words

$$\mathcal{H}_1 = m.$$

But since we computed that $m(X) = 1$, we conclude that

The Hausdorff dimension of $(X, d_{\frac{1}{2}})$ is 1.

Now let us turn to $(X, d_{\frac{1}{3}})$. Then the diameter $\mathrm{diam}_{\frac{1}{2}}$ relative to the metric $d_{\frac{1}{2}}$ and the diameter $\mathrm{diam}_{\frac{1}{3}}$ relative to the metric $d_{\frac{1}{3}}$ are given by

$$\mathrm{diam}_{\frac{1}{2}}([\alpha]) = \left(\frac{1}{2}\right)^k, \quad \mathrm{diam}_{\frac{1}{3}}([\alpha]) = \left(\frac{1}{3}\right)^k, \quad k = |\alpha|.$$

If we choose s so that $2^{-k} = (3^{-k})^s$ then

$$\mathrm{diam}_{\frac{1}{2}}([\alpha]) = (\mathrm{diam}_{\frac{1}{3}}([\alpha]))^s.$$

This says that relative to the metric $d_{\frac{1}{3}}$, the previous computation yields

$$\mathcal{H}_s(X) = 1.$$

Hence $s = \log 2 / \log 3$ is the Hausdorff dimension of X.

The material above (with some slight changes in notation) was taken from the book *Measure, Topology, and Fractal Geometry* by Gerald Edgar, where a thorough and delightfully clear discussion can be found of the subjects listed in the title.

4.11 Push forward.

The above discussion is a sampling of introductory material to what is known as "geometric measure theory". However the construction of measures that we will be mainly working with will be an abstraction of the "simulation" approach that we have been developing in the problem sets. The setup is as follows: Let $(X, \mathcal{F}, m)$ be a set with a σ-field and a measure on it, and let $(Y, \mathcal{G})$ be some other set with a σ-field on it. A map

$$f : X \rightarrow Y$$

is called **measurable** if

$$f^{-1}(B) \in \mathcal{F} \quad \forall \, B \in \mathcal{G}.$$

We may then define a measure $f_*(m)$ on $(Y, \mathcal{G})$ by

$$(f_*)m(B) = m(f^{-1}(B)).$$

For example, if Y_λ is the Poisson random variable from the exercises, and u is the uniform measure (the restriction of Lebesgue measure to) on $[0, 1]$, then $f_*(u)$ is the measure on the non-negative integers given by

$$f_*(u)(\{k\}) = e^{-\lambda} \frac{\lambda^k}{k!}.$$

It will be this construction of measures and variants on it which will occupy us over the next few weeks.

4.12 The Hausdorff dimension of fractals

4.12.1 Similarity dimension.

Contracting ratio lists.

A finite collection of real numbers

$$(r_1, \ldots, r_n)$$

is called a **contracting ratio list** if

$$0 < r_i < 1 \quad \forall \, i = 1, \ldots, n.$$

Proposition 4.12.1 *Let $(r_1, \ldots, r_n)$ be a contracting ratio list. There exists a unique non-negative real number s such that*

$$\sum_{i=1}^{n} r_i^s = 1. \qquad (4.20)$$

The number s is 0 if and only if $n = 1$.

Proof. If $n = 1$ then $s = 0$ works and is clearly the only solution. If $n > 1$, define the function f on $[0, \infty)$ by

$$f(t) := \sum_{i=1}^{n} r_i^t.$$

We have

$$f(0) = n \quad \text{and} \quad \lim_{t \to \infty} f(t) = 0 < 1.$$

Since f is continuous, there is some postive solution to (4.20). To show that this solution is unique, it is enough to show that f is monotone decreasing. This follows from the fact that its derivative is

$$\sum_{i=1}^{n} r_i^t \log r_i < 0.$$

QED

Definition 4.12.1 *The number s in (4.20) is called the **similarity dimension** of the ratio list $(r_1, \ldots, r_n)$.*

Iterated function systems and fractals.

A map $f : X \to Y$ between two metric spaces is called a **similarity** with similarity ratio r if

$$d_Y(f(x_1), f(x_2)) = r d_X(x_1, x_2) \quad \forall\ x_1,\ x_2 \in X.$$

(Recall that a map is called **Lipschitz** with Lipschitz constant r if we only had an inequality, $\leq$, instead of an equality in the above.)

Let X be a complete metric space, and let $(r_1, \ldots, r_n)$ be a contracting ratio list. A collection

$$(f_1, \ldots, f_n), \quad f_i :\ X \to X$$

is called an **iterated function system** which **realizes** the contracting ratio list if

$$f_i : X \to X, \quad i = 1, \ldots, n$$

is a similarity with ratio r_i. We also say that $(f_1, \ldots, f_n)$ is a **realization** of the ratio list $(r_1, \ldots, r_n)$.

It is a consequence of *Hutchinson's theorem*, see below, that

Proposition 4.12.2 *If $(f_1, \ldots, f_n)$ is a realization of the contracting ratio list $(r_1, \ldots, r_n)$ on a complete metric space, X, then there exists a unique non-empty compact subset $K \subset X$ such that*

$$K = f_1(K) \cup \cdots \cup f_n(K).$$

In fact, Hutchinson's theorem asserts the corresponding result where the f_i are merely assumed to be Lipschitz maps with Lipschitz constants $(r_1, \ldots, r_n)$.

The set K is sometimes called the **fractal** associated with the realization $(f_1, \ldots, f_n)$ of the contracting ratio list $(r_1, \ldots, r_n)$. The facts we want to establish are: First,

$$\dim(K) \leq s \tag{4.21}$$

where dim denotes Hausdorff dimension, and s is the similarity dimension of $(r_1, \ldots, r_n)$. In general, we can only assert an inequality here, for the the set K does not fix $(r_1, \ldots, r_n)$ or its realization. For example, we can repeat some of the r_i and the corresponding f_i. This will give us a longer list, and hence a larger s, but will not change K. But we can demand a rather strong form of non-redundancy known as **Moran's condition**: There exists an open set O such that

$$O \supset f_i(O) \ \forall \ i \ \text{ and } \ f_i(O) \cap f_j(O) = \emptyset \ \ \forall \ i \neq j. \tag{4.22}$$

Then

Theorem 4.12.1 *If $(f_1, \ldots, f_n)$ is a realization of $(r_1, \ldots, r_n)$ on $\mathbf{R}^d$ and if Moran's condition holds then*

$$\dim K = s.$$

The method of proof of (4.21) will be to construct a "model" complete metric space E with a realization $(g_1, \ldots, g_n)$ of $(r_1, \ldots, r_n)$ on it, which is "universal" in the sense that

- E is itself the fractal associated to $(g_1, \ldots, g_n)$.

- The Hausdorff dimension of E is s.

- If $(f_1, \ldots, f_n)$ is a realization of $(r_1, \ldots, r_n)$ on a complete metric space X then there exists a unique continuous map

$$h : E \to X$$

  such that

$$h \circ g_i = f_i \circ h. \tag{4.23}$$

- The image $h(E)$ of h is K.

- The map h is Lipschitz.

This is clearly enough to prove (4.21). A little more work will then prove Moran's theorem.

4.12.2   The string model.

Construction of the model.

Let $(r_1, \ldots, r_n)$ be a contracting ratio list, and let $\mathcal{A}$ denote the alphabet consisting of the letters $\{1, \ldots, n\}$. Let E denote the space of one sided infinite strings of letters from the alphabet $\mathcal{A}$. If α denotes a finite string (word) of letters from $\mathcal{A}$, we let w_α denote the product over all i occurring in α of the r_i. Thus

$$w_\emptyset = 1$$

where $\emptyset$ is the empty string, and, inductively,

$$w_{\alpha e} = w_\alpha \cdot w_e, \quad e \in \mathcal{A}.$$

If $x \neq y$ are two elements of E, they will have a longest common initial string α, and we then define

$$d(x, y) := w_\alpha.$$

This makes E into a complete ultrametic space. Define the maps $g_i : E \to E$ by

$$g_i(x) = ix.$$

That is, g_i shifts the infinite string one unit to the right and inserts the letter i in the initial position. In terms of our metric, clearly $(g_1, \ldots, g_n)$ is a realization of $(r_1, \ldots, r_n)$ and the space E itself is the corresponding fractal set.

We let $[\alpha]$ denote the set of all strings beginning with α, i.e. whose first word (of length equal to the length of α) is α. The diameter of this set is w_α.

The Hausdorff dimension of E is s.

We begin with a lemma:

Lemma 4.12.1 *Let $A \subset E$ have positive diameter. Then there exists a word α such that $A \subset [\alpha]$ and*

$$\mathrm{diam}(A) = \mathrm{diam}[\alpha] = w_\alpha.$$

Proof. Since A has at least two elements, there will be a γ which is a prefix of one and not the other. So there will be an integer n (possibly zero) which is the length of the longest common prefix of all elements of A. Then every element of A will begin with this common prefix α which thus satisfies the conditions of the lemma. QED

The lemma implies that in computing the Hausdorff measure or dimension, we need only consider covers by sets of the form $[\alpha]$. Now if we choose s to be the solution of (4.20), then

$$(\mathrm{diam}[\alpha])^s = \sum_{i=1}^n (\mathrm{diam}[\alpha i])^s = (\mathrm{diam}[\alpha])^s \sum_{i=1}^n r_i^s.$$

This means that the method II outer measure assosicated to the function $A \mapsto (\operatorname{diam} A)^s$ coincides with the method I outer measure and assigns to each set $[\alpha]$ the measure w_α^s. In particular the measure of E is one, and so the Hausdorff dimension of E is s.

The universality of E.

Let $(f_1, \ldots, f_n)$ a realization of $(r_1, \ldots, r_n)$ on a complete metric space X. Choose a point $a \in X$ and define $h_0 : E \to X$ by

$$h_0(z) :\equiv a.$$

Inductively define the maps h_p by defining h_{p+1} on each of the open sets $[\{i\}]$ by

$$h_{p+1}(iz) := f_i(h_p(z)).$$

The sequence of maps $\{h_p\}$ is Cauchy in the uniform norm. Indeed, if $y \in [\{i\}]$ so $y = g_i(z)$ for some $z \in E$ then

$$d_X(h_{p+1}(y), h_p(y)) = d_X(f_i(h_p(z)), f_i(h_{p-1}(z))) = r_i d_X(h_p(z), h_{p-1}(z))).$$

So if we let $c := \max_i(r_i)$ so that $0 < c < 1$, we have

$$\sup_{y \in E} d_X(h_{p+1}(y), h_p(y)) \leq c \sup_{x \in E} d_X(h_p(x), h_{p-1}(x))$$

for $p \geq 1$ and hence

$$\sup_{y \in E} d_X(h_{p+1}(y), h_p(y)) < Cc^p$$

for a suitable constant C. This shows that the h_p converge uniformly to a limit h which satisfies

$$h \circ g_i = f_i \circ h.$$

Now

$$h_{k+1}(E) = \bigcup_i f_i(h_k(E)),$$

and the proof of Hutchinson's theorem given below - using the contraction fixed point theorem for compact sets under the Hausdorff metric - shows that the sequence of sets $h_k(E)$ converges to the fractal K.

Since the image of h is K which is compact, the image of $[\alpha]$ is $f_\alpha(K)$ where we are using the obvious notation $f_{ij} = f_i \circ f_j$, $f_{ijk} = f_i \circ f_j \circ f_k$ etc. The set $f_\alpha(K)$ has diameter $w_\alpha \cdot \operatorname{diam}(K)$. Thus h is Lipschitz with Lipschitz constant $\operatorname{diam}(K)$.

The uniqueness of the map h follows from the above sort of argument.

4.13   The Hausdorff metric and Hutchinson's theorem.

Let X be a complete metric space. Let $\mathcal{H}(X)$ denote the space of non-empty compact subsets of X. For any $A \in \mathcal{H}(X)$ and any positive number ϵ, let

$$A_\epsilon = \{x \in X | d(x, y) \leq \epsilon, \text{for some } y \in A\}.$$

We call A_ϵ the ϵ-*collar* of A. Recall that we defined

$$d(x, A) = \inf_{y \in A} d(x, y)$$

to be the distance from any $x \in X$ to A, then we can write the definition of the ϵ-collar as

$$A_\epsilon = \{x | d(x, A) \leq \epsilon\}.$$

Notice that the infimum in the definition of $d(x, A)$ is actually achieved, that is, there is some point $y \in A$ such that

$$d(x, A) = d(x, y).$$

This is because A is compact. For a pair of non-empty compact sets, A and B, define

$$d(A, B) = \max_{x \in A} d(x, B).$$

So

$$d(A, B) \leq \epsilon \quad \Leftrightarrow \quad A \subset B_\epsilon.$$

Notice that this condition is not symmetric in A and B. So Hausdorff introduced

$$
\begin{aligned}
h(A, B) &= \max\{d(A, B), d(B, A)\} & (4.24)\\
&= \inf\{\epsilon \mid A \subset B_\epsilon \text{ and } B \subset A_\epsilon\}. & (4.25)
\end{aligned}
$$

as a distance on $\mathcal{H}(X)$. He proved

Proposition 4.13.1 *The function h on $\mathcal{H}(X) \times \mathcal{H}(X)$ satsifies the axioms for a metric and makes $\mathcal{H}(X)$ into a complete metric space. Furthermore, if*

$$A, B, C, D \in \mathcal{H}(X)$$

then

$$h(A \cup B, C \cup D) \leq \max\{h(A, C), h(B, D)\}. \qquad (4.26)$$

Proof. We begin with (4.26). If ϵ is such that $A \subset C_\epsilon$ and $B \subset D_\epsilon$ then clearly $A \cup B \subset C_\epsilon \cup D_\epsilon = (C \cup D)_\epsilon$. Repeating this argument with the roles of A, C and B, D interchanged proves (4.26).

We prove that h is a metric: h is symmetric, by definition. Also, $h(A, A) = 0$, and if $h(A, B) = 0$, then every point of A is within zero distance of B, and

hence must belong to B since B is compact, so $A \subset B$ and similarly $B \subset A$. So $h(A, B) = 0$ implies that $A = B$.

We must prove the triangle inequality. For this it is enough to prove that

$$d(A, B) \le d(A, C) + d(C, B),$$

because interchanging the role of A and B gives the desired result. Now for any $a \in A$ we have

$$
\begin{aligned}
d(a, B) &= \min_{b \in B} d(a, b) \\
&\le \min_{b \in B} (d(a, c) + d(c, b)) \ \forall c \in C \\
&= d(a, c) + \min_{b \in B} d(c, b) \ \forall c \in C \\
&= d(a, c) + d(c, B) \ \forall c \in C \\
&\le d(a, c) + d(C, B) \ \forall c \in C.
\end{aligned}
$$

The second term in the last expression does not depend on c, so minimizing over c gives

$$d(a, B) \le d(a, C) + d(C, B).$$

Maximizing over a on the right gives

$$d(a, B) \le d(A, C) + d(C, B).$$

Maximizing on the left gives the desired

$$d(A, B) \le d(A, C) + d(C, A).$$

We sketch the proof of completeness. Let A_n be a sequence of compact non-empty subsets of X which is Cauchy in the Hausdorff metric. Define the set A to be the set of all $x \in X$ with the property that there exists a sequence of points $x_n \in A_n$ with $x_n \to x$. It is straighforward to prove that A is compact and non-empty and is the limit of the A_n in the Hausdorff metric.

Suppose that $\kappa : X \to X$ is a contraction. Then κ defines a transformation on the space of subsets of X (which we continue to denote by κ):

$$\kappa(A) = \{\kappa x | x \in A\}.$$

Since κ is continuous, it carries $\mathcal{H}(X)$ into itself. Let c be the Lipschitz constant of κ. Then

$$
\begin{aligned}
d(\kappa(A), \kappa(B)) &= \max_{a \in A}[\min_{b \in B} d(\kappa(a), \kappa(b))] \\
&\le \max_{a \in A}[\min_{b \in B} cd(a, b)] \\
&= cd(A, B).
\end{aligned}
$$

Similarly, $d(\kappa(B), \kappa(A)) \le c \, d(B, A)$ and hence

$$h(\kappa(A), \kappa(B)) \le c \, h(A, B). \tag{4.27}$$

In other words, a contraction on X induces a contraction on $\mathcal{H}(X)$.

The previous remark together with the following observation is the key to Hutchinson's remarkable construction of fractals:

Proposition 4.13.2 *Let $T_1, \ldots, T_n$ be a collection of contractions on $\mathcal{H}(X)$ with Lipschitz constants $c_1, \ldots, c_n$, and let $c = \max c_i$. Define the transformation T on $\mathcal{H}(X)$ by*

$$T(A) = T_1(A) \cup T_2(A) \cup \cdots \cup T_n(A).$$

Then T is a contraction with Lipschitz constant c.

Proof. By induction, it is enough to prove this for the case $n = 2$. By (4.26)

$$
\begin{aligned}
h(T(A), T(B)) &= h(T_1(A) \cup T_2(A), T_1(B) \cup T_2(B)) \\
&\leq \max\{h(T_1(A), h(T_1(B)), h(T_2(A), T_2(B))\} \\
&\leq \max\{c_1 h(A, B), c_2 h(A, B)\} \\
&= h(A, B) \max\{c_1, c_2\} = c \cdot h(A, B)
\end{aligned}
$$

.

Putting the previous facts together we get Hutchinson's theorem;

Theorem 4.13.1 *Let $T_1, \ldots, T_n$ be contractions on a complete metric space and let c be the maximum of their Lipschitz contants. Define the Hutchinoson operator T on $\mathcal{H}(X)$ by*

$$T(A) := T_1(A) \cup \cdots \cup T_n(A).$$

Then T is a contraction with Lipschtz constant c.

4.14   Affine examples

We describe several examples in which X is a subset of a vector space and each of the T_i in Hutchinson's theorem are affine transformations of the form

$$T_i : x \mapsto A_i x + b_i$$

where $b_i \in X$ and A_i is a linear transformation.

4.14.1   The classical Cantor set.

Take $X = [0, 1]$, the unit interval. Take

$$T_1 : x \mapsto \frac{x}{3}, \quad T_2 : x \mapsto \frac{x}{3} + \frac{2}{3}.$$

These are both contractions, so by Hutchinson's theorem there exists a unique closed fixed set C. This is the Cantor set.

To relate it to Cantor's original construction, let us go back to the proof of the contraction fixed point theorem applied to T acting on $\mathcal{H}(X)$. It says that if we start with any non-empty compact subset A_0 and keep applying T to it, i.e. set $A_n = T^n A_0$ then $A_n \to C$ in the Hausdorff metric, h. Suppose we take the interval I itself as our A_0. Then

$$A_1 = T(I) = [0, \frac{1}{3}] \cup [\frac{2}{3}, 1].$$

in other words, applying the Hutchinson operator T to the interval $[0, 1]$ has the effect of deleting the "middle third" open interval $(\frac{1}{3}, \frac{2}{3})$. Applying T once more gives

$$A_2 = T^2[0, 1] = [0, \frac{1}{9}] \cup [\frac{2}{9}, \frac{1}{3}] \cup [\frac{2}{3}, \frac{7}{9}] \cup [\frac{8}{9}, 1].$$

In other words, A_2 is obtained from A_1 by deleting the middle thirds of each of the two intervals of A_1 and so on. This was Cantor's original construction. Since $A_{n+1} \subset A_n$ for this choice of initial set, the Hausdorff limit coincides with the intersection.

But of course Hutchinson's theorem (and the proof of the contractions fixed point theorem) says that we can start with *any* non-empty closed set as our initial "seed" and then keep applying T. For example, suppose we start with the one point set $B_0 = \{0\}$. Then $B_1 = TB_0$ is the two point set

$$B_1 = \{0, \frac{2}{3}\},$$

B_2 consists of the four point set

$$B_2 = \{0, \frac{2}{9}, \frac{2}{3}, \frac{8}{9}\}$$

and so on. We then must take the Hausdorff limit of this increasing collection of sets.

To describe the limiting set c from this point of view, it is useful to use triadic expansions of points in $[0, 1]$. Thus

$$
\begin{array}{rcl}
0 & = & .0000000\cdots \\
2/3 & = & .2000000\cdots \\
2/9 & = & .0200000\cdots \\
8/9 & = & .2200000\cdots
\end{array}
$$

and so on. Thus the set B_n will consist of points whose triadic expansion has only zeros or twos in the first n positions followed by a string of all zeros. Thus a point will lie in C (be the limit of such points) if and only if it has a triadic expansion consisting entirely of zeros or twos. This includes the possibility of an infinite string of all twos at the tail of the expansion. for example, the point 1 which belongs to the Cantor set has a triadic expansion $1 = .222222\cdots$. Similarly the point $\frac{2}{3}$ has the triadic expansion $\frac{2}{3} = .0222222\cdots$ and so is in

the limit of the sets B_n. But a point such as $.101 \cdots$ is not in the limit of the B_n and hence not in C. This description of C is also due to Cantor. Notice that for any point a with triadic expansion $a = .a_1 a_2 a_2 \cdots$

$$T_1 a = .0 a_1 a_2 a_3 \cdots, \quad \text{while} \quad T_2 a = .2 a_1 a_2 a_3 \cdots.$$

Thus if all the entries in the expansion of a are either zero or two, this will also be true for $T_1 a$ and $T_2 a$. This shows that the C (given by this second Cantor description) satisfies $TC \subset C$. On the other hand,

$$T_1(.a_2 a_3 \cdots) = .0 a_2 a_3 \cdots, \quad T_2(.a_2 a_3 \cdots) = .2 a_2 a_3 \cdots$$

which shows that $.a_1 a_2 a_3 \cdots$ is in the image of T_1 if $a_1 = 0$ or in the image of T_2 if $a_1 = 2$. This shows that $TC = C$. Since C (according to Cantor's second description) is closed, the uniqueness part of the fixed point theorem guarantees that the second description coincides with the first.

The statement that $TC = C$ implies that C is "self-similar".

4.14.2   The Sierpinski Gasket

Consider the three affine transformations of the plane:

$$T_1 : \begin{pmatrix} x \\ y \end{pmatrix} \mapsto \frac{1}{2} \begin{pmatrix} x \\ y \end{pmatrix}, \quad T_2 : \begin{pmatrix} x \\ y \end{pmatrix} \mapsto \frac{1}{2} \begin{pmatrix} x \\ y \end{pmatrix} + \frac{1}{2} \begin{pmatrix} 1 \\ 0 \end{pmatrix},$$

$$T_3 : \begin{pmatrix} x \\ y \end{pmatrix} \mapsto \frac{1}{2} \begin{pmatrix} x \\ y \end{pmatrix} + \frac{1}{2} \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

The fixed point of the Hutchinson operator for this choice of T_1, T_2, T_3 is called the Sierpinski gasket, S. If we take our initial set A_0 to be the right triangle with vertices at

$$\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \text{and} \quad \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

then each of the $T_i A_0$ is a similar right triangle whose linear dimensions are one-half as large, and which shares one common vertex with the original triangle. In other words,

$$A_1 = T A_0$$

is obtained from our original triangle be deleting the interior of the (reversed) right triangle whose vertices are the midpoints of our orignal triangle. Just as in the case of the Cantor set, successive applications of T to this choice of original set amounts to successive deletions of the "middle" and the Hausdorff limit is the intersection of all of them: $S = \bigcap A_i$.

We can also start with the one element set

$$B_0 \left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix} \right\}$$

Using a binary expansion for the x and y coordinates, application of T to B_0 gives the three element set

$$\left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} .1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ .1 \end{pmatrix} \right\}.$$

The set $B_2 = TB_1$ will contain nine points, whose binary expansions are obtained from the above three by shifting the x and y exapnsions one unit to the right and either inserting a 0 before both expansions (the effect of T_1), insert a 1 before the expansion of x and a zero before the y or vice versa. Proceding in this fashion, we see that B_n consists of 3^n points which have all 0 in the binary expansion of the x and y coordinates, past the n-th position, and which are further constrained by the condition that at no earler point do we have both $x_i = 1$ and $y_i = 1$. Passing to the limit shows that S consists of all points for which we can find (possible inifinite) binary expansions of the x and y coordinates so that $x_i = 1 = y_i$ never occurs. (For example $x = \frac{1}{2}, y = \frac{1}{2}$ belongs to S because we can write $x = .10000\cdots, y = .011111\ldots$). Again, from this (second) description of S in terms of binary expansions it is clear that $TS = S$.

4.14.3 Moran's theorem

If A is any set such that $f[A] \subset A$, then clearly $f^p[A] \subset A$ by induction. If A is non-empty and closed, then for any $a \in A$, and any $x \in E$, the limit of the $f_\gamma(a)$ belongs to K as γ ranges over the first words of size p of x, and so belongs to K and also to A. Since these points consititute all of K, we see that

$$K \subset A$$

and hence

$$f_\beta(K) \subset f_\beta(A) \tag{4.28}$$

for any word β.

Now suppose that Moran's open set condition is satisfied, and let us write

$$O_\alpha := f_\alpha(O).$$

Then

$$O_\alpha \cap O_\beta = \emptyset$$

if α is not a prefix of β or β is not a prefix of α. Furthermore,

$$\overline{f_\beta(O)} = f_\beta(\overline{O})$$

so we can use the symbol

$$\overline{O}_\beta$$

unambiguously to denote these two equal sets. By virtue of (4.28) we have

$$K_\beta \subset \overline{O}_\beta$$

where we use K_β to denote $f_\beta(K)$. Suppose that α is not a prefix of β or vice versa. Then $K_\beta \cap O\alpha = \emptyset$ since $\overline{O}_\beta \cap O_\alpha = \emptyset$.

Let $\mathfrak{m}$ denote the measure on the string model E that we constructed above, so that $\mathfrak{m}(E) = 1$ and more generally $\mathfrak{m}([\alpha]) = w_\alpha^s$. Then we will have proved that the Hausdorff dimension of K is $\geq s$, and hence $= s$ if we can prove that there exists a constant b such that for every Borel set $B \subset K$

$$\mathfrak{m}(h^{-1}(B)) \leq b \cdot \operatorname{diam}(B)^s, \tag{4.29}$$

where $h : E \to K$ is the map we constructed above from the string model to K.

Let us introduce the following notation: For any (finite) non-empty string α, let α^- denote the string (of cardinality one less) obtained by removing the last letter in α.

Lemma 4.14.1 *There exists an integer N such that for any subset $B \subset K$ the set Q_B of all finite strings α such that*

$$\overline{O_\alpha} \cap B \neq \emptyset$$

and

$$\operatorname{diam} O_\alpha < \operatorname{diam} B \leq \operatorname{diam} O_{\alpha^-}$$

has at most N elements.

Proof. Let

$$D := \operatorname{diam} O.$$

The map f_α is a similarity with similarity ratio $\operatorname{diam}[\alpha]$ so

$$\operatorname{diam} O_\alpha = D \cdot \operatorname{diam}[\alpha].$$

Let $r := \min_i r_i$. Then if $\alpha \in Q_B$ we have

$$\operatorname{diam} O_\alpha = D \cdot \operatorname{diam}[\alpha] \geq D \cdot r \operatorname{diam}[\alpha^-] = r \operatorname{diam} O_{\alpha^-} \geq r \operatorname{diam} B.$$

Let V denote the volume of O relative to the d-dimensional Hausdorff measure of $\mathbb{R}^d$, i.e., up to a constant factor the Lebesgue measure. Let V_α denote the volume of O_α so that $V_\alpha = w_\alpha^d V = V \cdot (\operatorname{diam} O_\alpha)/\operatorname{diam} O)^d$. From the preceding displayed equation it follows that

$$V_\alpha \geq \frac{V r^d}{D^d} (\operatorname{diam} B)^d.$$

If $x \in B$, then every $y \in O_\alpha$ is within a distance $\operatorname{diam} B + \operatorname{diam} O_\alpha \leq 2 \operatorname{diam} B$ of x. So if m denotes the number of elements in Q_B, we have m disjoint sets with volume at least $\frac{V r^d}{D^d} (\operatorname{diam} B)^d$ all within a ball of radius $2 \cdot \operatorname{diam} B$. We have normalized our volume so that the unit ball has volume one, and hence the ball of radius $2 \cdot \operatorname{diam} B$ has volume $2^d (\operatorname{diam} B)^d$. Hence

$$m \cdot \frac{V r^d}{D^d} (\operatorname{diam} B)^d \leq 2^d (\operatorname{diam} B)^d$$

or

$$m \le \frac{2^d D^d}{V r^d}.$$

So any integer greater that the right hand side of this inequality (which is independent of B) will do. $\square$

Now we turn to the proof of (4.29) which will then complete the proof of Moran's theorem. Let B be a Borel subset of K. Then

$$B \subset \bigcup_{\alpha \in Q_B} \overline{O_\alpha}$$

so

$$h^{-1}(B) \subset \bigcup_{\alpha \in Q_B} [\alpha].$$

Now

$$([\alpha]) = (\operatorname{diam}[\alpha])^s = \left(\frac{1}{D} \operatorname{diam}(O_\alpha) \right)^s \le \frac{1}{D^s} (\operatorname{diam} B)^s$$

and so

$$\mathfrak{m}(h^{-1}(B)) \le \sum_{\alpha \in Q_B} \mathfrak{m}(\alpha) \le N \cdot \frac{1}{D^s} (\operatorname{diam} B)^s$$

and hence we may take

$$b = N \cdot \frac{1}{D^s} (\operatorname{diam} B)^s$$

and then (4.29) will hold.

Chapter 5

The Lebesgue integral.

In what follows, $(X, \mathcal{F}, m)$ is a space with a σ-field of sets, and m a measure on $\mathcal{F}$. The purpose of this chapter is to develop the theory of the Lebesgue integral for functions defined on X. The theory starts with **simple** functions, that is functions which take on only finitely many non-zero values, say $\{a_1, \ldots, a_n\}$ and where

$$A_i := f^{-1}(a_i) \in \mathcal{F}.$$

In other words, we start with functions of the form

$$\phi(x) = \sum_{i=1}^{n} a_i \mathbf{1}_{A_i} \qquad A_i \in \mathcal{F}. \tag{5.1}$$

Then, for any $E \in \mathcal{F}$ we would like to define the integral of a simple function ϕ over E as

$$\int_E \phi \, dm = \sum_{i=1}^{n} a_i m(A_i \cap E) \tag{5.2}$$

and extend this definition by some sort of limiting process to a broader class of functions.

I haven't yet specified what the range of the functions should be. Certainly, even to get started, we have to allow our functions to take values in a vector space over $\mathbf{R}$, in order that the expression on the right of (5.2) make sense. In fact, I will eventually allow f to take values in a Banach space. However the theory is a bit simpler for real valued functions, where the linear order of the reals makes some arguments easier. Of course it would then be no problem to pass to any finite dimensional space over the reals. But we will on occasion need integrals in infinite dimensional Banach spaces, and that will require a little reworking of the theory.

5.1   Real valued measurable functions.

Recall that if $(X, \mathcal{F})$ and $(Y, \mathcal{G})$ are spaces with σ-fields, then

$$f : X \to Y$$

is called measurable if

$$f^{-1}(E) \in \mathcal{F} \quad \forall \ E \in \mathcal{G}. \tag{5.3}$$

Notice that the collection of subsets of Y for which (5.3) holds is a σ-field, and hence if it holds for some collection $\mathcal{C}$, it holds for the σ-field generated by $\mathcal{C}$. For the next few sections we will take $Y = \mathbf{R}$ and $\mathcal{G} = \mathcal{B}$, the Borel field. Since the collection of open intervals on the line generate the Borel field, a real valued function $f : X \to \mathbf{R}$ is measurable if and only if

$$f^{-1}(I) \in \mathcal{F} \quad \text{for all open intervals} \ I.$$

Equally well, it is enough to check this for intervals of the form $(-\infty, a)$ for all real numbers a.

Proposition 5.1.1 *If* $F : \mathbf{R}^2 \to \mathbf{R}$ *is a continuous function and* f, g *are two measurable real valued functions on* X, *then* $F(f, g)$ *is measurable.*

Proof.   The set $F^{-1}(-\infty, a)$ is an open subset of the plane, and hence can be written as the countable union of products of open intervals $I \times J$. So if we set $h = F(f, g)$ then $h^{-1}((-\infty, a))$ is the countable union of the sets $f^{-1}(I) \cap g^{-1}(J)$ and hence belongs to $\mathcal{F}$. QED

From this elementary proposition we conclude that if f and g are measurable real valued functions then

- $f + g$ is measurable (since $(x, y) \mapsto x + y$ is continuous),

- fg is measurable (since $(x, y) \mapsto xy$ is continuous), hence

- $f\mathbf{1}_A$ is measurable for any $A \in \mathcal{F}$ hence

- f^+ is measurable since $f^{-1}([0, \infty]) \in \mathcal{F}$ and similarly for f^- so

- $|f|$ is measurable and so is $|f - g|$. Hence

- $f \wedge g$ and $f \vee g$ are measurable

and so on.

5.2   The integral of a non-negative function.

We are going to allow for the possibility that a function value or an integral might be infinite. We adopt the convention that

$$0 \cdot \infty = 0.$$

Recall that ϕ is **simple** if ϕ takes on a finite number of distinct non-negative (finite) values, $a_1, \ldots, a_n$, and that each of the sets

$$A_i = \phi^{-1}(a_i)$$

is measurable. These sets partition X:

$$X = A_1 \cup \cdots \cup A_n.$$

Of course since the values are distinct,

$$A_i \cap A_j = \emptyset \ \text{ for } \ i \neq j.$$

With this definition, a simple function can be written as in (5.1) and this expression is unique. So we may take (5.2) as the definition of the integral of a simple function. We now extend the definition to an arbitrary ($[0, \infty]$ valued) function f by

$$\int_E f dm := \sup I(E, f) \tag{5.4}$$

where

$$I(E, f) = \left\{ \int_E \phi dm : 0 \leq \phi \leq f, \ \phi \text{ simple} \right\}. \tag{5.5}$$

In other words, we take all integrals of expressions of simple functions ϕ such that $\phi(x) \leq f(x)$ at all x. We then define the integral of f as the supremum of these values.

Notice that if $A := f^{-1}(\infty)$ has positive measure, then the simple functions $n\mathbf{1}_A$ are all $\leq f$ and so $\int_X f dm = \infty$.

Proposition 5.2.1 *For simple functions, the definition (5.4) coincides with definition (5.2).*

Proof. Since ϕ is $\leq$ itself, the right hand side of (5.2) belongs to $I(E, \phi)$ and hence is $\leq \int_E \phi dm$ as given by (5.5). We must show the reverse inequality: Suppose that $\psi = \sum b_j \mathbf{1}_{B_j} \leq \phi$. We can write the right hand side of (5.2) as

$$\sum b_j m(E \cap B_j) = \sum_{i,j} b_j m(E \cap A_i \cap B_j)$$

since $E \cap B_j$ is the disjoint union of the sets $E \cap A_i \cap B_j$ because the A_i partition X, and m is additive on disjoint finite (even countable) unions. On each of the sets $A_i \cap B_j$ we must have $b_j \leq a_i$. Hence

$$\sum_{i,j} b_j m(E \cap A_i \cap B_j) \leq \sum_{i,j} a_i m(E \cap A_i \cap B_j) = \sum a_i m(E \cap A_i)$$

since the B_j partition X. QED

In the course of the proof of the above proposition we have also established

$$\psi \leq \phi \ \text{ for simple functions implies } \int_E \psi dm \leq \int_E \phi dm. \tag{5.6}$$

Suppose that E and F are disjoint measurable sets. Then

$$m(A_i \cap (E \cup F)) = m(A_i \cap E) + m(A_i \cap F)$$

so each term on the right of (5.2) breaks up into a sum of two terms and we conclude that

If ϕ is simple and $E \cap F = \emptyset$, then $\displaystyle\int_{E \cup F} \phi dm = \int_E \phi dm + \int_F \phi dm.$ (5.7)

Also, it is immediate from (5.2) that if $a \geq 0$ then

If ϕ is simple then $\displaystyle\int_E a\phi dm = a \int_E \phi dm.$ (5.8)

It is now immediate that these results extend to all non-negative measurable functions. We list the results and then prove them. In what follows f and g are non-negative measurable functions, $a \geq 0$ is a real number and E and F are measurable sets:

$$f \leq g \quad \Rightarrow \quad \int_E f dm \leq \int_E g dm. \tag{5.9}$$

$$\int_E f dm \quad = \quad \int_X \mathbf{1}_E f dm \tag{5.10}$$

$$E \subset F \quad \Rightarrow \quad \int_E f dm \leq \int_F f dm. \tag{5.11}$$

$$\int_E a f dm \quad = \quad a \int_E f dm. \tag{5.12}$$

$$m(E) = 0 \quad \Rightarrow \quad \int_E f dm = 0. \tag{5.13}$$

$$E \cap F = \emptyset \quad \Rightarrow \quad \int_{E \cup F} f dm = \int_E f dm + \int_F f dm. \tag{5.14}$$

$$f = 0 \;\; \text{a.e.} \quad \Leftrightarrow \quad \int_X f dm = 0. \tag{5.15}$$

$$f \leq g \;\; \text{a.e.} \quad \Rightarrow \quad \int_X f dm \leq \int_X g dm. \tag{5.16}$$

Proofs.
(5.9):$I(E, f) \subset I(E, g)$.
(5.10): If ϕ is a simple function with $\phi \leq f$, then multiplying ϕ by $\mathbf{1}_E$ gives a function which is still $\leq f$ and is still a simple function. The set $I(E, f)$ is unchanged by considering only simple functions of the form $\mathbf{1}_E \phi$ and these constitute all simple functions $\leq \mathbf{1}_E f$.
(5.11): We have $\mathbf{1}_E f \leq \mathbf{1}_F f$ and we can apply (5.9) and (5.10).
(5.12): $I(E, af) = aI(E, f)$.
(5.13): In the definition (5.2) all the terms on the right vanish since $m(E \cap A_i) = 0$. So $I(E, f)$ consists of the single element 0.

(5.14): This is true for simple functions, so $I(E \cup F, f) = I(E, f) + I(F, f)$ meaning that every element of $I(E \cup F, f)$ is a sum of an element of $I(E, f)$ and an element of $I(F, f)$. Thus the sup on the left is $\leq$ the sum of the sups on the right, proving that the left hand side of (5.14) is $\leq$ its right hand side. To prove the reverse inequality, choose a simple function $\phi \leq \mathbf{1}_E f$ and a simple function $\psi \leq \mathbf{1}_F f$. Then $\phi + \psi \leq \mathbf{1}_{E \cup F} f$ since $E \cap F = \emptyset$. So $\phi + \psi$ is a simple function $\leq f$ and hence

$$\int_E \phi dm + \int_F \psi dm \leq \int_{E \cup F} f dm.$$

If we now maximize the two summands separately we get

$$\int_E f dm + \int_F f dm \leq \int_{E \cup F} f dm$$

which is what we want.

(5.15): If $f = 0$ almost everywhere, and $\phi \leq f$ then $\phi = 0$ a.e. since $\phi \geq 0$. This means that all sets which enter into the right hand side of (5.2) with $a_i \neq 0$ have measure zero, so the right hand side vanishes. So $I(X, f)$ consists of the single element 0. This proves $\Rightarrow$ in (5.15). We wish to prove the reverse implication. Let $A = \{x | f(x) > 0\}$. We wish to show that $m(A) = 0$. Now

$$A = \bigcup A_n \quad \text{where} \quad A_n := \{x | f(x) > \frac{1}{n}\}.$$

The sets A_n are increasing, so we know that $m(A) = \lim_{n \to \infty} m(A_n)$. So it is enough to prove that $m(A_n) = 0$ for all n. But

$$\frac{1}{n} \mathbf{1}_{A_n} \leq f$$

and is a simple function. So

$$\frac{1}{n} \int_X \mathbf{1}_{A_n} dm = \frac{1}{n} m(A_n) \leq \int_X f dm = 0$$

implying that $m(A_n) = 0$.

(5.16): Let $E = \{x | f(x) \leq g(x)\}$. Then E is measurable and E^c is of measure zero. By definition, $\mathbf{1}_E f \leq \mathbf{1}_E g$ everywhere, hence by (5.11)

$$\int_X \mathbf{1}_E f dm \leq \int_X \mathbf{1}_E g dm.$$

But

$$\int_X \mathbf{1}_E f dm + \int_X \mathbf{1}_{E^c} f dm = \int_E f dm + \int_{E^c} f dm = \int_X f dm$$

where we have used (5.14) and (5.13). Similarly for g. QED

5.3   Fatou's lemma.

This says:

Theorem 5.3.1 *If* $\{f_n\}$ *is a sequence of non-negative functions, then*

$$\lim_{n\to\infty} \inf_{k\ge n} \int f_k dm \ge \int \left(\lim_{n\to\infty} \inf_{k\ge n} f_k \right) dm. \qquad (5.17)$$

Recall that the limit inferior of a sequence of numbers $\{a_n\}$ is defined as follows:
Set

$$b_n := \inf_{k\ge n} a_k$$

so that the sequence $\{b_n\}$ is non-decreasing, and hence has a limit (possibly infinite) which is defined as the lim inf. For a sequence of functions, lim inf f_n is obtained by taking lim inf $f_n(x)$ for every x.

Consider the sequence of simple functions $\{\mathbf{1}_{[n,n+1]}\}$. At each point x the lim inf is 0, in fact $\mathbf{1}_{[n,n+1]}(x)$ becomes and stays 0 as soon as $n > x$. Thus the right hand side of (5.17) is zero. The numbers which enter into the left hand side are all 1, so the left hand side is 1.

Similarly, if we take $f_n = n\mathbf{1}_{(0,1/n]}$, the left hand side is 1 and the right hand side is 0. So without further assumptions, we generally expect to get strict inequality in Fatou's lemma.

Proof: Set

$$g_n := \inf_{k\ge n} f_k$$

so that

$$g_n \le g_{n+1}$$

and set

$$f := \lim_{n\to\infty} \inf_{k\ge n} f_n = \lim_{n\to\infty} g_n.$$

Let

$$\phi \le f$$

be a simple function. We must show that

$$\int \phi dm \le \lim_{n\to\infty} \inf_{k\ge n} \int f_k dm. \qquad (5.18)$$

There are two cases to consider:

a) $m\left(\{x : \phi(x) > 0\}\right) = \infty$. In this case $\int \phi dm = \infty$ and hence $\int f dm = \infty$ since $\phi \le f$. We must show that $\liminf \int f_n dm = \infty$. Let

$$D := \{x : \phi(x) > 0\} \quad \text{so } m(D) = \infty.$$

Choose some positive number $b <$ all the positive values taken by ϕ. This is possible since there are only finitely many such values.

Let
$$D_n := \{x | g_n(x) > b\}.$$
The $D_n \nearrow D$ since $b < \phi(x) \leq \lim_{n \to \infty} g_n(x)$ at each point of D. Hence $m(D_n) \to m(D) = \infty$. But
$$bm(D_n) \leq \int_{D_n} g_n dm \leq \int_{D_n} f_k dm \quad k \geq n$$
since $g_n \leq f_k$ for $k \geq n$. Now
$$\int f_k dm \geq \int_{D_n} f_k dm$$
since f_k is non-negative. Hence $\liminf \int f_n dm = \infty$.

b) $m(\{x : \phi(x) > 0\}) < \infty$. Choose $\epsilon > 0$ so that it is less than the minimum of the positive values taken on by ϕ and set
$$\phi_\epsilon(x) = \begin{cases} \phi(x) - \epsilon & \text{if} \quad \phi(x) > 0 \\ 0 & \text{if} \quad \phi(x) = 0. \end{cases}$$

Let
$$C_n := \{x | g_n(x) \geq \phi_\epsilon\}$$
and
$$C = \{x : f(x) \geq \phi_\epsilon\}.$$
Then $C_n \nearrow C$. We have
$$\begin{aligned} \int_{C_n} \phi_\epsilon dm & \leq \int_{C_n} g_n dm \\ & \leq \int_{C_n} f_k dm \quad k \geq n \\ & \leq \int_C f_k dm \quad k \geq n \\ & \leq \int f_k dm \quad k \geq n. \end{aligned}$$

So
$$\int_{C_n} \phi_\epsilon dm \leq \liminf \int f_k dm.$$
We will next let $n \to \infty$: Let c_i be the non-zero values of ϕ_ϵ so
$$\phi_\epsilon = \sum c_i \mathbf{1}_{B_i}$$
for some measurable sets $B_i \subset C$. Then
$$\int_{C_n} \phi_\epsilon dm = \sum c_i m(B_i \cap C_n) \to \sum c_i m(B_i) = \int \phi_\epsilon dm$$

since $(B_i \cap C_n) \nearrow B_i \cap C = B_i$. So

$$\int \phi_\epsilon dm \leq \liminf \int f_k dm.$$

Now

$$\int \phi_\epsilon dm = \int \phi dm - \epsilon m \left(\{x|\phi(x) > 0\}\right).$$

Since we are assuming that $m\left(\{x|\phi(x) > 0\}\right) < \infty$, we can let $\epsilon \to 0$ and conclude that $\int \phi dm \leq \liminf \int f_k dm$. QED

5.4   The monotone convergence theorem.

We assume that $\{f_n\}$ is a sequence of non-negative measurable functions, and that $f_n(x)$ is an increasing sequence for each x. Define $f(x)$ to be the limit (possibly $+\infty$) of this sequence. We describe this situation by $f_n \nearrow f$. The monotone convergence theorem asserts that:

$$f_n \geq 0, \quad f_n \nearrow f \quad \Rightarrow \quad \lim_{n\to\infty} \int f_n dm = \int f dm. \qquad (5.19)$$

The f_n are increasing and all $\leq f$ so the $\int f_n dm$ are monotone increasing and all $\leq \int f dm$. So the limit exists and is $\leq \int f dm$. On the other hand, Fatou's lemma gives

$$\int f dm \leq \liminf \int f_n dm = \lim \int f_n dm.$$

QED

   In the monotone convergence theorem we need only know that

$$f_n \nearrow f \quad \text{a.e.}$$

Indeed, let C be the set where convergence holds, so $m(C^c) = 0$. Let $g_n = \mathbf{1}_C f_n$ and $g = \mathbf{1}_C f$. Then $g_n \nearrow g$ everywhere, so we may apply (5.19) to g_n and g. But $\int g_n dm = \int f_n dm$ and $\int g dm = \int f dm$ so the theorem holds for f_n and f as well.

5.5   The space $\mathcal{L}_1(X, \mathbf{R})$.

We will say an $\mathbf{R}$ valued measurable function is **integrable** if both $\int f^+ dm < \infty$ and $\int f^- dm < \infty$. If this happens, we set

$$\int f dm := \int f^+ dm - \int f^- dm. \qquad (5.20)$$

Since both numbers on the right are finite, this difference makes sense. Some authors prefer to allow one or the other numbers (but not both) to be infinite,

in which case the right hand side of (5.20) might be $= \infty$ or $-\infty$. We will stick with the above convention.

We will denote the set of all (real valued) integrable functions by $\mathcal{L}_1$ or $\mathcal{L}_1(X)$ or $\mathcal{L}_1(X, \mathbf{R})$ depending on how precise we want to be.

Notice that if $f \leq g$ then $f^+ \leq g^+$ and $f^- \geq g^-$ all of these functions being non-negative. So

$$\int f^+ dm \leq \int g^+ dm, \quad \int f^- dm \geq \int g^- dm$$

hence

$$\int f^+ dm - \int f^- dm \leq \int g^+ dm - \int g^- dm$$

or

$$f \leq g \quad \Rightarrow \quad \int f dm \leq \int g dm. \tag{5.21}$$

If a is a non-negative number, then $(af)^{\pm} = af^{\pm}$. If $a < 0$ then $(af)^{\pm} = (-a)f^{\mp}$ so in all cases we have

$$\int af dm = a \int f dm. \tag{5.22}$$

We now wish to establish

$$f, g \in \mathcal{L}_1 \Rightarrow f + g \in \mathcal{L}_1 \text{ and } \int (f + g) dm = \int f dm + \int g dm. \tag{5.23}$$

Proof. We prove this in stages:

- First assume $f = \sum a_i \mathbf{1}_{A_i}$, $g = \sum b_i \mathbf{1}_{B_i}$ are non-negative simple functions, where the A_i partition X as do the B_j. Then we can decompose and recombine the sets to yield:

$$\begin{aligned}
\int (f + g) dm &= \sum_{i,j} (a_i + b_j) m(A_i \cap B_j) \\
&= \sum_i \sum_j a_i m(A_i \cap B_j) + \sum_j \sum_i b_j m(A_i \cap B_j) \\
&= \sum_i a_i m(A_i) + \sum_j b_j m(B_j) \\
&= \int f dm + \int g dm
\end{aligned}$$

  where we have used the fact that m is additive and the $A_i \cap B_j$ are disjoint sets whose union over j is A_i and whose union over i is B_j.

- Next suppose that f and g are non-negative measurable functions with finite integrals. Set

$$f_n := \sum_{k=0}^{2^{2n}} \frac{k}{2^n} \mathbf{1}_{f^{-1}[\frac{k}{2^n}, \frac{k+1}{2^n}]}.$$

Each f_n is a simple function $\le f$, and passing from f_n to f_{n+1} involves splitting each of the sets $f^{-1}([\frac{k}{2^n}, \frac{k+1}{2^n}])$ in the sum into two, and choosing a larger value on the second portion. So the f_n are increasing. Also, if $f(x) < \infty$, then $f(x) < 2^m$ for some m, and for any $n > m$ $f_n(x)$ differs from $f(x)$ by at most 2^{-n}. Hence $f_n \nearrow f$ a.e., since f is finite a.e because its integral is finite. Similarly we can construct $g_n \nearrow g$. Also $(f_n + g_n) \nearrow f + g$ a.e.

By the a.e. monotone convergence theorem

$$\int (f+g)dm = \lim \int (f_n+g_n)dm = \lim \int f_n dm + \lim \int g_n dm = \int f dm + \int g dm,$$

where we have used (5.23) for simple functions. This argument shows that $\int (f+g)dm < \infty$ if both integrals $\int f dm$ and $\int g dm$ are finite.

- For any $f \in \mathcal{L}_1$ we conclude from the preceding that

$$\int |f| dm = \int (f^+ + f^-)dm < \infty.$$

Similarly for g. Since $|f + g| \le |f| + |g|$ we conclude that both $(f+g)^+$ and $(f+g)^-$ have finite integrals. Now

$$(f+g)^+ - (f+g)^- = f + g = (f^+ - f^-) + (g^+ - g^-)$$

or

$$(f+g)^+ + f^- + g^- = f^+ + g^+ + (f+g)^-.$$

All expressions are non-negative and integrable. So integrate both sides to get (5.23).QED

We have thus established

Theorem 5.5.1 *The space $\mathcal{L}_1(X, \mathbf{R})$ is a real vector space and $f \mapsto \int f dm$ is a linear function on $\mathcal{L}_1(X, \mathbf{R})$.*

We also have

Proposition 5.5.1 *If $h \in \mathcal{L}_1$ and $\int_A h dm \ge 0$ for all $A \in \mathcal{F}$ then $h \ge 0$ a.e.*

Proof: Let $A_n : \{x | h(x) \le -\frac{1}{n}\}$. Then

$$\int_{A_n} h dm \le \int_{A_n} \frac{-1}{n} dm = -\frac{1}{n} m(A_n)$$

so $m(A_n) = 0$. But if we let $A := \{x | h(x) < 0\}$ then $A_n \nearrow A$ and hence $m(A) = 0$.QED

We have defined the integral of any function f as $\int f dm = \int f^+ dm - \int f^- dm$, and $\int |f| dm = \int f^+ dm + \int f^- dm$. Since for any two non-negative real numbers $a - b \leq a + b$ we conclude that

$$\left| \int f dm \right| \leq \int |f| dm. \tag{5.24}$$

If we define

$$\|f\|_1 := \int |f| dm$$

we have verified that

$$\|f + g\|_1 \leq \|f\|_1 + \|g\|_1,$$

and have also verified that

$$\|cf\|_1 = |c| \|f\|_1.$$

In other words, $\| \cdot \|_1$ is a semi-norm on $\mathcal{L}_1$. From the preceding proposition we know that $\|f\|_1 = 0$ if and only if $f = 0$ a.e. The question of whether we want to pass to the quotient and identify two functions which differ on a set of measure zero is a matter of taste.

5.6   The dominated convergence theorem.

This says that

Theorem 5.6.1 *Let f_n be a sequence of measurable functions such that*

$$|f_n| \leq g \ a.e., \quad g \in \mathcal{L}_1.$$

Then

$$f_n \to f \quad a.e. \ \Rightarrow \ f \in \mathcal{L}_1 \ and \ \int f_n dm \to \int f dm.$$

Proof.   The functions f_n are all integrable, since their positive and negative parts are dominated by g. Assume for the moment that $f_n \geq 0$. Then Fatou's lemma says that

$$\int f dm \leq \liminf \int f_n dm.$$

Fatou's lemma applied to $g - f_n$ says that

$$\int (g - f) dm \leq \liminf \int (g - f_n) dm = \liminf \left(\int g dm - \int f_n dm \right)$$

$$= \int g dm - \limsup \int f_n dm.$$

Subtracting $\int g \, dm$ gives

$$\limsup \int f_n dm \leq \int f dm.$$

So

$$\limsup \int f_n dm \leq \int f dm \leq \liminf \int f_n dm$$

which can only happen if all three are equal.

We have proved the result for non-negative f_n. For general f_n we can write our hypothesis as

$$-g \leq f_n \leq g \quad \text{a.e..}$$

Adding g to both sides gives

$$0 \leq f_n + g \leq 2g \quad \text{a.e..}$$

We now apply the result for non-negative sequences to $g + f_n$ and then subtract off $\int g \, dm$.

5.7   Riemann integrability.

Suppose that $X = [a, b]$ is an interval. What is the relation between the Lebesgue integral and the Riemann integral? Let us suppose that $[a, b]$ is bounded and that f is a bounded function, say $|f| \leq M$. Each partition

$$P : a = a_0 < a_1 < \cdots < a_n = b$$

into intervals $I_i = [a_{i-1}, a_i]$ with

$$m_i := m(I_i) = a_i - a_{i-1}, \quad i = 1, \ldots, n$$

defines a Riemann lower sum

$$L_P = \sum k_i m_i \quad k_i = \inf_{x \in I_i} f(x)$$

and a Riemann upper sum

$$U_P = \sum M_i m_i \quad M_i := \sup_{x \in I_i} f(x)$$

which are the Lebesgue integrals of the simple functions

$$\ell_P := \sum k_i \mathbf{1}_{I_i} \quad \text{and} \quad u_P := \sum M_i \mathbf{1}_{I_i}$$

respectively.

According to Riemann, we are to choose a sequence of partitions P_n which refine one another and whose maximal interval lengths go to zero. Write ℓ_i for ℓ_{P_i} and u_i for u_{P_i}. Then

$$\ell_1 \leq \ell_2 \leq \cdots \leq f \leq \cdots \leq u_2 \leq u_1.$$

Suppose that f is measurable. All the functions in the above inequality are Lebesgue integrable, so dominated convergence implies that

$$\lim U_n = \lim \int_a^b u_n dx = \int_a^b u dx$$

where $u = \lim u_n$ with a similar equation for the lower bounds. The Riemann integral is defined as the common value of $\lim L_n$ and $\lim U_n$ whenever these limits are equal.

Proposition 5.7.1 *f is Riemann integrable if and only if f is continuous almost everywhere.*

Proof. Notice that if x is not an endpoint of any interval in the partitions, then f is continuous at x if and only if $u(x) = \ell(x)$. Riemann's condition for integrability says that $\int (u-\ell)dm = 0$ which implies that f is continuous almost everywhere.

Conversely, if f is continuous a.e. then $u = f = \ell$ a.e.. Since u is measurable so is f, and since we are assuming that f is bounded, we conclude that f Lebesgue integrable. As $\ell = f = u$ a.e. their Lebesgue integrals coincide. But the statement that the Lebesgue integral of u is the same as that of ℓ is precisely the statement of Riemann integrability.QED

Notice that in the course of the proof we have also shown that the Lebesgue and Riemann integrals coincide when both exist.

5.8 The Beppo - Levi theorem.

We begin with a lemma:

Lemma 5.8.1 *Let $\{g_n\}$ be a sequence of non-negative measurable functions. Then*

$$\int \sum_{n=1}^{\infty} g_n \, dm = \sum_{n=1}^{\infty} \int g_n dm.$$

Proof. We have

$$\int \sum_{k=1}^{n} g_k dm = \sum_{k=1}^{n} \int g_k dm$$

for finite n by the linearity of the integral. Since the $g_k \geq 0$, the sums under the integral sign are increasing, and by definition converge to $\sum_{k=1}^{\infty} g_k$. The monotone convergence theorem implies the lemma. QED

But both sides of the equation in the lemma might be infinite.

Theorem 5.8.1 Beppo-Levi. *Let $f_n \in \mathcal{L}_1$ and suppose that*

$$\sum_{k=1}^{\infty} \int |f_k| dm < \infty.$$

Then $\sum f_k(x)$ converges to a finite limit for almost all x, the sum is integrable, and

$$\int \sum_{k=1}^{\infty} f_k \ dm = \sum_{k=1}^{\infty} \int f_k dm.$$

Proof. Take $g_n := |f_n|$ in the lemma. If we set $g = \sum_{n=1}^{\infty} g_n = \sum_{n=1}^{\infty} |f_n|$ then the lemma says that

$$\int g dm = \sum_{n=1}^{\infty} \int |f_n| dm,$$

and we are assuming that this sum is finite. So g is integrable, in particular the set of x for which $g(x) = \infty$ must have measure zero. In other words,

$$\sum_{n=1}^{\infty} |f_n(x)| < \infty \quad \text{a.e.} \ .$$

If a series is absolutely convergent, then it is convergent, so we can say that $\sum f_n(x)$ converges almost everywhere. Let

$$f(x) = \sum_{n=1}^{\infty} f_n(x)$$

at all points where the series converges, and set $f(x) = 0$ at all other points. Now

$$\left| \sum_{n=0}^{\infty} f_n(x) \right| \leq g(x)$$

at all points, and hence by the dominated convergence theorem, $f \in \mathcal{L}_1$ and

$$\int f dm = \int \lim_{n \to \infty} \sum_{k=1}^{n} f_k \ dm = \lim_{n \to \infty} \sum \int f_k dm = \sum_{k=1}^{\infty} \int f_k dm$$

QED

5.9 $\mathcal{L}_1$ is complete.

This is an immediate corollary of the Beppo-Levi theorem and Fatou's lemma. Indeed, suppose that $\{h_n\}$ is a Cauchy sequence in $\mathcal{L}_1$. Choose n_1 so that

$$\|h_n - h_{n_1}\| \leq \frac{1}{2} \quad \forall \, n \geq n_1.$$

Then choose $n_2 > n_1$ so that

$$\|h_n - h_{n_2}\| \leq \frac{1}{2^2} \quad \forall \, n \geq n_2.$$

Continuing this way, we have produced a subsequence h_{n_j} such that

$$\|h_{n_{j+1}} - h_{n_j}\| \le \frac{1}{2^j}.$$

Let

$$f_j := h_{n_{j+1}} - h_{n_j}.$$

Then

$$\int |f_j| dm < \frac{1}{2^j}$$

so the hypotheses of the Beppo-Levy theorem are satisfied, and $\sum f_j$ converges almost everywhere to some limit $f \in \mathcal{L}_1$. But

$$h_{n_1} + \sum_{j=1}^{k} f_j = h_{n_{k+1}}.$$

So the subsequence h_{n_k} converges almost everywhere to some $h \in \mathcal{L}_1$.

We must show that this h is the limit of the h_n in the $\| \cdot \|_1$ norm. For this we will use Fatou's lemma.

For a given $\epsilon > 0$, choose N so that $\|h_n - h_m\| < \epsilon$ for $k, n > N$. Since $h = \lim h_{n_j}$ we have, for $k > N$,

$$\|h - h_k\|_1 = \int |h - h_k| dm = \int \lim_{j \to \infty} |h_{n_j} - h_k| dm \le \liminf \int |h_{n_j} - h_k| dm$$

$$= \liminf \|h_{n_j} - h_k\| < \epsilon.$$

QED

5.10 Dense subsets of $\mathcal{L}_1(\mathbf{R}, \mathbf{R})$.

Up until now we have been studying integration on an arbitrary measure space $(X, \mathcal{F}, m)$. In this section and the next, we will take $X = \mathbf{R}$, $\mathcal{F}$ to be the σ-field of Lebesgue measurable sets, and m to be Lebesgue measure, in order to simplify some of the formulations and arguments.

Suppose that f is a Lebesgue integrable non-negative function on $\mathbf{R}$. We know that for any $\epsilon > 0$ there is a simple function ϕ such that

$$\phi \le f$$

and

$$\int f dm - \int \phi dm = \int (f - \phi) dm = \|f - \phi\|_1 < \epsilon.$$

To say that ϕ is simple implies that

$$\phi = \sum a_i \mathbf{1}_{A_i}$$

(finite sum) where each of the $a_i > 0$ and since $\int \phi \, dm < \infty$ each A_i has finite measure. Since $m(A_i \cap [-n, n]) \to m(A_i)$ as $n \to \infty$, we may choose n sufficiently large so that

$$\|f - \psi\|_1 < 2\epsilon \quad \text{where} \quad \psi = \sum a_i \mathbf{1}_{A_i \cap [-n,n]}.$$

For each of the sets $A_i \cap [-n, n]$ we can find a bounded open set U_i which contains it, and such that $m(U_i / A_i)$ is as small as we please. So we can find finitely many bounded open sets U_i such that

$$\|f - \sum a_i \mathbf{1}_{U_i}\|_1 < 3\epsilon.$$

Each U_i is a countable union of disjoint open intervals, $U_i = \bigcup_j I_{i,j}$, and since $m(U_i) = \sum_j m(I_{i,j})$, we can find finitely many $I_{i,j}$, j ranging over a finite set of integers, J_i such that $m\left(\bigcup_{j \in J_i}\right)$ is as close as we like to $m(U_i)$. So let us call a **step function** a function of the form $\sum b_i \mathbf{1}_{I_i}$ where the I_i are bounded intervals. We have shown that we can find a step function with positive coefficients which is as close as we like in the $\|\cdot\|_1$ norm to f. If f is not necessarily non-negative, we know (by definition!) that f^+ and f^- are in $\mathcal{L}_1$, and so we can approximate each by a step function. the triangle inequality then gives

Proposition 5.10.1 *The step functions are dense in $\mathcal{L}_1(\mathbf{R}, \mathbf{R})$.*

If $[a, b]$, $a < b$ is a finite interval, we can approximate $\mathbf{1}_{[a,b]}$ as closely as we like in the $\|\cdot\|_1$ norm by continuous functions: just choose n large enough so that $\frac{2}{n} < b - a$, and take the function which is 0 for $x < a$, rises linearly from 0 to 1 on $[a, a + \frac{1}{n}]$, is identically 1 on $[a + \frac{1}{n}, b - \frac{1}{n}]$, and goes down linearly from 1 to 0 from $b - \frac{1}{n}$ to b and stays 0 thereafter. As $n \to \infty$ this clearly tends to $\mathbf{1}_{[a,b]}$ in the $\|\cdot\|_1$ norm. So

Proposition 5.10.2 *The continuous functions of compact support are dense in $\mathcal{L}_1(\mathbf{R}, \mathbf{R})$.*

As a consequence of this proposition, we see that we could have avoided all of measure theory if our sole purpose was to define the space $\mathcal{L}_1(\mathbf{R}, \mathbf{R})$. We could have defined it to be the completion of the space of continuous functions of compact support relative to the $\|\cdot\|_1$ norm.

5.11   The Riemann-Lebesgue Lemma.

We will state and prove this in the "generalized form". Let h be a bounded measurable function on $\mathbf{R}$. We say that h satisfies the **averaging condition** if

$$\lim_{|c| \to \infty} \frac{1}{|c|} \int_0^c h \, dm \to 0. \tag{5.25}$$

For example, if $h(t) = \cos \xi t$, $\xi \neq 0$, then the expression under the limit sign in the averaging condition is

$$\frac{1}{c\xi} \sin \xi t$$

which tends to zero as $|c| \to \infty$. Here the oscillations in h are what give rise to the averaging condition. As another example, let

$$h(t) = \left\{ \begin{array}{ll} 1 & |t| \leq t \\ 1/|t| & |t| \geq 1. \end{array} \right.$$

Then the left hand side of (5.25) is

$$\frac{1}{|c|} (1 + \log|c|), \qquad |c| \geq 1.$$

Here the averaging condition is satisfied because the integral in (5.25) grows more slowly that $|c|$.

Theorem 5.11.1 [Generalized Riemann-Lebesgue Lemma].
   Let $f \in \mathcal{L}_1([c,d], \mathbf{R})$, $-\infty \leq c < d \leq \infty$. If h satisfies the averaging condition (5.25) then

$$\lim_{r \to \infty} \int_c^d f(t)h(rt)dt = 0. \tag{5.26}$$

Proof. Our proof will use the density of step functions, Proposition 5.10.1. We first prove the theorem when $f = \mathbf{1}_{[a,b]}$ is the indicator function of a finite interval. Suppose for example that $0 \leq a < b$. Then the integral on the right hand side of (5.26) is

$$\int_0^\infty \mathbf{1}_{[a.b]} h(rt)dt \;\; = \;\; \int_a^b h(rt)dt, \;\; \text{or setting } x = rt$$

$$= \;\; \frac{1}{r} \int_0^{br} h(x)dx - \frac{1}{r} \int_0^{ra} h(x)dx$$

and each of these terms tends to 0 by hypothesis. The same argument will work for any bounded interval $[a, b]$ we will get a sum or difference of terms as above. So we have proved (5.26) for indicator functions of intervals and hence for step functions.

   Now let M be such that $|h| \leq M$ everywhere (or almost everywhere) and choose a step function s so that

$$\|f - s\|_1 \leq \frac{\epsilon}{2M}.$$

Then $fh = (f - s)h + sh$

$$\left| \int f(t)h(rt)dt \right| \;\; = \;\; \left| \int (f(t) - s(t)h(rt)dt + \int s(t)h(rt)dt \right|$$

$$\leq \;\; \left| \int (f(t) - s(t))h(rt)dt \right| + \left| \int s(t)h(rt)dt \right|$$

$$\leq \;\; \frac{\epsilon}{2M} M + \left| \int s(t)h(rt)dt \right|.$$

We can make the second term $< \frac{\epsilon}{2}$ by choosing r large enough. QED

5.11.1   The Cantor-Lebesgue theorem.

This says:

Theorem 5.11.2 *If a trigonometric series*

$$\frac{a_0}{2} + \sum_n d_n \cos(nt - \phi_n) \qquad d_n \in \mathbf{R}$$

converges on a set E of positive Lebesgue measure then

$$d_n \to 0.$$

(I have written the general form of a real trigonometric series as a cosine series with phases since we are talking about only real valued functions at the present. Of course, applied to the real and imaginary parts, the theorem asserts that if $\sum a_n e^{inx}$ converges on a set of positive measure, then the $a_n \to 0$. Also, the notation suggests - and this is my intention - that the n's are integers. But in the proof below all that we will need is that the n's are any sequence of real numbers tending to ∞.)

Proof. The proof is a nice application of the dominated convergence theorem, which was invented by Lebesgue in part precisely to prove this theorem.

We may assume (by passing to a subset if necessary) that E is contained in some finite interval $[a, b]$. If $d_n \not\to 0$ then there is an $\epsilon > 0$ and a subsequence $|d_{n_k}| > \epsilon$ for all k. If the series converges, all its terms go to 0, so this means that

$$\cos(n_k t - \phi_k) \to 0 \quad \forall\, t \in E.$$

So

$$\cos^2(n_k t - \phi_k) \to 0 \quad \forall\, t \in E.$$

Now $m(E) < \infty$ and $\cos^2(n_k t - \phi_k) \le 1$ and the constant 1 is integrable on $[a, b]$. So we may take the limit under the integral sign using the dominated convergence theorem to conclude that

$$\lim_{k \to \infty} \int_E \cos^2(n_k t - \phi_k)dt = \int_E \lim_{k \to \infty} \cos^2(n_k t - \phi_k)dt = 0.$$

But

$$\cos^2(n_k t - \phi_k) = \frac{1}{2}[1 + \cos 2(n_k t - \phi_k)]$$

so

$$
\begin{aligned}
\int_E \cos^2(n_k t - \phi_k)dt &= \frac{1}{2}\int_E [1 + \cos 2(n_k t - \phi_k)]dt \\
&= \frac{1}{2}\left[m(E) + \int_E \cos 2(n_k t - \phi_k) \right] \\
&= \frac{1}{2}m(E) + \frac{1}{2}\int_{\mathbf{R}} \mathbf{1}_E \cos 2(n_k t - \phi_k)dt.
\end{aligned}
$$

But $\mathbf{1}_E \in \mathcal{L}_1(\mathbf{R}, \mathbf{R})$ so the second term on the last line goes to 0 by the Riemann Lebesgue Lemma. So the limit is $\frac{1}{2}m(E)$ instead of 0, a contradiction. QED

5.12 Fubini's theorem.

This famous theorem asserts that under suitable conditions, a double integral is equal to an iterated integral. We will prove it for real (and hence finite dimensional) valued functions on arbitrary measure spaces. (The proof for Banach space valued functions is a bit more tricky, and we shall omit it as we will not need it. This is one of the reasons why we have developed the real valued theory first.) We begin with some facts about product σ-fields.

5.12.1 Product σ-fields.

Let $(X, \mathcal{F})$ and $(Y, \mathcal{G})$ be spaces with σ-fields. On $X \times Y$ we can consider the collection $\mathcal{P}$ of all sets of the form

$$A \times B, \quad A \in \mathcal{F},\ B \in \mathcal{G}.$$

The σ-field generated by $\mathcal{P}$ will, by abuse of language, be denoted by

$$\mathcal{F} \times \mathcal{G}.$$

If E is any subset of $X \times Y$, by an even more serious abuse of language we will let

$$E_x := \{y | (x, y) \in E\}$$

and (contradictorily) we will let

$$E_y := \{x | (x, y) \in E\}.$$

The set E_x will be called the x-**section** of E and the set E_y will be called the y-section of E.

Finally we will let $\mathcal{C} \subset \mathcal{P}$ denote the collection of **cylinder sets**, that is sets of the form

$$A \times Y \quad A \in \mathcal{F}$$

or

$$X \times B, \quad B \in \mathcal{G}.$$

In other words, an element of $\mathcal{P}$ is a cylinder set when one of the factors is the whole space.

Theorem 5.12.1 .

- *$\mathcal{F} \times \mathcal{G}$ is generated by the collection of cylinder sets $\mathcal{C}$.*

- $\mathcal{F} \times \mathcal{G}$ *is the smallest σ-field on $X \times Y$ such that the projections*

$$\text{pr}_X : X \times Y \to X \qquad \text{pr}_X(x,y) = x$$
$$\text{pr}_Y : X \times Y \to Y \qquad \text{pr}_Y(x,y) = y$$

  are measurable maps.

- *For each $E \in \mathcal{F} \times \mathcal{G}$ and all $x \in X$ the x-section E_x of E belongs to $\mathcal{G}$ and for all $y \in Y$ the y-section E_y of E belongs to $\mathcal{F}$.*

Proof. $A \times B = (A \times Y) \cap (X \times B)$ so any σ-field containing $\mathcal{C}$ must also contain $\mathcal{P}$. This proves the first item.

Since $\text{pr}_X^{-1}(A) = A \times Y$, the map pr_X is measurable, and similarly for Y. But also, any σ-field containing all $A \times Y$ and $X \times B$ must contain $\mathcal{P}$ by what we just proved. This proves the second item.

As to the third item, any set E of the form $A \times B$ has the desired section properties, since its x section is B if $x \in A$ or the empty set if $x \notin A$. Similarly for its y sections. So let $\mathcal{H}$ denote the collection of subsets E which have the property that all $E_x \in \mathcal{G}$ and all $E_y \in \mathcal{F}$. If we show that $\mathcal{H}$ is a σ-field we are done.

Now $E_x^c = (E_x)^c$ and similarly for y, so $\mathcal{G}$ is closed under taking complements. Similarly for countable unions:

$$\left(\bigcup_n E_n \right)_x = \bigcup_n (E_n)_x.$$

QED

5.12.2 π-systems and λ-systems.

Recall that the σ-field $\sigma(\mathcal{C})$ generated by a collection $\mathcal{C}$ of subsets of X is the intersection of all the σ-fields containing $\mathcal{C}$. Sometimes the collection $\mathcal{C}$ is closed under finite intersection. In that case, we call $\mathcal{C}$ a π-system. Examples:

- X is a topological space, and $\mathcal{C}$ is the collection of open sets in X.

- $X = \mathbf{R}$, and $\mathcal{C}$ consists of all half infinite intervals of the form $(-\infty, a]$. We will denote this π system by $\pi(\mathbf{R})$.

A collection $\mathcal{H}$ of subsets of X will be called a λ-system if

1. $X \in \mathcal{H}$,

2. $A, B \in \mathcal{H}$ with $A \cap B = \emptyset \Rightarrow A \cup B \in \mathcal{H}$,

3. $A, B \in \mathcal{H}$ and $B \subset A \Rightarrow (A \setminus B) \in \mathcal{H}$, and

4. $\{A_n\}_1^\infty \subset \mathcal{H}$ and $A_n \nearrow A \Rightarrow A \in \mathcal{H}$.

From items 1) and 3) we see that a λ-system is closed under complementation, and since $\emptyset = X^c$ it contains the empty set. If $\mathcal{B}$ is both a π-system and a λ system, it is closed under any finite union, since $A \cup B = A \cup (B/(A \cap B))$ which is a disjoint union. Any countable union can be written in the form $A = \nearrow A_n$ where the A_n are finite disjoint unions as we have already argued. So we have proved

Proposition 5.12.1 *If $\mathcal{H}$ is both a π-system and a λ-system then it is a σ-field.*

Also, we have

Proposition 5.12.2 [Dynkin's lemma.] *If $\mathcal{C}$ is a π-system, then the σ-field generated by $\mathcal{C}$ is the smallest λ-system containing $\mathcal{C}$.*

Let $\mathcal{M}$ be the σ-field generated by $\mathcal{C}$, and $\mathcal{H}$ the smallest λ-system containing $\mathcal{C}$. So $\mathcal{M} \supset \mathcal{H}$. By the preceding proposition, all we need to do is show that $\mathcal{H}$ is a π-system.
Let

$$\mathcal{H}_1 := \{A | \ A \cap C \in \mathcal{H} \ \forall C \in \mathcal{C}\}.$$

Clearly $\mathcal{H}_1$ is a λ-system containing $\mathcal{C}$, so $\mathcal{H} \subset \mathcal{H}_1$ which means that $A \cap C \in \mathcal{H}$ for all $A \in \mathcal{H}$ and $C \in \mathcal{C}$.
Let

$$\mathcal{H}_2 := \{A | \ A \cap H \in \mathcal{H} \ \ \forall H \in \mathcal{H}\}.$$

$\mathcal{H}_2$ is again a λ-system, and it contains $\mathcal{C}$ by what we have just proved. So $\mathcal{H}_2 \supset \mathcal{H}$, which means that the intersection of two elements of $\mathcal{H}$ is again in $\mathcal{H}$, i.e. $\mathcal{H}$ is a π-system. QED

5.12.3   The monotone class theorem.

Theorem 5.12.2 *Let $\mathbf{B}$ be a class of bounded real valued functions on a space Z satisfying*

1. *$\mathbf{B}$ is a vector space over $\mathbf{R}$.*

2. *The constant function $\mathbf{1}$ belongs to $\mathbf{B}$.*

3. *$\mathbf{B}$ contains the indicator functions $\mathbf{1}_A$ for all A belonging to a π-system $\mathcal{I}$.*

4. *If $\{f_n\}$ is a sequence of non-negative functions in $\mathbf{B}$ and $f_n \nearrow f$ where f is a bounded function on Z, then $f \in \mathbf{B}$.*

Then $\mathbf{B}$ contains every bounded $\mathcal{M}$ measurable function, where $\mathcal{M}$ is the σ-field generated by $\mathcal{I}$.

Proof.   Let $\mathcal{H}$ denote the class of subsets of Z whose indicator functions belong to $\mathbf{B}$. Then $Z \in \mathcal{H}$ by item 2). If $B \subset A$ are both in $\mathcal{H}$, then $\mathbf{1}_{A \setminus B} = \mathbf{1}_A - \mathbf{1}_B$ and so $A \setminus B$ belongs to $\mathcal{H}$ by item 1). Similarly, if $A \cap B = \emptyset$ then $\mathbf{1}_{A \cup B} = \mathbf{1}_A + \mathbf{1}_B$

and so if A and B belong to $\mathcal{H}$ so does $A \cup B$ when $A \cap B = \emptyset$. Finally, condition 4) in the theorem implies condition 4) in the definition of a λ-system. So we have proved that that $\mathcal{H}$ is a λ-system containing $\mathcal{I}$. So by Dynkin's lemma, it contains $\mathcal{M}$.

Now suppose that $0 \leq f \leq K$ is a bounded $\mathcal{M}$ measurable function, where we may take K to be an integer. For each integer $n \geq 0$ divide the interval $[0, K]$ up into subintervals of size 2^{-n}, and let

$$A(n, i) := \{z | i2^{-n} \leq f(z) < (i+1)2^{-n}\}$$

where i ranges from 0 to $K2^n$. Let

$$s_n(z) := \sum_{i=0}^{K2^n} \frac{i}{2^n} \mathbf{1}_{A(n,i)}.$$

Since f is assumed to be $\mathcal{M}$-measurable, each $A(n, i) \in \mathcal{M}$, so by the preceding, and condition 1), $f_n \in \mathbf{B}$. But $0 \leq s_n \nearrow f$, and hence by condition 4), $f \in B$.

For a general bounded $\mathcal{M}$ measurable f, both f^+ and f^- are bounded and $\mathcal{M}$ measurable, and hence by the preceding and condition 1), $f = f^+ - f^- \in \mathbf{B}$.QED

We now want to apply the monotone class theorem to our situation of a product space. So $Z = X \times Y$, where $(X, \mathcal{F})$ and $(Y, \mathcal{G})$ are spaces with σ-fields, and where we take $\mathcal{I} = \mathcal{P}$ to be the π-system consisting of the product sets $A \times B$, $A \in \mathcal{F}$, $B \in \mathcal{G}$.

Proposition 5.12.3 *Let* $\mathbf{B}$ *consist of all bounded real valued functions* f *on* $X \times Y$ *which are* $\mathcal{F} \times \mathcal{G}$-*measurable, and which have the property that*

- *for each* $x \in X$, *the function* $y \mapsto f(x, y)$ *is* $\mathcal{G}$-*measurable, and*

- *for each* $y \in Y$ *the function* $x \mapsto f(x, y)$ *is* $\mathcal{F}$-*measurable.*

Then $\mathbf{B}$ *consists of all bounded* $\mathcal{F} \times \mathcal{G}$ *measurable functions.*

Indeed, $y \mapsto \mathbf{1}_{A \times B}(x, y) = \mathbf{1}_B(y)$ if $x \in A$ and $= 0$ otherwise; and similarly for $x \mapsto \mathbf{1}_{A \times B}(x, y)$. So condition 3) of the monotone class theorem is satisfied, and the other conditions are immediate. Since $\mathcal{F} \times \mathcal{G}$ was defined to be the σ-field generated by $\mathcal{P}$, the proposition is an immediate consequence of the monotone class theorem.

5.12.4   Fubini for finite measures and bounded functions.

Let $(X, \mathcal{F}, m)$ and $(Y, \mathcal{G}, n)$ be measure spaces with $m(X) < \infty$ and $n(Y) < \infty$. For every bounded $\mathcal{F} \times \mathcal{G}$-measurable function f, we know that the function

$$f(x, \cdot) : y \mapsto f(x, y)$$

is bounded and $\mathcal{G}$ measurable. Hence it has an integral with respect to the measure n, which we will denote by

$$\int_Y f(x,y)n(dy).$$

This is a bounded function of x (which we will prove to be $\mathcal{F}$ measurable in just a moment). Similarly we can form

$$\int_X f(x,y)m(dx)$$

which is a function of y.

Proposition 5.12.4 *Let* **B** *denote the space of bounded $\mathcal{F} \times \mathcal{G}$ measurable functions such that*

- *$\int_Y f(x,y)n(dy)$ is a $\mathcal{F}$ measurable function on X,*

- *$\int_X f(x,y)m(dx)$ is a $\mathcal{G}$ measurable function on Y and*

-
$$\int_X \left(\int_Y f(x,y)n(dy) \right) m(dx) = \int_Y \left(\int_X f(x,y)m(dx) \right) n(dy). \quad (5.27)$$

Then **B** *consists of all bounded $\mathcal{F} \times \mathcal{G}$ measurable functions.*

Proof. We have verified that the first two items hold for $\mathbf{1}_{A \times B}$. Both sides of (5.27) equal $m(A)n(B)$ as is clear from the proof of Proposition 5.12.3. So conditions 1-3 of the monotone class theorem are clearly satisfied, and condition 4) is a consequence of two double applications of the monotone convergence theorem. QED

Now for any $C \in \mathcal{F} \times \mathcal{G}$ we define

$$(m \times n)(C) := \int_X \left(\int_Y \mathbf{1}_C(x,y)n(dy) \right) m(dx) = \int_Y \left(\int_X \mathbf{1}_C(x,y)m(dx) \right) n(dy),$$
$$(5.28)$$

both sides being equal on account of the preceding proposition. This measure assigns the value $m(A)n(B)$ to any set $A \times B \in \mathcal{P}$, and since $\mathcal{P}$ generates $\mathcal{F} \times \mathcal{G}$ as a sigma field, any two measures which agree on $\mathcal{P}$ must agree on $\mathcal{F} \times \mathcal{G}$. Hence $m \times n$ is the unique measure which assigns the value $m(A)n(B)$ to sets of $\mathcal{P}$.

Furthermore, we know that

$$\int_{X \times Y} f(x,y)(m \times n) = \int_X \left(\int_Y f(x,y)n(dy) \right) m(dx) = \int_Y \left(\int_X f(x,y)m(dx) \right) n(dy)$$
$$(5.29)$$

is true for functions of the form $\mathbf{1}_{A \times B}$ and hence by the monotone class theorem it is true for all bounded functions which are measurable relative to $\mathcal{F} \times \mathcal{G}$.

The above assertions are the content of Fubini's theorem for bounded measures and functions. We summarize:

Theorem 5.12.3 *Let $(X, \mathcal{F}, m)$ and $(Y, \mathcal{G}, n)$ be measure spaces with $m(X) < \infty$ and $n(Y) < \infty$. There exists a unique measure on $\mathcal{F} \times \mathcal{G}$ with the property that*

$$(m \times n)(A \times B) = m(A)n(B) \quad \forall\ A \times B \in \mathcal{P}.$$

For any bounded $\mathcal{F} \times \mathcal{G}$ measurable function, the double integral is equal to the iterated integral in the sense that (5.29) holds.

5.12.5   Extensions to unbounded functions and to σ-finite measures.

Suppose that we temporarily keep the condition that $m(X) < \infty$ and $n(Y) < \infty$. Let f be any non-negative $\mathcal{F} \times \mathcal{G}$-measurable function. We know that (5.29) holds for all bounded measurable functions, in particular for all simple functions. We know that we can find a sequence of simple functions s_n such that $s_n \nearrow f$. Hence by several applications of the monotone convergence theorem, we know that (5.29) is true for all non-negative $\mathcal{F} \times \mathcal{G}$-measurable functions in the sense that all three terms are infinite together, or finite together and equal. Now we have agreed to call a $\mathcal{F} \times \mathcal{G}$-measurable function f integrable if and only if f^+ and f^- have finite integrals. In this case (5.29) holds.

A measure space $(X, \mathcal{F}, m)$ is called σ-**finite** if $X = \bigcup_n X_n$ where $m(X_n) < \infty$. In other words, X is σ-finite if it is a countable union of finite measure spaces. As usual, we can then write X as a countable union of disjoint finite measure spaces. So if X and Y are σ-finite, we can write the various integrals that occur in (5.29) as sums of integrals which occur over finite measure spaces. A bit of standard argumentation shows that Fubini continues to hold in this case.

If X or Y is not σ-finite, or, even in the finite case, if f is not non-negative or $m \times n$ integrable, then Fubini need not hold. I hope to present the standard counter-examples in the problem set.

Chapter 6

The Daniell integral.

Daniell's idea was to take the axiomatic properties of the integral as the starting point and develop integration for broader and broader classes of functions. Then derive measure theory as a consequence. Much of the presentation here is taken from the book *Abstract Harmonic Analysis* by Lynn Loomis. Some of the lemmas, propositions and theorems indicate the corresponding sections in Loomis's book.

6.1   The Daniell Integral

Let L be a vector space of *bounded* real valued functions on a set S closed under $\wedge$ and $\vee$. For example, S might be a complete metric space, and L might be the space of continuous functions of compact support on S.

A map
$$I : L \to \mathbf{R}$$

is called an **Integral** if

1. I is linear: $I(af + bg) = aI(f) + bI(g)$

2. I is non-negative: $f \geq 0 \Rightarrow I(f) \geq 0$ or equivalently $f \geq g \Rightarrow I(f) \geq I(g)$.

3. $f_n \searrow 0 \Rightarrow I(f_n) \searrow 0$.

For example, we might take $S = \mathbf{R}^n$, $L =$ the space of continuous functions of compact support on $\mathbf{R}^n$, and I to be the Riemann integral. The first two items on the above list are clearly satisfied. As to the third, we recall Dini's lemma from the notes on metric spaces, which says that a sequence of continuous functions of compact support $\{f_n\}$ on a metric space which satisfies $f_n \searrow 0$ actually converges uniformly to 0. Furthermore the supports of the f_n are all contained in a fixed compact set - for example the support of f_1. This establishes the third item.

The plan is now to successively increase the class of functions on which the integral is defined.

Define

$$U := \{\text{limits of monotone non-decreasing sequences of elements of } L\}.$$

We will use the word "increasing" as synonymous with "monotone non-decreasing" so as to simplify the language.

Lemma 6.1.1 *If f_n is an increasing sequence of elements of L and if $k \in L$ satisfies $k \le \lim f_n$ then $\lim I(f_n) \ge I(k)$.*

Proof. If $k \in L$ and $\lim f_n \ge k$, then

$$f_n \wedge k \le k \quad \text{and } f_n \ge f_n \wedge k$$

so $I(f_n) \ge I(f_n \wedge k)$ while

$$[k - (f_n \wedge k)] \searrow 0$$

so

$$I([k - f_n \wedge k]) \searrow 0$$

by 3) or

$$I(f_n \wedge k) \nearrow I(k).$$

Hence $\lim I(f_n) \ge \lim I(f_n \wedge k) = I(k)$. QED

Lemma 6.1.2 [12C] *If $\{f_n\}$ and $\{g_n\}$ are increasing sequences of elements of L and $\lim g_n \le \lim f_n$ then $\lim I(g_n) \le \lim I(f_n)$.*

Proof. Fix m and take $k = g_m$ in the previous lemma. Then $I(g_m) \le \lim I(f_n)$. Now let $m \to \infty$. QED

Thus

$$f_n \nearrow f \text{ and } g_n \nearrow f \quad \Rightarrow \lim I(f_n) = \lim I(g_n)$$

so we may extend I to U by setting

$$I(f) := \lim I(f_n) \quad \text{for} \quad f_n \nearrow f.$$

If $f \in L$, this coincides with our original I, since we can take $g_n = f$ for all n in the preceding lemma.

We have now extended I from L to U. The next lemma shows that if we now start with I on U and apply the same procedure again, we do not get any further.

Lemma 6.1.3 [12D] *If $f_n \in U$ and $f_n \nearrow f$ then $f \in U$ and $I(f_n) \nearrow I(f)$.*

Proof. For each fixed n choose $g_n^m \nearrow_m f_n$. Set

$$h_n := g_1^n \vee \cdots \vee g_n^n$$

so

$$h_n \in L \quad \text{and } h_n \text{ is increasing}$$

with

$$g_i^n \leq h_n \leq f_n \quad \text{for } i \leq n.$$

Let $n \to \infty$. Then

$$f_i \leq \lim h_n \leq f.$$

Now let $i \to \infty$. We get

$$f \leq \lim h_n \leq f.$$

So we have written f as a limit of an increasing sequence of elements of L, So $f \in U$. Also

$$I(g_i^n) \leq I(h_n) \leq I(f)$$

so letting $n \to \infty$ we get

$$I(f_i) \leq I(f) \leq \lim I(f_n)$$

so passing to the limits gives $I(f) = \lim I(f_n)$. QED

We have

$$I(f + g) = I(f) + I(g) \quad \text{for } f, g \in U.$$

Define

$$-U := \{-f \mid f \in U\}$$

and

$$I(f) := -I(-f) \quad f \in -U.$$

If $f \in U$ and $-f \in U$ then $I(f) + I(-f) = I(f - f) = I(0) = 0$ so $I(-f) = -I(f)$ in this case. So the definition is consistent.

$-U$ is closed under monotone decreasing limits. etc.

If $g \in -U$ and $h \in U$ with $g \leq h$ then $-g \in U$ so $h - g \in U$ and $h - g \geq 0$ so $I(h) - I(g) = I(h + (-g)) = I(h - g) \geq 0$.

A function f is called *I-summable* if for every $\epsilon > 0$, $\exists g \in -U$, $h \in U$ with

$$g \leq f \leq h, \quad |I(g)| < \infty, \ |I(h)| < \infty \ \text{ and } I(h - g) \leq \epsilon.$$

For such f define

$$I(f) = \text{glb } I(h) = \text{lub } I(g).$$

If $f \in U$ take $h = f$ and $f_n \in L$ with $f_n \nearrow f$. Then $-f_n \in L \subset U$ so $f_n \in -U$. If $I(f) < \infty$ then we can choose n sufficiently large so that $I(f) - I(f_n) < \epsilon$. The space of summable functions is denoted by $\overline{L}_1$. It is clearly a vector space, and I satisfies conditions 1) and 2) above, i.e. is linear and non-negative.

Theorem 6.1.1 [12G] Monotone convergence theorem. $f_n \in \overline{L}_1$, $f_n \nearrow f$ *and* $\lim I(f_n) < \infty \Rightarrow f \in \overline{L}_1$ *and* $I(f) = \lim I(f_n)$.

Proof. Replacing f_n by $f_n - f_0$ we may assume that $f_0 = 0$. Choose

$$h_n \in U, \text{ such that } f_n - f_{n-1} \leq h_n \text{ and } I(h_n) \leq I(f_n - f_{n-1}) + \frac{\epsilon}{2^n}.$$

Then

$$f_n \leq \sum_1^n h_i \quad \text{and} \quad \sum_{i=1}^n I(h_i) \leq I(f_n) + \epsilon.$$

Since U is closed under monotone increasing limits,

$$h := \sum_{i=1}^\infty h_i \in U, \quad f \leq h \quad \text{and } I(h) \leq \lim I(f_n) + \epsilon.$$

Since $f_m \in \overline{L}_1$ we can find a $g_m \in -U$ with $I(f_m) - I(g_m) < \epsilon$ and hence for m large enough $I(h) - I(g_m) < 2\epsilon$. So $f \in \overline{L}_1$ and $I(f) = \lim I(f_n)$. QED

6.2 Monotone class theorems.

A collection of functions which is closed under monotone increasing and monotone decreasing limits is called a **monotone class**. $\mathcal{B}$ is defined to be the smallest monotone class containing L.

Lemma 6.2.1 *Let* $h \leq k$. *If* $\mathcal{M}$ *is a monotone class which contains* $(g \vee h) \wedge k$ *for every* $g \in L$, *then* $\mathcal{M}$ *contains all* $(f \vee h) \wedge k$ *for all* $f \in \mathcal{B}$.

Proof. The set of f such that $(f \vee h) \wedge k \in \mathcal{M}$ is a monotone class containing L by the distributive laws.QED

Taking $h = k = 0$ this says that the smallest monotone class containing L^+, the set of non-negative functions in L, is the set $\mathcal{B}^+$, the set of non-negative functions in $\mathcal{B}$.

Here is a series of monotone class theorem style arguments:

Theorem 6.2.1 $f, g \in \mathcal{B} \Rightarrow af + bg \in \mathcal{B}, f \vee g \in \mathcal{B}$ *and* $f \wedge g \in \mathcal{B}$.

For $f \in \mathcal{B}$, let
$$\mathcal{M}(f) := \{g \in \mathcal{B} | f + g, f \vee g, f \wedge g \in \mathcal{B}\}.$$

$\mathcal{M}(f)$ is a monotone class. If $f \in L$ it includes all of L, hence all of $\mathcal{B}$. But

$$g \in \mathcal{M}(f) \iff f \in \mathcal{M}(g).$$

So $L \subset \mathcal{M}(g)$ for any $g \in \mathcal{B}$, and since it is a monotone class $\mathcal{B} \subset \mathcal{M}(g)$. This says that $f, g \in \mathcal{B} \Rightarrow f + g \in \mathcal{B}$, $f \wedge g \in \mathcal{B}$ and $f \vee g \in \mathcal{B}$. Similarly, let $\mathcal{M}$ be the class of functions for which $cf \in \mathcal{B}$ for all real c. This is a monotone class containing L hence contains $\mathcal{B}$. QED

Lemma 6.2.2 *If $f \in \mathcal{B}$ there exists a $g \in U$ such that $f \leq g$.*

Proof. The limit of a monotone increasing sequence of functions in U belongs to U. Hence the set of f for which the lemma is true is a monotone class which contains L. hence it contains $\mathcal{B}$. QED

A function f is **L-bounded** if there exists a $g \in L^+$ with $|f| \leq g$. A class $\mathcal{F}$ of functions is said to be L-monotone if $\mathcal{F}$ is closed under monotone limits of L-bounded functions.

Theorem 6.2.2 *The smallest L-monotone class including L^+ is $\mathcal{B}^+$.*

Proof. Call this smallest family $\mathcal{F}$. If $g \in L^+$, the set of all $f \in \mathcal{B}^+$ such that $f \wedge g \in \mathcal{F}$ form a monotone class containing L^+, hence containing $\mathcal{B}^+$ hence equal to $\mathcal{B}^+$. If $f \in \mathcal{B}^+$ and $f \leq g$ then $f \wedge g = f \in \mathcal{F}$. So $\mathcal{F}$ contains all L bounded functions belonging to $\mathcal{B}^+$. Let $f \in \mathcal{B}^+$. By the lemma, choose $g \in U$ such that $f \leq g$, and choose $g_n \in L^+$ with $g_n \nearrow g$. Then $f \wedge g_n \leq g_n$ and so is L bounded, so $f \wedge g_n \in \mathcal{F}$. Since $(f \wedge g_n) \to f$ we see that $f \in \mathcal{F}$. So

$$\mathcal{B}^+ \subset \mathcal{F}.$$

We know that $\mathcal{B}^+$ is a monotone class, in particular an L-monotone class. Hence $\mathcal{F} = \mathcal{B}^+$. QED

Define

$$L^1 := \overline{L}_1 \cap \mathcal{B}.$$

Since $\overline{L}_1$ and $\mathcal{B}$ are both closed under the lattice operations,

$$f \in L^1 \Rightarrow f^\pm \in L^1 \Rightarrow |f| \in L^1.$$

Theorem 6.2.3 *If $f \in \mathcal{B}$ then $f \in L^1 \Leftrightarrow \exists g \in L^1$ with $|f| \leq g$.*

We have proved $\Rightarrow$: simply take $g = |f|$. For the converse we may assume that $f \geq 0$ by applying the result to f^+ and f^-. The family of all $h \in \mathcal{B}^+$ such that $h \wedge g \in L^1$ is monotone and includes L^+ so includes $\mathcal{B}^+$. So $f = f \wedge g \in L^1$. QED

Extend I to all of $\mathcal{B}^+$ be setting it $= \infty$ on functions which do not belong to L^1.

6.3 Measure.

Loomis calls a set A **integrable** if $\mathbf{1}_A \in \mathcal{B}$. The monotone class properties of $\mathcal{B}$ imply that the integrable sets form a σ-field. Then define

$$\mu(A) := \int \mathbf{1}_A$$

and the monotone convergence theorem guarantees that μ is a measure.

Add **Stone's axiom**

$$f \in L \Rightarrow f \wedge \mathbf{1} \in L.$$

Then the monotone class property implies that this is true with L replaced by $\mathcal{B}$.

Theorem 6.3.1 $f \in \mathcal{B}$ and $a > 0 \Rightarrow$ then

$$A_a := \{p | f(p) > a\}$$

is an integrable set. If $f \in L^1$ then

$$\mu(A_a) < \infty.$$

Proof. Let

$$f_n := [n(f - f \wedge a)] \wedge 1 \in \mathcal{B}.$$

Then

$$f_n(x) = \begin{cases} 1 & \text{if} & f(x) \geq a + \frac{1}{n} \\ 0 & \text{if} & f(x) \leq a \\ n(f(x) - a) & \text{if} & a < f(x) < a + \frac{1}{n} \end{cases}.$$

We have

$$f_n \nearrow \mathbf{1}_{A_a}$$

so $\mathbf{1}_{A_a} \in \mathcal{B}$ and $0 \leq \mathbf{1}_{A_a} \leq \frac{1}{a} f^+$. QED

Theorem 6.3.2 If $f \geq 0$ and A_a is integrable for all $a > 0$ then $f \in \mathcal{B}$.

Proof. For $\delta > 1$ define

$$A_m^\delta := \{x | \delta^m < f(x) \leq \delta^{m+1}\}$$

for $m \in \mathbb{Z}$ and

$$f_\delta := \sum_m \delta^m \mathbf{1}_{A_m^\delta}.$$

Each $f_\delta \in \mathcal{B}$. Take

$$\delta_n = 2^{2^{-n}}.$$

Then each successive subdivision divides the previous one into "octaves" and $f_{\delta_m} \nearrow f$. QED

Also

$$f_\delta \leq f \leq \delta f_\delta$$

and

$$I(f_\delta) = \sum \delta^n \mu(A_m^\delta) = \int f_\delta d\mu.$$

So we have

$$I(f_\delta) \leq I(f) \leq \delta I(f_\delta)$$

and

$$\int f_\delta d\mu \le \int f d\mu \le \delta \int f_\delta d\mu.$$

So if either of $I(f)$ or $\int f d\mu$ is finite they both are and

$$\left| I(f) - \int f d\mu \right| \le (\delta - 1)I(f_\delta) \le (\delta - 1)I(f).$$

So

$$\int f d\mu = I(f).$$

If $f \in \mathcal{B}^+$ and $a > 0$ then

$$\{x | f(x)^a > b\} = \{x | f(x) > b^{\frac{1}{a}}\}.$$

So $f \in \mathcal{B}^+ \Rightarrow f^a \in \mathcal{B}^+$ and hence the product of two elements of $\mathcal{B}^+$ belongs to $\mathcal{B}^+$ because

$$fg = \frac{1}{4} \left[(f+g)^2 - (f-g)^2 \right].$$

6.4 Hölder, Minkowski , L^p and L^q.

The numbers $p, q > 1$ are called **conjugate** if

$$\frac{1}{p} + \frac{1}{q} = 1.$$

This is the same as

$$pq = p + q$$

or

$$(p - 1)(q - 1) = 1.$$

This last equation says that if

$$y = x^{p-1}$$

then

$$x = y^{q-1}.$$

The area under the curve $y = x^{p-1}$ from 0 to a is

$$A = \frac{a^p}{p}$$

while the area between the same curve and the y-axis up to $y = b$

$$B = \frac{b^q}{q}.$$

Suppose $b < a^{p-1}$ to fix the ideas. Then area ab of the rectangle is less than $A + B$ or

$$\frac{a^p}{p} + \frac{b^q}{q} \geq ab$$

with equality if and only if $b = a^{p-1}$. Replacing a by $a^{\frac{1}{p}}$ and b by $b^{\frac{1}{q}}$ gives

$$a^{\frac{1}{p}} b^{\frac{1}{q}} \leq \frac{a}{p} + \frac{b}{q}.$$

Let L^p denote the space of functions such that $|f|^p \in L^1$. For $f \in L^p$ define

$$\|f\|_p := \left(\int |f|^p d\mu \right)^{\frac{1}{p}}.$$

We will soon see that if $p \geq 1$ this is a (semi-)norm.

If $f \in L^p$ and $g \in L^q$ with $\|f\|_p \neq 0$ and $\|g\|_q \neq 0$ take

$$a = \frac{|f|^p}{\|f\|_p}, \quad b = \frac{|g|^q}{\|g\|_q}$$

as functions. Then

$$\int (|f||g|) d\mu \leq \|f\|_p \|g\|_q \left(\frac{1}{p} \frac{1}{\|f\|_p^p} \int |f|^p d\mu + \frac{1}{q} \frac{1}{\|g\|_q^q} \int |g|^q d\mu \right) = \|f\|_p \|g\|_q.$$

This shows that the left hand side is integrable and that

$$\left| \int fg d\mu \right| \leq \|f\|_p \|g\|_q \tag{6.1}$$

which is known as **Hölder's inequality**. (If either $\|f\|_p$ or $\|g\|_q = 0$ then $fg = 0$ a.e. and Hölder's inequality is trivial.)

We write

$$(f, g) := \int fg d\mu.$$

Proposition 6.4.1 [Minkowski's inequality] *If $f, g \in L^p$, $p \geq 1$ then $f + g \in L^p$ and*

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

For $p = 1$ this is obvious. If $p > 1$

$$|f + g|^p \leq [2 \max(|f|, |g|)]^p \leq 2^p [|f|^p + |g|^p]$$

implies that $f + g \in L^p$. Write

$$\|f + g\|_p^p \leq I(|f + g|^{p-1}|f|) + I(|f + g|^{p-1}|g|).$$

Now

$$q(p - 1) = qp - q = p$$

so
$$|f + g|^{p-1} \in L_q$$

and its $\| \cdot \|_q$ norm is

$$I(|f + g|^p)^{\frac{1}{q}} = I(|f + q|^p)^{1 - \frac{1}{p}} = I(|f + g|^p)^{\frac{p-1}{p}} = \|f + g\|_p^{p-1}.$$

So we can write the preceding inequality as

$$\|f + g\|_p^p \leq (|f|, |f + g|^{p-1}) + (|g|, |f + g|^{p-1})$$

and apply Hölder's inequality to conclude that

$$\|f + g\|^p \leq \|f + g\|^{p-1}(\|f\|_p + \|g\|_p).$$

We may divide by $\|f + g\|_p^{p-1}$ to get Minkowski's inequality unless $\|f + g\|_p = 0$ in which case it is obvious. QED

Theorem 6.4.1 L^p *is complete.*

Proof. Suppose $f_n \geq 0$, $f_n \in L^p$, and $\sum \|f_n\|_p < \infty$ Then

$$k_n := \sum_1^n f_j \in L^p$$

by Minkowski and since $k_n \nearrow f$ we have $|k_n|^p \nearrow f^p$ and hence by the monotone convergence theorem $f := \sum_{j=1}^\infty f_n \in L^p$ and $\|f\|_p = \lim \|k_n\|_p \leq \sum \|f_j\|_p$.

Now let $\{f_n\}$ be any Cauchy sequence in L^p. By passing to a subsequence we may assume that

$$\|f_{n+1} - f_n\|_p < \frac{1}{2^n}.$$

So $\sum_n^\infty |f_{i+1} - f_i| \in L^p$ and hence

$$g_n := f_n - \sum_n^\infty |f_{i+1} - f_i| \in L^p \quad \text{and} \quad h_n := f_n + \sum_n^\infty |f_{i+1} - f_i| \in L^p.$$

We have
$$g_{n+1} - g_n = f_{n+1} - f_n + |f_{n+1} - f_n| \geq 0$$

so g_n is increasing and similarly h_n is decreasing. Hence $f := \lim g_n \in L^p$ and $\|f - f_n\|_p \leq \|h_n - g_n\|_p \leq 2^{-n+2} \to 0$. So the subsequence has a limit which then must be the limit of the original sequence. QED

Proposition 6.4.2 L *is dense in* L^p *for any* $1 \leq p < \infty$.

Proof. For $p = 1$ this was a defining property of L^1. More generally, suppose that $f \in L^p$ and that $f \geq 0$. Let

$$A_n := \{x : \frac{1}{n} < f(x) < n\},$$

and let
$$g_n := f \cdot \mathbf{1}_{A_n}.$$

Then $(f-g_n) \searrow 0$ as $n \to \infty$. Choose n sufficiently large so that $\|f-g_n\|_p < \epsilon/2$. Since
$$0 \leq g_n \leq n\mathbf{1}_{A_n} \quad \text{and} \quad \mu(A_n) < n^p I(|f|^p) < \infty$$

we conclude that
$$g_n \in L^1.$$

Now choose $h \in L^+$ so that
$$\|h - g_n\|_1 < \left(\frac{\epsilon}{2n}\right)^p$$

and also so that $h \leq n$. Then
$$
\begin{aligned}
\|h - g_n\|_p &= (I(|h - g_n|^p))^{1/p} \\
&= \left(I(|h - g_n|^{p-1}|h - g_n|)\right)^{1/p} \\
&\leq \left(I(n^{p-1}|h - g_n|)\right)^{1/p} \\
&= \left(n^{p-1}\|h - g_n\|_1\right)^{1/p} \\
&< \epsilon/2.
\end{aligned}
$$

So by the triangle inequality $\|f - h\| < \epsilon$. QED

In the above, we have not bothered to pass to the quotient by the elements of norm zero. In other words, we have not identified two functions which differ on a set of measure zero. We will continue with this ambiguity. But equally well, we could change our notation, and use L^p to denote the quotient space (as we did earlier in class) and denote the space before we pass to the quotient by $\mathcal{L}^p$ to conform with our earlier notation. I will continue to be sloppy on this point, in conformity to Loomis' notation.

6.5 $\|\cdot\|_\infty$ is the essential sup norm.

Suppose that $f \in \mathcal{B}$ has the property that it is equal almost everywhere to a function which is bounded above. We call such a function **essentially bounded** (from above). We can then define the **essential least upper bound** of f to be the smallest number which is an upper bound for a function which differs from f on a set of measure zero. If $|f|$ is essentially bounded, we denote its essential least upper bound by $\|f\|_\infty$. Otherwise we say that $\|f\|_\infty = \infty$. We let $\mathcal{L}^\infty$ denote the space of $f \in \mathcal{B}$ which have $\|f\|_\infty < \infty$. It is clear that $\|\cdot\|_\infty$ is a semi-norm on this space. The justification for this notation is

Theorem 6.5.1 [14G] *If $f \in L^p$ for some $p > 0$ then*

$$\|f\|_\infty = \lim_{q \to \infty} \|f\|_q. \tag{6.2}$$

Remark. In the statement of the theorem, both sides of (6.2) are allowed to be ∞.

Proof. If $\|f\|_\infty = 0$, then $\|f\|_q = 0$ for all $q > 0$ so the result is trivial in this case. So let us assume that $\|f\|_\infty > 0$ and let a be any positive number smaller that $\|f\|_\infty$. In other words,

$$0 < a < \|f\|_\infty.$$

Let

$$A_a := \|x : |f(x)| > a\}.$$

This set has positive measure by the choice of a, and its measure is finite since $f \in L^p$. Also

$$\|f\|_q \geq \left(\int_{A_a} |f|^q \right)^{1/q} \geq a\mu(A_a)^{1/q}.$$

Letting $q \to \infty$ gives

$$\lim_{q \to \infty} \inf \|f\|_q \geq a$$

and since a can be any number $< \|f\|_\infty$ we conclude that

$$\lim_{q \to \infty} \inf \|f\|_q \geq \|f\|_\infty.$$

So we need to prove that

$$\lim \|f\|_q \leq \|f\|_\infty.$$

This is obvious if $\|f\|_\infty = \infty$. So suppose that $\|f\|_\infty$ is finite. Then for $q > p$ we have

$$|f|^q \leq |f|^p (\|f\|_\infty)^{q-p}$$

almost everywhere. Integrating and taking the q-th root gives

$$\|f\|_q \leq (\|f\|_p)^{\frac{p}{q}} (\|f\|_\infty)^{1 - \frac{p}{q}}.$$

Letting $q \to \infty$ gives the desired result. QED

6.6 The Radon-Nikodym Theorem.

Suppose we are given two integrals, I and J on the same space L. That is, both I and J satisfy the three conditions of linearity, positivity, and the monotone limit property that went into our definition of the term "integral". We say that J is **absolutely continuous** with respect to I if every set which is I null (i.e. has measure zero with respect to the measure associated to I) is J null.

The integral I is said to be **bounded** if

$$I(\mathbf{1}) < \infty,$$

or, what amounts to the same thing, that

$$\mu_I(S) < \infty$$

where μ_I is the measure associated to I.

We will first formulate the Radon-Nikodym theorem for the case of bounded integrals, where there is a very clever proof due to von-Neumman which reduces it to the Riesz representation theorem in Hilbert space theory.

Theorem 6.6.1 [Radon-Nikodym] *Let I and J be bounded integrals, and suppose that J is absolutely continuous with respect to I. Then there exists an element $f_0 \in \mathcal{L}^1(I)$ such that*

$$J(f) = I(ff_0) \quad \forall \ f \in \mathcal{L}^1(J). \tag{6.3}$$

The element f_0 is unique up to equality almost everywhere (with respect to μ_I).

Proof.(After von-Neumann.) Consider the linear function

$$K := I + J$$

on L. Then K satisfies all three conditions in our definition of an integral, and in addition is bounded. We know from the case $p = 2$ of Theorem 6.4.1 that $L^2(K)$ is a (real) Hilbert space. (Assume for this argument that we have passed to the quotient space so an element of $L^2(K)$ is an equivalence class of of functions.) The fact that K is bounded, says that $\mathbf{1} := \mathbf{1}_S \in L^2(K)$. If $f \in L^2(K)$ then the Cauchy-Schwartz inequality says that

$$K(|f|) = K(|f| \cdot \mathbf{1}) = (|f|, \mathbf{1})_{2,K} \leq \|f\|_{2,K} \|\mathbf{1}\|_{2,K} < \infty$$

so $|f|$ and hence f are elements of $L^1(K)$.

Furthermore,

$$|J(f)| \leq J(|f|) \leq K(|f|) \leq \|f\|_{2,K} \|\mathbf{1}\|_{2,K}$$

for all $f \in L$. Since we know that L is dense in $L^2(K)$ by Proposition 6.4.2, J extends to a unique continuous linear functional on $L^2(K)$. We conclude from the real version of the Riesz representation theorem, that there exists a unique $g \in L^2(K)$ such that

$$J(f) = (f, g)_{2,K} = K(fg).$$

If A is any subset of S of positive measure, then $J(\mathbf{1}_A) = K(\mathbf{1}_A g)$ so g is non-negative. (More precisely, g is equivalent almost everywhere to a function which is non-negative.) We obtain inductively

$$
\begin{aligned}
J(f) &= K(fg) = \\
I(fg) + J(fg) &= I(fg) + I(fg^2) + J(fg^2) = \\
&\vdots \\
&= I\left(f \cdot \sum_{i=1}^n g^i\right) + J(fg^n).
\end{aligned}
$$

Let N be the set of all x where $g(x) \geq 1$. Taking $f = \mathbf{1}_N$ in the preceding string of equalities shows that

$$J(\mathbf{1}_N) \geq nI(\mathbf{1}_N).$$

Since n is arbitrary, we have proved

Lemma 6.6.1 *The set where $g \geq 1$ has I measure zero.*

We have not yet used the assumption that J is absolutely continuous with respect to I. Let us now use this assumption to conclude that N is also J-null. This means that if $f \geq 0$ and $f \in L^1(J)$ then $fg^n \searrow 0$ almost everywhere (J), and hence by the dominated convergence theorem

$$J(fg^n) \searrow 0.$$

Plugging this back into the above string of equalities shows (by the monotone convergence theorem for I) that

$$f \sum_{i=1}^{\infty} g^n$$

converges in the $L^1(I)$ norm to $J(f)$. In particular, since $J(\mathbf{1}) < \infty$, we may take $f = \mathbf{1}$ and conclude that $\sum_{i=1}^{\infty} g^i$ converges in $L^1(I)$. So set

$$f_0 := \sum_{i=1}^{\infty} g^i \in L^1(I).$$

We have

$$f_0 = \frac{1}{1-g} \quad \text{almost everywhere}$$

so

$$g = \frac{f_0 - 1}{f_0} \quad \text{almost everywhere}$$

and

$$J(f) = I(ff_0)$$

for $f \geq 0$, $f \in L^1(J)$. By breaking any $f \in L^1(J)$ into the difference of its positive and negative parts, we conclude that (6.3) holds for all $f \in L^1(J)$. The uniqueness of f_0 (almost everywhere (I)) follows from the uniqueness of g in $L^2(K)$. QED

The Radon Nikodym theorem can be extended in two directions. First of all, let us continue with our assumption that I and J are bounded, but drop the absolute continuity requirement. Let us say that an integral H is **absolutely singular** with respect to I if there is a set N of I-measure zero such that $J(h) = 0$ for any h vanishing on N.

Let us now go back to Lemma 6.6.1. Define J_{sing} by

$$J_{sing}(f) = J(\mathbf{1}_N f).$$

Then J_{sing} is singular with respect to I, and we can write

$$J = J_{cont} + J_{sing}$$

where

$$J_{cont} = J - J_{sing} = J(\mathbf{1}_{N^c} \cdot).$$

Then we can apply the rest of the proof of the Radon Nikodym theorem to J_{cont} to conclude that

$$J_{cont}(f) = I(ff_0)$$

where $f_0 = \sum_{i=1}^{\infty} (\mathbf{1}_{N^c} g)^i$ is an element of $L^1(I)$ as before. In particular, J_{cont} is absolutely continuous with respect to I.

A second extension is to certain situations where S is not of finite measure. We say that a function f is **locally** L^1 if $f\mathbf{1}_A \in L^1$ for every set A with $\mu(A) < \infty$. We say that S is σ-**finite** with respect to μ if S is a countable union of sets of finite μ measure. This is the same as saying that $\mathbf{1} = \mathbf{1}_S \in \mathcal{B}$. If S is σ-finite then it can be written as a disjoint union of sets of finite measure. If S is σ-finite with respect to both I and J it can be written as the disjoint union of countably many sets which are both I and J finite. So if J is absolutely continuous with respect I, we can apply the Radon-Nikodym theorem to each of these sets of finite measure, and conclude that there is an f_0 which is locally L^1 with respect to I, such that $J(f) = I(ff_0)$ for all $f \in L^1(J)$, and f_0 is unique up to almost everywhere equality.

6.7   The dual space of L^p.

Recall that Hölder's inequality (6.1) says that

$$\left| \int fg d\mu \right| \leq \|f\|_p \|g\|_q$$

if $f \in L^p$ and $g \in L^q$ where

$$\frac{1}{p} + \frac{1}{q} = 1.$$

For the rest of this section we will assume without further mention that this relation between p and q holds. Hölder's inequality implies that we have a map from

$$L^q \to (L^p)^*$$

sending $g \in L^q$ to the continuous linear function on L^p which sends

$$f \mapsto I(fg) = \int fg d\mu.$$

Furthermore, Hölder's inequality says that the norm of this map from $L^q \to (L^p)^*$ is ≤ 1. In particular, this map is injective.

The theorem we want to prove is that under suitable conditions on S and I (which are more general even that σ-finiteness) this map is surjective for $1 \leq p < \infty$.

We will first prove the theorem in the case where $\mu(S) < \infty$, that is when I is a bounded integral. For this we will will need a lemma:

6.7.1 The variations of a bounded functional.

Suppose we start with an arbitrary L and I. For each $1 \leq p \leq \infty$ we have the norm $\| \cdot \|_p$ on L which makes L into a real normed linear space. Let F be a linear function on L which is bounded with respect to this norm, so that

$$|F(f)| \leq C\|f\|_p$$

for all $f \in L$ where C is some non-negative constant. The least upper bound of the set of C which work is called $\|F\|_p$ as usual. If $f \geq 0 \in L$, define

$$F^+(f) := \mathrm{lub}\{F(g) : 0 \leq g \leq f, \quad g \in L\}.$$

Then

$$F^+(f) \geq 0$$

and

$$F^+(f) \leq \|F\|_p \|f\|_p$$

since $F(g) \leq |F(g)| \leq \|F\|_p \|g\|_p \leq \|F\|_p \|f\|_p$ for all $0 \leq g \leq f$, $g \in L$, since $0 \leq g \leq f$ implies $|g|^p \leq |f|^p$ for $1 \leq p < \infty$ and also implies $\|g\|_\infty \leq \|f\|_\infty$. Also

$$F^+(cf) = cF^+(f) \quad \forall\, c \geq 0$$

as follows directly from the definition. Suppose that f_1 and f_2 are both non-negative elements of L. If g_1, $g_2 \in L$ with

$$0 \leq g_1 \leq f_1 \quad \text{and } 0 \leq g_2 \leq f_2$$

then

$$F^+(f_1 + f_2) \geq \mathrm{lub}\, F(g_1 + g_1) = \mathrm{lub}\, F(g_1) + \mathrm{lub}\, F(g_2) = F^+(f_1) + F^+(f_2).$$

On the other hand, if $g \in L$ satisfies $0 \leq g \leq (f_1 + f_2)$ then $0 \leq g \wedge f_1 \leq f_1$, and $g \wedge f_1 \in L$. Also $g - g \wedge f_1 \in L$ and vanishes at points x where $g(x) \leq f_1(x)$ while at points where $g(x) > f_1(x)$ we have $g(x) - g \wedge f_1(x) = g(x) - f_1(x) \leq f_2(x)$. So

$$g - g \wedge f_1 \leq f_2$$

and so

$$F^+(f_1 + f_2) = \mathrm{lub}\, F(g) \leq \mathrm{lub}\, F(g \wedge f_1) + \mathrm{lub}\, F(g - g \wedge f_1) \leq F^+(f_1) + F^+(f_2).$$

So

$$F^+(f_1 + f_2) = F^+(f_1) + F^+(f_2)$$

if both f_1 and f_2 are non-negative elements of L. Now write any $f \in L$ as $f = f_1 - g_1$ where f_1 and g_1 are non-negative. (For example we could take $f_1 = f^+$ and $g_1 = f^-$.) Define

$$F^+(f) = F^+(f_1) - F^+(g_1).$$

This is well defined, for if we also had $f = f_2 - g_2$ then $f_1 + g_2 = f_2 + g_1$ so

$$F^+(f_1) + F^+(g_2) = F^+(f_1 + g_2) = F^+(f_2 + g_1) = F^+(f_2) + F^+(g_1)$$

so

$$F^+(f_1) - F^+(g_1) = F^+(f_2) - F^+(g_2).$$

From this it follows that F^+ so extended is linear, and

$$|F^+(f)| \leq F^+(|f|) \leq \|F\|_p \|f\|_p$$

so F^+ is bounded.

Define F^- by

$$F^-(f) := F^+(f) - F(f).$$

As F^- is the difference of two linear functions it is linear. Since by its definition, $F^+(f) \geq F(f)$ if $f \geq 0$, we see that $F^-(f) \geq 0$ if $f \geq 0$. Clearly $\|F^-\| \leq \|F^+\|_p + \|F\| \leq 2\|F\|_p$. We have proved:

Proposition 6.7.1 *Every linear function on L which is bounded with respect to the $\| \cdot \|_p$ norm can be written as the difference $F = F^+ - F^-$ of two linear functions which are bounded and take non-negative values on non-negative functions.*

In fact, we could formulate this proposition more abstractly as dealing with a normed vector space which has an order relation consistent with its metric but we shall refrain from this more abstract formulation.

6.7.2   Duality of L^p and L^q when $\mu(S) < \infty$.

Theorem 6.7.1 *Suppose that $\mu(S) < \infty$ and that F is a bounded linear function on L^p with $1 \leq p < \infty$. Then there exists a unique $g \in L^q$ such that*

$$F(f) = (f, g) = I(fg).$$

Here $q = p/(p-1)$ if $p > 1$ and $q = \infty$ if $p = 1$.

Proof. Consider the restriction of F to L. We know that $F = F^+ - F^-$ where both F^+ and F^- are linear and non-negative and are bounded with respect to the $\| \cdot \|_p$ norm on L. The monotone convergence theorem implies that if $f_n \searrow 0$ then $\|f_n\|_p \to 0$ and the boundedness of F^+ with respect to the $\| \cdot \|_p$ says that

$$\|f_n\|_p \to 0 \quad \Rightarrow \quad F^+(f_n) \to 0.$$

So F^+ satisfies all the axioms for an integral, and so does F^-. If f vanishes outside a set of I measure zero, then $\|f\|_p = 0$. Applied to a function of the form $f = \mathbf{1}_A$ we conclude that if A has $\mu = \mu_I$ measure zero, then A has measure zero with respect to the measures determined by F^+ or F^-. We can apply the

Radon-Nikodym theorem to conclude that there are functions g^+ and g^- which belong to $L^1(I)$ and such that

$$F^{\pm}(f) = I(fg^{\pm})$$

for every f which belongs to $L^1(F^{\pm})$. In particular, if we set $g := g^+ - g^-$ then

$$F(f) = I(fg)$$

for every function f which is integrable with respect to both F^+ and F^-, in particular for any $f \in L^p(I)$. We must show that $g \in L^q$.

We first treat the case where $p > 1$. Suppose that $0 \le f \le |g|$ and that f is bounded. Then

$$I(f^q) \le I(f^{q-1} \cdot \mathrm{sgn}(g)g) = F(f^{q-1} \cdot \mathrm{sgn}(g)) \le \|F\|_p \|f^{q-1}\|_p.$$

So

$$I(f^q) \le \|F\|_p (I(f^{(q-1)p}))^{\frac{1}{p}}.$$

Now $(q-1)p = q$ so we have

$$I(f^q) \le \|F\|_p I(f^q)^{\frac{1}{p}} = \|F\|_q I(f^q)^{1-\frac{1}{q}}.$$

This gives

$$\|f\|_q \le \|F\|_p$$

for all $0 \le f \le |g|$ with f bounded. We can choose such functions f_n with $f_n \nearrow |g|$. It follows from the monotone convergence theorem that $|g|$ and hence $g \in L^q(I)$. This proves the theorem for $p > 1$.

Let us now give the argument for $p = 1$. We want to show that $\|g\|_\infty \le \|F\|_1$. Suppose that $\|g\|_\infty \ge \|F\|_1 + \epsilon$ where $\epsilon > 0$. Consider the function $\mathbf{1}_A$ where

$$A := \{x : |g(x)| \ge \|F\|_1 + \frac{\epsilon}{2}\}.$$

Then

$$(\|F\|_1 + \frac{\epsilon}{2})\mu(A) \le I(\mathbf{1}_A|g|) = I(\mathbf{1}_A\mathrm{sgn}(g)g) = F(\mathbf{1}_A\mathrm{sgn}(g))$$

$$\le \|F\|_1\|\mathbf{1}_A\mathrm{sgn}(g)\|_1 = \|F\|_1\mu(A)$$

which is impossible unless $\mu(A) = 0$, contrary to our assumption. QED

6.7.3   The case where $\mu(S) = \infty$.

Here the cases $p > 1$ and $p = 1$ may be different, depending on "how infinite S is".

Let us first consider the case where $p > 1$. If we restrict the functional F to any subspace of L^p its norm can only decrease. Consider a subspace consisting of all functions which vanish outside a subset S_1 where $\mu(S_1) < \infty$. We get

a corresponding function g_1 defined on S_1 (and set equal to zero off S_1) with $\|g_1\|_q \leq \|F\|_p$ and $F(f) = I(fg_1)$ for all f belonging to this subspace. If (S_2, g_2) is a second such pair, then the uniqueness part of the theorem shows that $g_1 = g_2$ almost everywhere on $S_1 \cap S_2$. Thus we can consistently define g_{12} on $S_1 \cup S_2$. Let

$$b := \text{lub}\{\|g_\alpha\|_q\}$$

taken over all such g_α. Since this set of numbers is bounded by $\|F\|_p$ this least upper bound is finite. We can therefore find a nested sequence of sets S_n and corresponding functions g_n such that

$$\|g_n\|_q \nearrow b.$$

By the triangle inequality, if $n > m$ then

$$\|g_n - g_m\|_q \leq \|g_n\|_q - \|g_m\|_q$$

and so, as in your proof of the L^2 Martingale convergence theorem, this sequence is Cauchy in the $\|\cdot\|_q$ norm. Hence there is a limit $g \in L^q$ and g is supported on

$$S_0 := \bigcup S_n.$$

There can be no pair (S', g') with S disjoint from S_0 and $g' \neq 0$ on a subset of positive measure of S'. Indeed, if this were the case, then we could consider $g + g'$ on $S \cup S'$ and this would have a strictly larger $\|\cdot\|_q$ norm than $\|g\|_q = b$, contradicting the definition of b. (It is at this point in the argument that we use $q < \infty$ which is the same as $p > 1$.) Thus F vanishes on any function which is supported outside S_0. We have thus reduced the theorem to the case where S is σ-finite.

If S is σ-finite, decompose S into a disjoint union of sets A_i of finite measure. Let f_m denote the restriction of $f \in L^p$ to A_m and let h_m denote the restriction of g to A_m. Then

$$\sum_{m=1}^{\infty} f_m = f$$

as a convergent series in L^p and so

$$F(f) = \sum_m F(f_m) = \sum_m \int_{A_m} f_m h_m$$

and this last series converges to $I(fg)$ in L^1.

So we have proved that $(L^p)^* = L^q$ in complete generality when $p > 1$, and for σ-finite S when $p = 1$.

It may happen (and will happen when we consider the Haar integral on the most general locally compact group) that we don't even have σ-finiteness. But we will have the following more complicated condition: Recall that a set A is called **integrable** (by Loomis) if $\mathbf{1}_A \in \mathcal{B}$. Now suppose that

$$S = \bigcup_\alpha S_\alpha$$

where this union is disjoint, but possibly uncountable, of integrable sets, and with the property that every integrable set is contained in at most a countable union of the S_α. A set A is called **measurable** if the intersections $A \cap S_\alpha$ are all integrable, and a function is called **measurable** if its restriction to each S_α has the property that the restriction of f to each S_α belongs to $\mathcal{B}$, and further, that either the restriction of f^+ to every S_α or the restriction of f^- to every S_α belongs to L^1.

If we find ourselves in this situation, then we can find a g_α on each S_α since S_α is σ-finite, and piece these all together to get a g defined on all of S. If $f \in L^1$ then the set where $f \neq 0$ can have intersections with positive measure with only countably many of the S_α and so we can apply the result for the σ-finite case for $p = 1$ to this more general case as well.

6.8 Integration on locally compact Hausdorff spaces.

Suppose that S is a locally compact Hausdorff space. As in the case of $\mathbf{R}^n$, we can (and will) take L to be the space of continuous functions of compact support. Dini's lemma then says that if $f_n \in L \searrow 0$ then $f_n \to 0$ in the uniform topology.

If A is any subset of S we will denote the set of $f \in L$ whose support is contained in A by L_A.

Lemma 6.8.1 *A non-negative linear function I is bounded in the uniform norm on L_C whenever C is compact.*

Proof. Choose $g \geq 0 \in L$ so that $g(x) \geq 1$ for $x \in C$. If $f \in L_C$ then

$$|f| \leq \|f\|_\infty g$$

so

$$|I(f)| \leq I(|f|) \leq I(g) \cdot \|f\|_\infty. \quad \text{QED.}$$

6.8.1 Riesz representation theorems.

This is the same Riesz, but two more theorems.

Theorem 6.8.1 *Every non-negative linear functional I on L is an integral.*

Proof. This is Dini's lemma together with the preceding lemma. Indeed, by Dini we know that $f_n \in L \searrow 0$ implies that $\|f_n\|_\infty \searrow 0$. Since f_1 has compact support, let C be its support, a compact set. All the succeeding f_n are then also supported in C and so by the preceding lemma $I(f_n) \searrow 0$. QED

Theorem 6.8.2 *Let F be a bounded linear function on L (with respect to the uniform norm). Then there are two integrals I^+ and I^- such that*

$$F(f) = I^+(f) - I^-(f).$$

Proof. We apply Proposition 6.7.1 to the case of our L and with the uniform norm, $\| \cdot \|_\infty$. We get

$$F = F^+ - F^-$$

and an examination of he proof will show that in fact

$$\|F^\pm\|_\infty \leq \|F\|_\infty.$$

By the preceding theorem, $F^\pm$ are both integrals. QED

6.8.2   Fubini's theorem.

Theorem 6.8.3 *Let S_1 and S_2 be locally compact Hausdorff spaces and let I and J be non-negative linear functionals on $L(S_1)$ and $L(S_2)$ respectively. Then*

$$I_x(J_y h(x,y)) = J_y(I_x(h(x,y))$$

for every $h \in L(S_1 \times S_2)$ in the obvious notation, and this common value is an integral on $L(S_1 \times S_2)$.

Proof via Stone-Weierstrass. The equation in the theorem is clearly true if $h(x,y) = f(x)g(y)$ where $f \in L(S_1)$ and $g \in L(S_2)$ and so it is true for any h which can be written as a finite sum of such functions. Let h be a general element of $L(S_1 \times S_2)$. then we can find compact subsets $C_1 \subset S_1$ and $C_2 \subset S_2$ such that h is supported in the compact set $C_1 \times C_2$. The functions of the form

$$\sum f_i(x) g_i(y)$$

where the f_i are all supported in C_1 and the g_i in C_2, and the sum is finite, form an algebra which separates points. So for any $\epsilon > 0$ we can find a k of the above form with

$$\|h - k\|_\infty < \epsilon.$$

Let B_1 and B_2 be bounds for I on $L(C_1)$ and J on $L(C_2)$ as provided by Lemma 6.8.1. Then

$$|J_y h(x,y) - \sum J(g_i) f_i(x)| = |[J_y(f-k)](x)| < \epsilon B_2.$$

This shows that $J_y h(x,y)$ is the uniform limit of continuous functions supported in C_1 and so $J_y h(x,y)$ is itself continuous and supported in C_1. It then follows that $I_x(J_y(h)$ is defined, and that

$$|I_x(J_y h(x,y) - \sum I(f)_i J(g_i)| \leq \epsilon B_1 B_2.$$

Doing things in the reverse order shows that

$$I_x(J_y h(x,y)) - J_y(I_x(h(x,y))| \leq 2\epsilon B_1 B_2.$$

Since ϵ is arbitrary, this gives the equality in the theorem. Since this (same) functional is non-negative, it is an integral by the first of the Riesz representation theorems above. QED

Let X be a locally compact Hausdorff space, and let L denote the space of continuous functions of compact support on X. Recall that the Riesz representation theorem (one of them) asserts that any non-negative linear function I on L satisfies the starting axioms for the Daniell integral, and hence corresponds to a measure μ defined on a σ-field, and such that $I(f)$ is given by integration of f relative to this measure for any $f \in L$.

6.9   The Riesz representation theorem redux.

I want to give an alternative proof of the Riesz representation theorem which will give some information about the possible σ-fields on which μ is defined. In particular, I want to show that we can find a μ (which is possibly an extension of the μ given by our previous proof of the Riesz representation theorem) which is defined on a σ-field which contains the Borel field $\mathcal{B}(X)$. Recall that $\mathcal{B}(X)$ is the smallest σ-field which contains the open sets.

Let $\mathcal{F}$ be a σ-field which contains $\mathcal{B}(X)$. A (non-negative valued) measure μ on $\mathcal{F}$ is called **regular** if

1. $\mu(K) < \infty$ for any compact subset $K \subset X$.

2. For any $A \in \mathcal{F}$

$$\mu(A) = \inf\{\mu(U) : \ A \subset U, \ U \text{ open}\}$$

3. If $U \subset X$ is open then

$$\mu(U) = \sup\{\mu(K) : K \subset U, \ \ K \text{ compact }\}.$$

The second condition is called **outer regularity** and the third condition is called **inner regularity**.

6.9.1   Statement of the theorem.

Here is the improved version of the Riesz representation theorem:

Theorem 6.9.1 *Let X be a locally compact Hausdorff space, L the space of continuous functions of compact support on X, and I a non-negative linear functional on L. Then there exists a σ-field $\mathcal{F}$ containing $\mathcal{B}(X)$ and a non-negative regular measure μ on $\mathcal{F}$ such that*

$$I(f) = \int f d\mu \tag{6.4}$$

for all $f \in L$. Furthermore, the restriction of μ to $\mathcal{B}(X)$ is unique.

The proof of this theorem hinges on some topological facts whose true place is in the chapter on metric spaces, but I will prove them here. The importance of the theorem is that it will allow us to derive some conclusions about spaces which are very huge (such as the space of "all" paths in $\mathbf{R}^n$) but are nevertheless locally compact (in fact compact) Hausdorff spaces. It is because we want to consider such spaces, that the earlier proof, which hinged on taking limits of *sequences* in the very definition of the Daniell integral, is insufficient to get at the results we want.

6.9.2   Propositions in topology.

Proposition 6.9.1 *Let X be a Hausdorff space, and let H and K be disjoint compact subsets of X. Then there exist disjoint open subsets U and V of X such that $H \subset U$ and $K \subset V$.*

This we actually did prove in the chapter on metric spaces.

Proposition 6.9.2 *Let X be a locally compact Hausdorff space, $x \in X$, and U an open set containing x. Then there exists an open set O such that*

- $x \in O$

- $\overline{O}$ *is compact, and*

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

Proof. Choose an open neighborhood W of x whose closure is compact, which is possible since we are assuming that X is locally compact. Let $Z = U \cap W$ so that $\overline{Z}$ is compact and hence so is $H := \overline{Z} \setminus Z$. Take $K := \{x\}$ in the preceding proposition. We then get an open set V containing x which is disjoint from an open set G containing $\overline{Z} \setminus Z$. Take $O := V \cap Z$. Then $x \in O$ and $\overline{O} \subset \overline{Z}$ is compact and O has empty intersection with $\overline{Z} \setminus Z$, and hence is is contained in $Z \subset U$. QED

Proposition 6.9.3 *Let X be a locally compact Hausdorff space, $K \subset U$ with K compact and U open subsets of X. Then there exists a V with*

$$K \subset V \subset \overline{V} \subset U$$

with V open and $\overline{V}$ compact.

Proof. Each $x \in K$ has a neighborhood O with compact closure contained in U, by the preceding proposition. The set of these O cover K, so a finite subcollection of them cover K and the union of this finite subcollection gives the desired V.

Proposition 6.9.4 *Let X be a locally compact Hausdorff space, $K \subset U$ with K compact and U open. Then there exists a continuous function h with compact support such that*

$$\mathbf{1}_K \le h \le \mathbf{1}_U$$

and

$$\operatorname{Supp}(h) \subset U.$$

Proof. Choose V as in Proposition 6.9.3. By Urysohn's lemma applied to the compact space $\overline{V}$ we can find a function $h : \overline{V} \to [0,1]$ such that $h = 1$ on K and $f = 0$ on $\overline{V} \setminus V$. Extend h to be zero on the complement of $\overline{V}$. Then h does the trick.

Proposition 6.9.5 *Let X be a locally compact Hausdorff space, $f \in L$, i.e. f is a continuous function of compact support on X. Suppose that there are open subsets $U_1, \ldots U_n$ such that*

$$\operatorname{Supp}(f) \subset \bigcup_{i=1}^{n} U_i.$$

Then there are $f_1, \ldots, f_n \in L$ such that

$$\operatorname{Supp}(f_i) \subset U_i$$

and

$$f = f_1 + \cdots + f_n.$$

If f is non-negative, the f_i can be chosen so as to be non-negative.

Proof. By induction, it is enough to consider the case $n = 2$. Let $K := \operatorname{Supp}(f)$, so $K \subset U_1 \cup U_2$. Let

$$L_1 := K \setminus U_1, \qquad L_2 := K \setminus U_2.$$

So L_1 and L_2 are disjoint compact sets. By Proposition 6.9.1 we can find disjoint open sets V_1, V_2 with

$$L_1 \subset V_1, \quad L_2 \subset V_2.$$

Set

$$K_1 := K \setminus V_1, \quad K_2 := K \setminus V_2.$$

Then K_1 and K_2 are compact, and

$$K = K_1 \cup K_2, \quad K_1 \subset U_1, \ K_2 \subset U_2.$$

Choose h_1 and h_2 as in Proposition 6.9.4. Then set

$$\phi_1 := h_1, \quad \phi_2 := h_2 - h_1 \wedge h_2.$$

Then $\operatorname{Supp}(\phi_1) = \operatorname{Supp}(h_1) \subset U_1$ by construction, and $\operatorname{Supp}(\phi_2) \subset \operatorname{Supp}(h_2) \subset U_2$, the ϕ_i take values in $[0,1]$, and, if $x \in K = \operatorname{Supp}(f)$

$$\phi_1(x) + \phi_2(x) = (h_1 \vee h_2)(x) = 1.$$

Then set

$$f_1 := \phi_1 f, \quad f_2 := \phi_2 f.$$

QED

6.9.3   Proof of the uniqueness of the μ restricted to $\mathcal{B}(X)$.

It is enough to prove that

$$
\begin{aligned}
\mu(U) &= \sup\{I(f): \ f \in L, \ 0 \le f \le \mathbf{1}_U\} & (6.5) \\
&= \sup\{I(f): \ f \in L, \ 0 \le f \le \mathbf{1}_U, \ \operatorname{Supp}(f) \subset U\} & (6.6)
\end{aligned}
$$

for any open set U, since either of these equations determines μ on any open set U and hence for the Borel field.

Since $f \le \mathbf{1}_U$ and both are measurable functions, it is clear that $\mu(U) = \int \mathbf{1}_U$ is at least as large as the expression on the right hand side of (6.5). This in turn is as least as large as the right hand side of (6.6) since the supremum in (6.6) is taken over as smaller set of functions that that of (6.5). So it is enough to prove that $\mu(U)$ is $\le$ the right hand side of (6.6).

Let $a < \mu(U)$. Interior regularity implies that we can find a compact set $K \subset U$ with

$$a < \mu(K).$$

Take the f provided by Proposition 6.9.4. Then $a < I(f)$, and so the right hand side of (6.6) is $\ge a$. Since a was any number $< \mu(U)$, we conclude that $\mu(U)$ is $\le$ the right hand side of 6.6). QED

6.10   Existence.

We will

- define a function m^* defined on all subsets,

- show that it is an outer measure,

- show that the set of measurable sets in the sense of Caratheodory include all the Borel sets, and that

- integration with respect to the associated measure μ assigns $I(f)$ to every $f \in L$.

6.10.1   Definition.

Define m^* on open sets by

$$
m^*(U) = \sup\{I(f): \ f \in L, 0 \le f \le \mathbf{1}_U, \ \operatorname{Supp}(f) \subset U\}. \qquad (6.7)
$$

Clearly, if $U \subset V$ are open subsets, $m^*(U) \le m^*(V)$. Next define m^* on an arbitrary subset by

$$
m^*(A) = \inf\{m^*(U): \ A \subset U, \ U \text{ open}\}. \qquad (6.8)
$$

Since U is contained in itself, this does not change the definition on open sets. It is clear that $m^*(\emptyset) = 0$ and that $A \subset B$ implies that $m^*(A) \le m^*(B)$. So

to prove that m^* is an outer measure we must prove countable subadditivity. We will first prove countable subadditivity on open sets, and then use the $\epsilon/2^n$ argument to conclude countable subadditivity on all sets:

Suppose $\{U_n\}$ is a sequence of open sets. We wish to prove that

$$m^*\left(\bigcup_n U_n\right) \leq \sum_n m^*(U_n). \tag{6.9}$$

Set

$$U := \bigcup_n U_n,$$

and suppose that

$$f \in L, 0 \leq f \leq \mathbf{1}_U, \ \mathrm{Supp}(f) \subset U.$$

Since $\mathrm{Supp}(f)$ is compact and contained in U, it is covered by finitely many of the U_i. In other words, there is some finite integer N such that

$$\mathrm{Supp}(f) \subset \bigcup_{n=1}^{N} U_n.$$

By Proposition 6.9.5 we can write

$$f = f_1 + \cdots + f_N, \ \mathrm{Supp}(f_i) \subset U_i, \ i = 1, \ldots, N.$$

Then

$$I(f) = \sum I(f_i) \leq \sum m^*(U_i),$$

using the definition (6.7). Replacing the finite sum on the right hand side of this inequality by the infinite sum, and then taking the supremum over f proves (6.9), where we use the definition (6.7) once again.

Next let $\{A_n\}$ be any sequence of subsets of X. We wish to prove that

$$m^*\left(\bigcup_n A_n\right) \leq \sum_n m^*(A_n).$$

This is automatic if the right hand side is infinite. So assume that

$$\sum_n m^*(A_n) < \infty$$

and choose open sets $U_n \supset A_n$ so that

$$m^*(U_n) \leq m^*(A_n) + \frac{\epsilon}{2^n}.$$

Then $U := \bigcup U_n$ is an open set containing $A := \bigcup A_n$ and

$$m^*(A) \leq m^*(U) \leq \sum m^*(U)_i \leq \sum_n m^*(A_n) + \epsilon.$$

Since ϵ is arbitrary, we have proved countable subadditivity.

6.10.2   Measurability of the Borel sets.

Let $\mathcal{F}$ denote the collection of subsets which are measurable in the sense of Caratheodory for the outer measure m^*. We wish to prove that $\mathcal{F} \supset \mathcal{B}(X)$. Since $\mathcal{B}(X)$ is the σ-field generated by the open sets, it is enough to show that every open set is measurable in the sense of Caratheodory, i.e. that

$$m^*(A) \geq m^*(A \cap U) + m^*(A \cap U^c) \tag{6.10}$$

for any open set U and any set A with $m^*(A) < \infty$: If $\epsilon > 0$, choose an open set $V \supset A$ with

$$m^*(V) \leq m^*(A) + \epsilon$$

which is possible by the definition (6.8). We will show that

$$m^*(V) \geq m^*(V \cap U) + m^*(V \cap U^c) - 2\epsilon. \tag{6.11}$$

This will then imply that

$$m^*(A) \geq m^*(A \cap U) + m^*(A \cap U^c) - 3\epsilon$$

and since $\epsilon > 0$ is arbitrary, this will imply (6.10).

Using the definition (6.7), we can find an $f_1 \in L$ such that

$$f_1 \leq \mathbf{1}_{V \cap U} \quad \text{and} \quad \mathrm{Supp}(f_1) \subset V \cap U$$

with

$$I(f_1) \geq m^*(V \cap U) - \epsilon.$$

Let $K := \mathrm{Supp}(f_1)$. Then $K \subset U$ and so $K^c \supset U^c$ and K^c is open. Hence $V \cap K^c$ is an open set and

$$V \cap K^c \supset V \cap U^c.$$

Using the definition (6.7), we can find an $f_2 \in L$ such that

$$f_2 \leq \mathbf{1}_{V \cap K^c} \quad \text{and} \quad \mathrm{Supp}(f_2) \subset V \cap K^c$$

with

$$I(f_2) \geq m^*(V \cap K^c) - \epsilon.$$

But $m^*(V \cap K^c) \geq m^*(V \cap U^c)$ since $V \cap K^c \supset V \cap U^c$. So

$$I(f_2) \geq m^*(V \cap U^c) - \epsilon.$$

So

$$f_1 + f_2 \leq \mathbf{1}_K + \mathbf{1}_{V \cap K^c} \leq \mathbf{1}_V$$

since $K = \mathrm{Supp}(f_1) \subset V$ and $\mathrm{Supp}(f_2) \subset V \cap K^c$. Also

$$\mathrm{Supp}(f_1 + f_2) \subset (K \cup V \cap K^c) = V.$$

Thus $f = f_1 + f_2 \in L$ and so by (6.7),

$$I(f_1 + f_2) \leq m^*(V).$$

This proves (6.11) and hence that all Borel sets are measurable.

6.10.3 Compact sets have finite measure.

Let μ be the measure associated to m on the σ-field $\mathcal{F}$ of measurable sets. We will now prove that μ is regular. The condition of outer regularity is automatic, since this was how we defined $\mu(A) = m^*(A)$ for a general set.

If K is a compact subset of X, we can find an $f \in L$ such that $\mathbf{1}_K \leq f$ by Proposition 6.9.4. Let $0 < \epsilon < 1$ and set

$$U_\epsilon := \{x : f(x) > 1 - \epsilon\}.$$

Then U_ϵ is an open set containing K. If $0 \leq g \in L$ satisfies $g \leq \mathbf{1}_{U_\epsilon}$, then $g = 0$ on U_ϵ^c, and for $x \in U_\epsilon$, $g(x) \leq 1$ while $f(x) > 1 - \epsilon$. So

$$g \leq \frac{1}{1 - \epsilon} f$$

and hence, by (6.7)

$$m^*(U_\epsilon) \leq \frac{1}{1 - \epsilon} I(f).$$

So, by (6.8)

$$\mu(K) \leq m^*(U_\epsilon) \leq \frac{1}{1 - \epsilon} I(f) < \infty.$$

Reviewing the preceding argument, we see that we have in fact proved the more general statement

Proposition 6.10.1 *If A is any subset of X and $f \in L$ is such that*

$$\mathbf{1}_A \leq f$$

then

$$m^*(A) \leq I(f).$$

6.10.4 Interior regularity.

We now prove interior regularity, which will be very important for us. We wish to prove that

$$\mu(U) = \sup\{\mu(K) : K \subset U, \quad K \text{ compact }\},$$

for any open set U, where, according to (6.7),

$$m^*(U) = \sup\{I(f) : f \in L, 0 \leq f \leq \mathbf{1}_U, \ \text{Supp}(f) \subset U\}.$$

Since $\text{Supp}(f)$ is compact, and contained in U, we will be done if we show that

$$f \in L, \quad 0 \leq f \leq 1 \Rightarrow I(f) \leq \mu(\text{Supp}(f)). \tag{6.12}$$

So let V be an open set containing $\text{Supp}(f)$. By definition (6.7),

$$\mu(V) \geq I(f)$$

and, since V is an arbitrary open set containing $\mathrm{Supp}(f)$, we have

$$\mu(\mathrm{Supp}(f)) \geq I(f)$$

using the definition (6.8) of $m^*(\mathrm{Supp}(f))$.

In the course of this argument we have proved

Proposition 6.10.2 *If $g \in L, 0 \leq g \leq \mathbf{1}_K$ where K is compact, then*

$$I(g) \leq \mu(K).$$

6.10.5   Conclusion of the proof.

Finally, we must show that all the elements of L are integrable with respect to μ and

$$I(f) = \int f d\mu. \tag{6.13}$$

Since the elements of L are continuous, they are Borel measurable. As every $f \in L$ can be written as the difference of two non-negative elements of L, and as both sides of (6.13) are linear in f, it is enough to prove (6.13) for non-negative functions.

Following Lebesgue, divide the "y-axis" up into intervals of size ϵ. That is, let ϵ be a positive number, and, for every positive integer n set

$$f_n(x) := \begin{cases} 0 & \text{if} & f(x) \leq (n-1)\epsilon \\ f(x) - (n-1)\epsilon & \text{if} & (n-1)\epsilon < f(x) \leq n\epsilon \\ \epsilon & \text{if} & n\epsilon < f(x) \end{cases}$$

If $(n-1)\epsilon \geq \|f\|_\infty$ only the first alternative can occur, so all but finitely many of the f_n vanish, and they all are continuous and have compact support so belong to L. Also

$$f = \sum f_n$$

this sum being finite, as we have observed, and so

$$I(f) = \sum I(f_n).$$

Set $K_0 := \mathrm{Supp}(f)$ and

$$K_n := \{x : f(x) \geq n\epsilon\} \quad n = 1, 2, \ldots.$$

Then the K_i are a nested decreasing collection of compact sets, and

$$\epsilon \mathbf{1}_{K_n} \leq f_n \leq \epsilon \mathbf{1}_{K_{n-1}}.$$

By Propositions 6.10.1 and 6.10.2 we have

$$\epsilon \mu(K_n) \leq I(f_n) \leq \epsilon \mu(K_{n-1}).$$

On the other hand, the monotonicity of the integral (and its definition) imply that

$$\epsilon\mu(K_n) \leq \int f_n d\mu \leq \epsilon\mu(K_{n-1}).$$

Summing these inequalities gives

$$\epsilon\sum_{i=1}^{N}\mu(K_n) \quad \leq I(f) \quad \leq \epsilon\sum_{i=0}^{N-1}\mu(K_n)$$

$$\epsilon\sum_{i=1}^{N}\mu(K_n) \quad \leq \int f d\mu \quad \epsilon\sum_{i=0}^{N-1}\mu(K_n)$$

where N is sufficiently large. Thus $I(f)$ and $\int f d\mu$ lie within a distance

$$\epsilon\sum_{i=0}^{N-1}\mu(K_n) - \epsilon\sum_{i=1}^{N}\mu(K_n) = \epsilon\mu(K_0) - \epsilon\mu(K_N) \leq \epsilon\mu(\mathrm{Supp}(f))$$

of one another. Since ϵ is arbitrary, we have proved (6.13) and completed the proof of the Riesz representation theorem.

Chapter 7

Wiener measure, Brownian motion and white noise.

7.1 Wiener measure.

We begin by constructing Wiener measure following a paper by Nelson, *Journal of Mathematical Physics* **5** (1964) 332-343.

7.1.1 The Big Path Space.

Let $\dot{\mathbf{R}}^{\mathbf{n}}$ denote the one point compactification of $\mathbf{R}^n$. Let

$$\Omega := \prod_{0 \le t < \infty} \dot{\mathbf{R}}^{\mathbf{n}} \tag{7.1}$$

be the product of copies of $\dot{\mathbf{R}}^{\mathbf{n}}$, one for each non-negative t. This is an uncountable product, and so a huge space, but by Tychonoff's theorem, it is compact and Hausdorff. We can think of a point ω of Ω as being a function from $\mathbf{R}_+$ to $\dot{\mathbf{R}}^{\mathbf{n}}$, i.e. as as a "curve" with no restrictions whatsoever.

Let F be a continuous function on the m-fold product:

$$F : \prod_{i=1}^{m} \dot{\mathbf{R}}^{\mathbf{n}} \to \mathbf{R},$$

and let $t_1 \le t_2 \le \cdots \le t_m$ be fixed "times". Define

$$\phi = \phi_{F;t_1,\ldots,t_m} : \ \Omega \to \mathbf{R}$$

by

$$\phi(\omega) := F(\omega(t_1), \ldots, \omega(t_m)).$$

We can call such a function a **finite** function since its value at ω depends only on the values of ω at finitely many points. The set of such functions satisfies our

183

abstract axioms for a space on which we can define integration. Furthermore, the set of such functions is an algebra containing 1 and which separates points, so is dense in $C(\Omega)$ by the Stone-Weierstrass theorem. Let us call the space of such functions $C_{fin}(\Omega)$.

If we define an integral I on $C_{fin}(\Omega)$ then, by the Stone-Weierstrass theorem it extends to $C(\Omega)$ and therefore, by the Riesz representation theorem, gives us a regular Borel measure on Ω.

For each $x \in \mathbf{R}^n$ we are going to define such an integral, I_x by

$$I_x(\phi) =$$

$$\int \cdots \int F(x_1, x_2, \ldots, x_m) p(x, x_1; t_1) p(x_1, x_2; t_2 - t_1) \cdots p(x_{m-1}, x_m, t_m - t_{m-1}) dx_1 \ldots dx_m$$

when $\phi = \phi_{F, t_1, \ldots, t_m}$ where

$$p(x, y; t) = \frac{1}{(2\pi t)^{n/2}} e^{-(x-y)^2/2t} \tag{7.2}$$

(with $p(x, \infty) = 0$) and all integrations are over $\dot{\mathbf{R}}^{\mathbf{n}}$. In order to check that this is well defined, we have to verify that if F does not depend on a given x_i then we get the same answer if we define ϕ in terms of the corresponding function of the remaining $m - 1$ variables. This amounts to the computation

$$\int p(x, y; s) p(y, z, t) dy = p(x, z; s + t).$$

If $n = 1$ this is the computation

$$\frac{1}{2\pi t} \int_{\mathbf{R}} e^{-(x-y)^2/2s} e^{-(y-z)^2 2t} dy = \frac{1}{2\pi(s + t)} e^{(x-z)^2/2(s+t)}.$$

If we make the change of variables $u = x - y$ this becomes

$$n_t \star n_s = n_{t+s}$$

where

$$n_r(x) := \frac{1}{\sqrt{r}} e^{-x^2/2r}.$$

In terms of our "scaling operator" S_a given by $S_a f(x) = f(ax)$ we can write

$$n_r = r^{-\frac{1}{2}} S_{r^{-\frac{1}{2}}} n$$

where n is the unit Gaussian $n(x) = e^{-x^2/2}$. Now the Fourier transform takes convolution into multiplication, satisfies

$$(S_a f)\hat{} = (1/a) S_{1/a} \hat{f},$$

and takes the unit Gaussian into the unit Gaussian. Thus upon Fourier transform, the equation $n_t \star n_s = n_{t+s}$ becomes the obvious fact that

$$e^{s\xi^2/2}e^{-t\xi^2/2} = e^{-(s+t)\xi^2/2}.$$

The same proof (or an iterated version of the one dimensional result) applies in n-dimensions.

So, for each $x \in \mathbf{R}^n$ we have defined a measure on Ω. We denote the measure corresponding to I_x by pr_x. It is a probability measure in the sense that $\mathrm{pr}_x(\Omega) = 1$.

The intuitive idea behind the definition of pr_x is that it assigns probability

$$\mathrm{pr}_x(E) :=$$

$$\int_{E_1} \cdots \int_{E_m} p(x, x_1; t_1) p(x_1, x_2; t_2 - t_1) \cdots p(x_{m-1}, x_m, t_m - t_{m-1}) dx_1 \dots dx_m$$

to the set of all paths ω which start at x and pass through the set E_1 at time t_1, the set E_2 at time t_2 etc. and we have denoted this set of paths by E.

7.1.2 The heat equation.

We pause to reflect upon the computation we did in the preceding section. Define the operator T_t on the space $\mathcal{S}$ (or on $\mathcal{S}'$) by

$$(T_t f)(x) = \int_{\mathbf{R}^n} p(x, y, t) f(y) dy. \tag{7.3}$$

In other words, T_t is the operation of convolution with

$$t^{-n/2} e^{-x^2/2t}.$$

We have verified that

$$T_t \circ T_s = T_{t+s}. \tag{7.4}$$

Also, we have verified that when we take Fourier transforms,

$$(T_t f)\hat{\ }(\xi) = e^{-t\xi^2/2} \hat{f}(\xi). \tag{7.5}$$

If we let $t \to 0$ in this equation we get

$$\lim_{t \to 0} T_t = \text{Identity.} \tag{7.6}$$

Using some language we will introduce later, conditions (7.4) and (7.6) say that the T_t form a continuous semi-group of operators. If we differentiate (7.5) with respect to t, and let

$$u(t, x) := (T_t f)(x)$$

we see that u is a solution of the "heat equation"

$$\frac{\partial^2 u}{(\partial t)^2} = \frac{\partial^2 u}{(\partial x^1)^2} + \cdots + \frac{\partial^2 u}{(\partial x^n)^2}$$

with the initial conditions $u(0, x) = f(x)$. In terms of the operator

$$\Delta := -\left(\frac{\partial^2}{(\partial x^1)^2} + \cdots + \frac{\partial^2}{(\partial x^n)^2}\right)$$

we are tempted to write

$$T_t = e^{-t\Delta},$$

in analogy to our study of elliptic operators on compact manifolds. We will spend lot of time justifying these kind of formulas in the non-compact setting later on in the course.

7.1.3 Paths are continuous with probability one.

The purpose of this subsection is to prove that if we use the measure pr_x, then the set of discontinuous paths has measure zero.

We begin with some technical issues. We recall that the statement that a measure μ is regular means that for any Borel set A

$$\mu(A) = \inf\{\mu(G) : \ A \subset G, \ G \text{ open }\}$$

and for any open set U

$$\mu(U) = \sup\{\mu(K) : \ K \subset U, \ K \text{ compact}\}.$$

This second condition has the following consequence: Suppose that Γ is any collection of open sets which is closed under finite union. If

$$O = \bigcup_{G \in \Gamma} G$$

then

$$\mu(O) = \sup_{G \in \Gamma} \mu(G)$$

since any compact subset of O is covered by finitely many sets belonging to Γ. The importance of this stems from the fact that we can allow Γ to consist of uncountably many open sets, and we will need to impose uncountably many conditions in singling out the space of continuous paths, for example. Indeed, our first task will be to show that the measure pr_x is concentrated on the space of continuous paths in $\mathbf{R}^n$ which do not go to infinity too fast.

We begin with the following computation in one dimension:

$$\mathrm{pr}_0(\{|\omega(t)| > r\}) = 2 \cdot \left(\frac{1}{2\pi t}\right)^{1/2} \int_r^\infty e^{-x^2/2t} dx \leq \left(\frac{2}{\pi t}\right)^{1/2} \int_r^\infty \frac{x}{r} e^{-x^2/2t} dx =$$

$$\left(\frac{2}{\pi t}\right)^{1/2} \frac{t}{r} \int_r^\infty \frac{x}{t} e^{-x^2/2t} dx = \left(\frac{2t}{\pi}\right)^{1/2} \frac{e^{-r^2/2t}}{r}.$$

For fixed r this tends to zero (very fast) as $t \to 0$. In n-dimensions $\|y\| > \epsilon$ (in the Euclidean norm) implies that at least one of its coordinates y_i satisfies $|y_i| > \epsilon/\sqrt{n}$ so we find that

$$\mathrm{pr}_x(\{|\omega(t) - x| > \epsilon\}) \le c e^{-\epsilon^2/2nt}$$

for a suitable constant depending only on n. In particular, if we let $\rho(\epsilon, \delta)$ denote the supremum of the above probability over all $0 < t \le \delta$ then

$$\rho(\epsilon, \delta) = o(\delta). \tag{7.7}$$

Lemma 7.1.1 *Let $0 \le t_1 \le \cdots \le t_m$ with $t_m - t_1 \le \delta$. Let*

$$A := \{\omega|\ |\omega(t_j) - \omega(t_1)| > \epsilon \quad \text{for some } j = 1, \ldots m\}.$$

Then

$$\mathrm{pr}_x(A) \le 2\rho(\frac{1}{2}\epsilon, \delta) \tag{7.8}$$

independently of the number m of steps.

Proof. Let

$$B := \{\omega|\ |\omega(t_1) - \omega(t_m)| > \frac{1}{2}\epsilon\}$$

let

$$C_i := \{\omega|\ |\omega(t_i) - \omega(t_m)| > \frac{1}{2}\epsilon\}$$

and let

$$D_i := \{\omega|\ |\omega(t_1) - \omega(t_i)| > \epsilon \text{ and } |\omega(t_1) - \omega(t_k)| \le \epsilon\ k = 1, \ldots i - 1\}.$$

If $\omega \in A$, then $\omega \in D_i$ for some i by the definition of A, by taking i to be the first j that works in the definition of A. If $\omega \notin B$ and $\omega \in D_i$ then $\omega \in C_i$ since it has to move a distance of at least $\frac{1}{2}\epsilon$ to get back from outside the ball of radius ϵ to inside the ball of radius $\frac{1}{2}\epsilon$. So we have

$$A \subset B \cup \bigcup_{i=1}^{m}(C_i \cap D_i)$$

and hence

$$\mathrm{pr}_x(A) \le \mathrm{pr}_x(B) + \sum_{i=1}^{m} \mathrm{pr}_x(C_i \cap D_i). \tag{7.9}$$

Now we can estimate $\mathrm{pr}_x(C_i \cap D_i)$ as follows. For ω to belong to this intersection, we must have $\omega \in D_i$ and then the path moves a distance at least $\frac{\epsilon}{2}$ in time $t_n - t_i$ and these two events are independent, so $\mathrm{pr}_x(C_i \cap D_i) \le \rho(\frac{\epsilon}{2}, \delta)\,\mathrm{pr}_x(D_i)$. Here is this argument in more detail: Let

$$F = \mathbf{1}_{\{(y,z)|\ |y-z| > \frac{1}{2}\epsilon\}}$$

so that
$$\mathbf{1}_{C_i} = \phi_{F,t_i,t_n}.$$

Similarly, let G be the indicator function of the subset of $\dot{\mathbf{R}}^n \times \dot{\mathbf{R}}^n \times \cdots \times \dot{\mathbf{R}}^n$ (i copies) consisting of all points with

$$|x_k - x_1| \le \epsilon, \quad k = 1, \ldots, i-1, \quad |x_1 - x_i| > \epsilon$$

so that
$$\mathbf{1}_{D_i} = \phi_{G,t_1,\ldots,t_j}.$$

Then
$$\mathrm{pr}_x(C_i \cap D_i) =$$

$$\int \cdots \int p(x, x_1; t_1) \cdots p(x_{i-1}, x_i; t_i - t_{i-1}) F(x_1, \ldots, x_i) G(x_i, x_n) p(x_i, x_n; t_n - t_i) dx_1 \cdots dx_n.$$

The last integral (with respect to x_n) is $\le \rho(\frac{1}{2}\epsilon, \delta)$. Thus

$$\mathrm{pr}_x(C_i \cap D_i) \le \rho(\frac{\epsilon}{2}, \delta) \, \mathrm{pr}_x(D_i).$$

The D_i are disjoint by definition, so

$$\sum \mathrm{pr}_x(D_i) \le \mathrm{pr}_x(\bigcup D_i) \le 1.$$

So
$$\mathrm{pr}_x(A) \le \mathrm{pr}_x(B) + \rho(\frac{1}{2}\epsilon, \delta) \le 2\rho(\frac{1}{2}\epsilon, \delta).$$

QED

Let
$$E : \{\omega | |\omega(t_i) - \omega(t_j)| > 2\epsilon \ \text{ for some } 1 \le j < k \le m\}.$$

Then $E \subset A$ since if $|\omega(t_j) - \omega(t_k)| > 2\epsilon$ then either $|\omega(t_1) - \omega(t_j)| > \epsilon$ or $|\omega(t_1) - \omega(t_k)| > \epsilon$ (or both). So

$$\mathrm{pr}_x(E) \le 2\rho(\frac{1}{2}\epsilon, \delta). \tag{7.10}$$

Lemma 7.1.2 *Let $0 \le a < b$ with $b - a \le \delta$. Let*

$$E(a, b, \epsilon) := \{\omega | \ |\omega(s) - \omega(t)| > 2\epsilon \ \text{ for some } \ s, t \in [a, b]\}.$$

Then
$$\mathrm{pr}_x(E(a, b.\epsilon)) \le 2\rho(\frac{1}{2}\epsilon, \delta).$$

Proof. Here is where we are going to use the regularity of the measure. Let S denote a finite subset of $[a, b]$ and and let

$$E(a, b, \epsilon, S) := \{\omega | \ |\omega(s) - \omega(t)| > 2\epsilon \ \text{ for some } \ s, t \in S\}.$$

Then $E(a, b, \epsilon, S)$ is an open set and $\mathrm{pr}_x(E(a, b, \epsilon, S)) < 2\rho(\frac{1}{2}\epsilon, \delta)$ for any S. The union over all S of the $E(a, b, \epsilon, S)$ is $E(a, b, \epsilon)$. The regularity of the measure now implies the lemma. QED

Let k and n be integers, and set

$$\delta := \frac{1}{n}.$$

Let

$$F(k, \epsilon, \delta) := \{\omega | \ |\omega(t) - \omega(s)| > 4\epsilon \ \text{ for some } t, s \in [0, k], \ \text{with } |t - s| < \delta\}.$$

Then we claim that

$$\mathrm{pr}_x(F(k, \epsilon, \delta)) < 2k\frac{\rho(\frac{1}{2}\epsilon, \delta)}{\delta}. \tag{7.11}$$

Indeed, $[0, k]$ is the union of the $nk = k/\delta$ subintervals $[0, \delta], [\delta, 2\delta], \ldots, [k - \delta, \delta]$. If $\omega \in F(k, \epsilon, \delta)$ then $|\omega(s) - \omega(t)| > 4\epsilon$ for some s and t which lie in either the same or in adjacent subintervals. So ω must lie in $E(a, b, \epsilon)$ for one of these subintervals, and there are kn of them. QED

Let $\omega \in \Omega$ be a continuous path in $\mathbf{R}^n$. Restricted to any interval $[0, k]$ it is uniformly continuous. This means that for any $\epsilon > 0$ it belongs to the complement of the set $F(k, \epsilon, \delta)$ for some δ. We can let $\epsilon = 1/p$ for some integer p. Let $\mathcal{C}$ denote the set of continuous paths from $[0, \infty)$ to $\mathbf{R}^n$. Then

$$\mathcal{C} = \bigcap_k \bigcap_\epsilon \bigcup_\delta F(k, \epsilon, \delta)^c$$

so the complement $\mathcal{C}^c$ of the set of continuous paths is

$$\bigcup_k \bigcup_\epsilon \bigcap_\delta F(k, \epsilon, \delta),$$

a countable union of sets of measure zero since

$$\mathrm{pr}_x\left(\bigcap_\delta F(k, \epsilon, \delta)\right) \leq \lim_{\delta \to 0} 2k\rho(\frac{1}{2}\epsilon, \delta)/\delta = 0.$$

We have thus proved a famous theorem of Wiener:

Theorem 7.1.1 [Wiener.] *The measure pr_x is concentrated on the space of continuous paths, i.e. $\mathrm{pr}_x(\mathcal{C}) = 1$. In particular, there is a probability measure on the space of continuous paths starting at the origin which assigns probability*

$$\mathrm{pr}_0(E) =$$

$$\int_{E_1} \cdots \int_{E_m} p(0, x_1; t_1) p(x_1, x_2; t_2 - t_1) \cdots p(x_{m-1}, x_m, t_m - t_{m-1}) dx_1 \ldots dx_m$$

to the set of all paths ω which start at 0 and pass through the set E_1 at time t_1, the set E_2 at time t_2 etc. and we have denoted this set of paths by E.

7.1.4 Embedding in $\mathcal{S}'$.

For convenience in notation let me now specialize to the case $n = 1$. Let

$$\mathcal{W} \subset \mathcal{C}$$

consist of those paths ω with $\omega(0) = 0$ and

$$\int_0^\infty (1+t)^{-2} w(t) dt < \infty.$$

Proposition 7.1.1 [Stroock] *The Wiener measure* pr_0 *is concentrated on* $\mathcal{W}$.

Indeed, we let $E(|\omega(t)|)$ denote the expectation of the function $|\omega(t)|$ of ω with respect to Wiener measure, so

$$E(|\omega(t)|) = \frac{1}{\sqrt{2\pi t}} \int_{\mathbf{R}} |x| e^{-x^2/2t} dx = \frac{1}{\sqrt{2\pi t}} \cdot t \int_0^\infty \frac{x}{t} e^{-x^2/t} dx = Ct^{1/2}.$$

Thus, by Fubini,

$$E\left(\int_0^\infty (1+t)^{-2} |w(t)| dt\right) = \int_0^\infty (1+t)^{-2} E(|w(t)|) < \infty.$$

Hence the set of ω with $\int_0^\infty (1+t)^{-2} |w(t)| dt = \infty$ must have measure zero. QED

Now each element of $\mathcal{W}$ defines a tempered distribution, i.e. an element of $\mathcal{S}'$ according to the rule

$$\langle \omega, \phi \rangle = \int_0^\infty \omega(t) \phi(t) dt.$$

This map from $\mathcal{W}$ to $\mathcal{S}'$ is continuous, hence

the Wiener measure pushes forward to give a measure on $\mathcal{S}'$.

We will want to examine this push forward measure due to Dan Stroock. First we need some results about Gaussian random variables.

7.2 Gaussian measures.

7.2.1 Generalities about expectation and variance.

Let V be a vector space (say over the reals and finite dimensional). Let X be a V-valued random variable. That is, we have some measure space $(M, \mathcal{F}, \mu)$ (which will be fixed and hidden in this section) where μ is a probability measure on M, and $X : M \to V$ is a measurable function. If X is integrable, then

$$E(X) := \int_M X d\mu$$

is called the **expectation** of X and is an element of V.

The function $X \otimes X$ is a $V \otimes V$ valued function, and if it is integrable, then

$$\mathrm{Var}(X) = E(X \otimes X) - E(X) \otimes E(X) = E\left(X - E(X)\right) \otimes \left((X - E(X))\right)$$

is called the **variance** of X and is an element of $V \otimes V$. It is by its definition a symmetric tensor, and so can be thought of as a quadratic form on V^*.

If $A : V \to W$ is a linear map, then AX is a W valued random variable, and

$$E(AX) = AE(X), \qquad \mathrm{Var}(AX) = (A \otimes A)\,\mathrm{Var}(X) \qquad (7.12)$$

assuming that $E(X)$ and $\mathrm{Var}(X)$ exist. We can also write this last equation as

$$\mathrm{Var}(AX)(\eta) = \mathrm{Var}(X)(A^*\eta), \qquad \eta \in W^* \qquad (7.13)$$

if we think of the variance as quadratic function on the dual space.

The function on V^* given by

$$\xi \mapsto E(e^{i\xi \cdot X})$$

is called the **characteristic function** associated to X and is denoted by ϕ_X. Here we have used the notation $\xi{\cdot}v$ to denote the value of $\xi \in V^*$ on $v \in V$. It is a version of the Fourier transform (with the conventions used by the probabilists). More precisely, let $X_*\mu$ denote the push forward of the measure μ by the map X, so that $X_*\mu$ is a probability measure on V. Then ϕ_X is the Fourier transform of this measure except that there are no powers of 2π in front of the integral and a plus rather than a minus sign is before the i in the exponent. These are the conventions of the probabilists. What is important for us is the fact that the Fourier transform determines the measure, i.e. ϕ_X determines $X_*\mu$. The probabilists would say that the *law* of the random variable (meaning $X_*\mu$) is determined by its characteristic function.

To get a feeling for (7.13) consider the case where $A = \xi$ is a linear map from V to $\mathbf{R}$. Then $\mathrm{Var}(X)(\xi) = \mathrm{Var}(\xi \cdot X)$ is the usual variance of the scalar valued random variable $\xi \cdot X$. Thus we see that $\mathrm{Var}(X)(\xi) \geq 0$, so $\mathrm{Var}(X)$ is non-negative definite symmetric bilinear form on V^*. The variance of a scalar valued random variable vanishes if and only if it is a constant. Thus $\mathrm{Var}(X)$ is positive definite unless X is concentrated on hyperplane.

Suppose that $A : V \to W$ is an isomorphism, and that $X_*\mu$ is absolutely continuous with respect to Lebesgue measure, so

$$X_*\mu = \rho dv$$

where ρ is some function on V (called the probability density of X). Then $(AX)_*\mu$ is absolutely continuous with respect to Lebesgue measure on W and its density σ is given by

$$\sigma(w) = \rho(A^{-1}w)|\det A|^{-1} \qquad (7.14)$$

as follows from the change of variables formula for multiple integrals.

7.2.2 Gaussian measures and their variances.

Let d be a positive integer. We say that N is a **unit** (d-dimensional) **Gaussian random variable** if N is a random variable with values in $\mathbf{R}^d$ with density

$$(2\pi)^{-d/2}e^{-(x_1^2\cdots+x_d^2)/2}.$$

It is clear that $E(N) = 0$ and, since

$$(2\pi)^{-d/2}\int x_i x_j e^{-(x_1^2\cdots+x_d^2)/2}dx = \delta_{ij},$$

that

$$\mathrm{Var}(N) = \sum_i \delta_i \otimes \delta_i \tag{7.15}$$

where $\delta_1, \ldots, \delta_d$ is the standard basis of $\mathbf{R}^d$. We will sometimes denote this tensor by I_d. In general we have the identification $V \otimes V$ with $\mathrm{Hom}(V^*, V)$, so we can think of the $\mathrm{Var}(X)$ as an element of $\mathrm{Hom}(V^*, V)$ if X is a V-valued random variable. If we identify $\mathbf{R}^d$ with its dual space using the standard basis, then I_d can be thought of as the identity matrix.

We can compute the characteristic function of N by reducing the computation to a product of one dimensional integrals yielding

$$\phi_N(t_1, \ldots, t_d) = e^{-(t_1^2+\cdots+t_d^2)/2}. \tag{7.16}$$

A V-valued random variable X is called **Gaussian** if (it is equal in law to a random variable of the form)

$$AN + a$$

where

$$A : \mathbf{R}^d \to V$$

is a linear map, where $a \in V$, and where N is a unit Gaussian random variable. Clearly

$$E(X) = a,$$

$$\mathrm{Var}(X) = (A \otimes A)(I_d)$$

or, put another way,

$$\mathrm{Var}(X)(\xi) = I_d(A^*\xi)$$

and hence

$$\phi_X(\xi) = \phi_N(A^*\xi)e^{i\xi\cdot a} = e^{-\frac{1}{2}I_d(A^*\xi)}e^{i\xi\cdot a}$$

or

$$\phi_X(\xi) = e^{-\mathrm{Var}(X)(\xi)/2+i\xi\cdot E(X)}. \tag{7.17}$$

It is a bit of a nuisance to carry along the $E(X)$ in all the computations, so we shall restrict ourselves to **centered Gaussian** random variables meaning that $E(X) = 0$. Thus for a centered Gaussian random variable we have

$$\phi_X(\xi) = e^{-\operatorname{Var}(X)(\xi)/2}. \tag{7.18}$$

Conversely, suppose that X is a V valued random variable whose characteristic function is of the form

$$\phi_X(\xi) = e^{-Q(\xi)/2},$$

where Q is a quadratic form. Since $|\phi_X(\xi)| \leq 1$ we see that Q must be non-negative definite. Suppose that we have chosen a basis of V so that V is identified with $\mathbf{R}^q$ where $q = \dim V$. By the principal axis theorem we can always find an orthogonal transformation (c_{ij}) which brings Q to diagonal form. In other words, if we set

$$\eta_j := \sum_i c_{ij}\xi_i$$

then

$$Q(\xi) = \sum_j \lambda_j \eta_j^2.$$

The λ_j are all non-negative since Q is non-negative definite. So if we set

$$a_{ij} := \lambda_j^{\frac{1}{2}} c_{ij}, \text{ and } A = (a_{ij})$$

we find that $Q(\xi) = I_q(A^*\xi)$. Hence X has the same characteristic function as a Gaussian random variable hence must be Gaussian.

As a corollary to this argument we see that

A random variable X is centered Gaussian if and only if $\xi \cdot X$ is a real valued Gaussian random variable with mean zero for each $\xi \in V^*$.

7.2.3 The variance of a Gaussian with density.

In our definition of a centered Gaussian random variable we were careful not to demand that the map A be an isomorphism. For example, if A were the zero map then we would end up with the δ function (at the origin for centered Gaussians) which (for reasons of passing to the limit) we want to consider as a Gaussian random variable.

But suppose that A *is* an isomorphism. Then by (7.14), X will have a density which is proportional to

$$e^{-S(v)/2}$$

where S is the quadratic form on V given by

$$S(v) = J_d(A^{-1}v)$$

and J_d is the unit quadratic form on $\mathbf{R}^d$:

$$J_d(x) = x_1^2 \cdots + x_d^2$$

or, in terms of the basis $\{\delta_i^*\}$ of the dual space to $\mathbf{R}^d$,

$$J_d = \sum_i \delta_i^* \otimes \delta_i^*.$$

Here $J_d \in (\mathbf{R}^d)^* \otimes (\mathbf{R}^d)^* = \mathrm{Hom}(\mathbf{R}^d, (\mathbf{R}^d)^*)$. It is the inverse of the map I_d. We can regard S as belonging to $\mathrm{Hom}(V, V^*)$ while we also regard $\mathrm{Var}(X) = (A \otimes A) \circ I_d$ as an element of $\mathrm{Hom}(V^*, V)$. I claim that $\mathrm{Var}(X)$ and S are inverses to one another. Indeed, dropping the subscript d which is fixed in this computation, $\mathrm{Var}(X)(\xi, \eta) = I(A^*\xi, A^*\eta) = \eta \cdot (A \circ I \circ A^*)\xi$ when thought of as a bilinear form on $V^* \otimes V^*$, and hence

$$\mathrm{Var}(X) = A \circ I \circ A^*$$

when thought of as an element of $\mathrm{Hom}(V^*, V)$. Similarly thinking of S as a bilinear form on V we have $S(v, w) = J(A^{-1}v, A^{-1}w) = J(A^{-1}v) \cdot A^{-1}w$ so

$$S = A^{-1*} \circ J \circ A^{-1}$$

when S is thought of as an element of $\mathrm{Hom}(V, V^*)$. Since I and J are inverses of one another, the two above displayed expressions for S and $\mathrm{Var}(X)$ show that these are inverses on one another.

This has the following very important computational consequence:

Suppose we are given a random variable X with (whose law has) a density proportional to $e^{-S(v)/2}$ where S is a quadratic form which is given as a "matrix" $S = (S_{ij})$ in terms of a basis of V^*. Then $\mathrm{Var}(X)$ is given by S^{-1} in terms of the dual basis of V.

7.2.4 The variance of Brownian motion.

For example, consider the two dimensional vector space with coordinates (x_1, x_2) and probability density proportional to

$$\exp -\frac{1}{2}\left(\frac{x_1^2}{s} + \frac{(x_2 - x_1)^2}{t - s}\right)$$

where $0 < s < t$. This corresponds to the matrix

$$\begin{pmatrix} \frac{t}{s(t-s)} & -\frac{1}{t-s} \\ -\frac{1}{t-s} & \frac{1}{t-s} \end{pmatrix} = \frac{1}{t-s}\begin{pmatrix} \frac{t}{s} & -1 \\ -1 & 1 \end{pmatrix}$$

whose inverse is

$$\begin{pmatrix} s & s \\ s & t \end{pmatrix}$$

which thus gives the variance.

So, if we let

$$B(s, t) := \min(s, t) \tag{7.19}$$

we can write the above variance as

$$\begin{pmatrix} B(s,s) & B(s,t) \\ B(t,s) & B(t,t) \end{pmatrix}.$$

Now suppose that we have picked some finite set of times $0 < s_1 < \cdots < s_n$ and we consider the corresponding Gaussian measure given by our formula for Brownian motion on a one-dimensional space for a path starting at the origin and passing successively through the points x_1 at time s_1, x_2 at time s_2 etc. We can compute the variance of this Gaussian to be

$$(B(s_i, s_j))$$

since the projection onto any coordinate plane (i.e. restricting to two values s_i and s_j) must have the variance given above.

Let $\phi \in \mathcal{S}$. We can think of ϕ as a (continuous) linear function on $\mathcal{S}'$. For convenience let us consider the real spaces $\mathcal{S}$ and $\mathcal{S}'$, so ϕ is a real valued linear function on $\mathcal{S}'$. Applied to Stroock's version of Brownian motion which is a probability measure living on $\mathcal{S}'$ we see that ϕ gives a real valued random variable. Recall that this was given by integrating $\phi \cdot \omega$ where ω is a continuous path of slow growth, and then integrating over Wiener measure on paths. Interchanging the order of integration we see that this is the limit of the Gaussian random variables

$$\frac{1}{n}(\phi(s_1)x_1 + \cdots + \phi(s_{n^2})x_{n^2})$$

where $s_k = k/n$, $k = 1, \ldots, n^2$, and hence ϕ defines a real valued centered Gaussian random variable whose variance is

$$\int_0^\infty \int_0^\infty \min(s,t)\phi(s)\phi(t)dsdt = 2 \int_0^\infty \int_{0 \leq s \leq t} s\phi(s)\phi(t)dsdt. \qquad (7.20)$$

Let us say that a probability measure μ on $\mathcal{S}'$ is a **centered Gaussian process** if every $\phi \in \mathcal{S}$, thought of as a function on the probability space $(\mathcal{S}', \mu)$ is a real valued centered Gaussian random variable; in other words $\phi_*(\mu)$ is a centered Gaussian probability measure on the real line. If we denote this process by Z, then we may write $Z(\phi)$ for the random variable given by ϕ. We clearly have $Z(a\phi + b\psi) = aZ(\phi) + bZ(\psi)$ in the sense of addition of random variables, and so we may think of Z as a rule which assigns, in a linear fashion, random variables to elements of $\mathcal{S}$. With some slight modification (we, following Stroock, are using $\mathcal{S}$ instead of $\mathcal{D}$ as our space of test functions) this notion was introduced by Gelfand some fifty years ago. (See Gelfand and Vilenkin, *Generalized Functions* volume IV.)

If we have generalized random process Z as above, we can consider its derivative in the sense of generalized functions, i.e.

$$\dot{Z}(\phi) := Z(-\dot{\phi}).$$

7.3 The derivative of Brownian motion is white noise.

To see how this derivative works, let us consider what happens for Brownian motion. Let ω be a continuous path of slow growth, and set

$$\omega_h(t) := \frac{1}{h}(\omega(t+h) - \omega(t)).$$

The paths ω are not differentiable (with probability one) so this limit does not exist as a function. But the limit does exist as a generalized function, assigning the value

$$\int_0^\infty -\dot{\phi}(t)\omega(t)dt$$

to ϕ. Now if $s < t$ the random variables $\omega(t+h) - \omega(t)$ and $\omega(s+h) - \omega(s)$ are independent of one another when $h < t - s$ since Brownian motion has independent increments. Hence we expect that this limiting process be independent at all points in some generalized sense. (No actual, as opposed to generalized, process can have this property. We will see more of this point in a moment when we compute the variance of $\dot{Z}$.)

In any event, $\dot{Z}(\phi)$ is a centered Gaussian random variable whose variance is given (according to (7.20)) by

$$2\int_0^\infty \left(\int_0^t s\dot{\phi}(s)ds\right)\dot{\phi}(t)dt.$$

We can integrate the inner integral by parts to obtain

$$\int_0^t s\dot{\phi}(s)ds = t\phi(t) - \int_0^t \phi(s)ds.$$

Integration by parts now yields

$$\int_0^\infty t\phi(t)\dot{\phi}(t)dt = -\frac{1}{2}\int_0^\infty \phi(t)^2 dt$$

and

$$-\int_0^\infty \left(\int_0^t \phi(s)ds\right)\dot{\phi}(t)dt = \int_0^\infty \phi(t)^2 dt.$$

We conclude that the variance of $\dot{Z}(\phi)$ is given by

$$\int_0^\infty \phi(t)^2 dt$$

which we can write as

$$\int_0^\infty \int_0^\infty \delta(s-t)\phi(s)\phi(t)dsdt.$$

Notice that now the "covariance function" is the generalized function $\delta(s - t)$. The generalized process (extended to the whole line) with this covariance is called white noise because it is a Gaussian process which is stationary under translations in time and its covariance "function" is $\delta(s-t)$, signifying independent variation at all times, and the Fourier transform of the delta function is a constant, i.e. assigns equal weight to all frequencies.

Chapter 8

Haar measure.

A **topological group** is a group G which is also a topological space such that the maps

$$G \times G \to G, \quad (x, y) \mapsto xy$$

and

$$G \to G, \quad x \mapsto x^{-1}$$

are continuous. If the topology on G is locally compact and Hausdorff, we say that G is a locally compact, Hausdorff, topological group. If $a \in G$ is fixed, then the map ℓ_a

$$\ell_a : G \to G, \quad \ell_a(x) = ax$$

is the composite of the multiplication map $G \times G \to G$ and the continuous map

$$G \to G \times G, \quad x \mapsto (a, x).$$

So ℓ_a is continuous, one to one, and with inverse $\ell_{a^{-1}}$. If μ is a measure G, then we can push it forward by ℓ_a, that is, consider the pushed forward measure $(\ell_a)_* \mu$. We say that the measure μ is **left invariant** if

$$(\ell_a)_* \mu = \mu \quad \forall \, a \in G.$$

The basic theorem on the subject, proved by Haar in 1933 is

Theorem 8.0.1 *If G is a locally compact Hausdorff topological group there exists a non-zero regular Borel measure μ which is left invariant. Any other such measure differs from μ by multiplication by a positive constant.*

This chapter is devoted to the proof of this theorem and the

8.1 Examples.

8.1.1 $\mathbf{R}^n$.

$\mathbf{R}^n$ is a group under addition, and Lebesgue measure is clearly left invariant. Similarly $\mathbf{T}^n$.

8.1.2   Discrete groups.

If G has the discrete topology then the counting measure which assigns the value one to every one element set $\{x\}$ is Haar measure.

8.1.3   Lie groups.

We can reformulate the condition of left invariance as follows: Let I denote the integral associated to the measure μ:

$$I(f) = \int f d\mu.$$

Then

$$\int f d(\ell_a)_* \mu = I(\ell_a^* f)$$

where

$$(\ell_a^* f)(x) = f(ax). \tag{8.1}$$

Indeed, this is most easily checked on indicator functions of sets, where

$$(\ell_a)^* \mathbf{1}_A = \mathbf{1}_{\ell_a^{-1} A}$$

and

$$\int \mathbf{1}_A d(\ell_a)_* \mu = ((\ell_a)_* \mu)(A) := \mu(\ell_a^{-1} A) = \int \mathbf{1}_{\ell_a^{-1} A} d\mu.$$

So the left invariance condition is

$$I(\ell_a^* f) = I(f) \quad \forall\, a \in G. \tag{8.2}$$

Suppose that G is a differentiable manifold and that the multiplication map $G \times G \to G$ and the inverse map $x \mapsto x^{-1}$ are differentiable.

   Now if G is n-dimensional, and we could find an n-form Ω which does not vanish anywhere, and such that

$$\ell_a^* \Omega = \Omega$$

(in the sense of pull-back on forms) then we can choose an orientation relative to which Ω becomes identified with a density, and then

$$I(f) = \int_G f\Omega$$

is the desired integral. Indeed,

$$
\begin{aligned}
I((\ell_a)^* f) &= \int ((\ell_a)^* f)\Omega \\
&= \int ((\ell_a)^* f)((\ell_a)^*\Omega) \quad \text{since } ((\ell_a)^*\Omega) = \Omega \\
&= \int (\ell_a)^*(f\Omega) \\
&= \int f\Omega \\
&= I(f).
\end{aligned}
$$

We shall replace the problem of finding a left invariant n-form Ω by the apparently harder looking problem of finding n left invariant one-forms $\omega_1, \ldots, \omega_n$ on G and then setting

$$
\Omega := \omega_1 \wedge \cdots \wedge \omega_n.
$$

The general theory of Lie groups says that such one forms (the Maurer-Cartan forms) always exist, but I want to sow how to compute them in important special cases.

Suppose that we can find a homomorphism

$$
M : G \to Gl(d)
$$

where $Gl(d)$ is the group of $d \times d$ invertible matrices (either real or complex). So M is a matrix valued function on G satisfying

$$
M(e) = \text{id}
$$

where e is the identity element of G and id is the identity matrix, and

$$
M(xy) = M(x)M(y)
$$

where multiplication on the left is group multiplication and multiplication on the right is matrix multiplication. We can think of M as a matrix valued function or as a matrix $M(x) = (M_{ij}(x))$ of real (or complex) valued functions. Suppose that all of these functions are differentiable. Then we can form

$$
dM := (dM_{ij})
$$

which is a matrix of linear differential forms on G, or, equivalently, a matrix valued linear differential form on G.

Finally, consider

$$
M^{-1}dM.
$$

Again, this is a matrix valued linear differential form on G (or what is the same thing a matrix of linear differential forms on G). Explicitly it is the matrix whose ik entry is

$$
\sum_j (M(x))^{-1})_{ij} dM_{jk}.
$$

I claim that every entry of this matrix is a left invariant linear differential form. Indeed,
$$(\ell_a^* M)(x) = M(ax) = M(a)M(x).$$

Let us write
$$A = M(a).$$

Since a is fixed, A is a constant matrix, and so
$$(\ell_a^* M)^{-1} = (AM)^{-1} = M^{-1}A^{-1}$$

while
$$\ell_a^* dM = d(AM) = AdM$$

since A is a constant. So
$$\ell_a^*(M^{-1}dM) = (M^{-1}A^{-1}AdM) = M^{-1}dM.$$

Of course, if the size of M is too small, there might not be enough linearly independent entries. (In the complex case we want to be able to choose the real and imaginary parts of these entries to be linearly independent.) But if, for example, the map $x \mapsto M(x)$ is an immersion, then there will be enough linearly independent entries to go around.

For example, consider the group of all two by two matrices of the form
$$\begin{pmatrix} a & b \\ 0 & 1 \end{pmatrix}.$$

This group is sometimes known as the "$ax + b$ group" since
$$\begin{pmatrix} a & b \\ 0 & 1 \end{pmatrix}\begin{pmatrix} x \\ 1 \end{pmatrix} = \begin{pmatrix} ax + b \\ 1 \end{pmatrix}.$$

In other words, G is the group of all translations and rescalings of the real line.

We have
$$\begin{pmatrix} a & b \\ 0 & 1 \end{pmatrix}^{-1} = \begin{pmatrix} a^{-1} & -a^{-1}b \\ 0 & 1 \end{pmatrix}$$

and
$$d\begin{pmatrix} a & b \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} da & db \\ 0 & 0 \end{pmatrix}$$

so
$$\begin{pmatrix} a & b \\ 0 & 1 \end{pmatrix}^{-1} d\begin{pmatrix} a & b \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} a^{-1}da & a^{-1}db \\ 0 & 1 \end{pmatrix}$$

and the Haar measure is (proportional to)

$$\frac{dadb}{a^2} \tag{8.3}$$

As a second example, consider the group $SU(2)$ of all unitary two by two matrices with determinant one. Each column of a unitary matrix is a unit vector, and the columns are orthogonal. We can write the first column of the matrix as

$$\begin{pmatrix} \overline{\alpha} \\ \beta \end{pmatrix}$$

where α and β are complex numbers with

$$|\alpha|^2 + |\beta|^2 = 1. \tag{8.4}$$

The second column must then be proportional to

$$\begin{pmatrix} -\beta \\ \alpha \end{pmatrix}$$

and the condition that the determinant be one fixes this constant of proportionality to be one. So we can write

$$M = \begin{pmatrix} \overline{\alpha} & -\beta \\ \overline{\beta} & \alpha \end{pmatrix}$$

where (8.4) is satisfied. So we can think of M as a complex matrix valued function on the group $SU(2)$. Since M is unitary, $M^{-1} = M^*$ so

$$M^{-1} = \begin{pmatrix} \alpha & \beta \\ -\overline{\beta} & \overline{\alpha} \end{pmatrix}$$

and

$$M^{-1}dM = \begin{pmatrix} \alpha & \beta \\ -\overline{\beta} & \overline{\alpha} \end{pmatrix} \begin{pmatrix} d\overline{\alpha} & -d\beta \\ d\overline{\beta} & d\alpha \end{pmatrix} = \begin{pmatrix} \alpha d\overline{\alpha} + \beta d\overline{\beta} & -\alpha d\beta + \beta d\alpha \\ -\overline{\beta}d\overline{\alpha} + \overline{\alpha}d\overline{\beta} & \overline{\alpha}d\alpha + \overline{\beta}d\beta \end{pmatrix}.$$

Each of the real and imaginary parts of the entries is a left invariant one form. But let us multiply three of these entries directly:

$$(\overline{\alpha}d\alpha + \overline{\beta}d\beta) \wedge (-\alpha d\beta + \beta d\alpha) \wedge (-\overline{\beta}d\overline{\alpha} + \overline{\alpha}d\overline{\beta})$$

$$= -(|\alpha|^2 + |\beta|^2)d\alpha \wedge d\beta \wedge (-\overline{\beta}d\overline{\alpha} + \overline{\alpha}d\overline{\beta})$$

$$-d\alpha \wedge d\beta \wedge (-\overline{\beta}d\overline{\alpha} + \overline{\alpha}d\overline{\beta}).$$

We can simplify this expression by differentiating the equation

$$\alpha\overline{\alpha} + \beta\overline{\beta} = 1$$

to get

$$\alpha d\overline{\alpha} + \overline{\alpha}d\alpha + \beta d\overline{\beta} + \overline{\beta}d\beta = 0.$$

So for $\beta \neq 0$ we can solve for $d\overline{\beta}$:

$$d\overline{\beta} = -\frac{1}{\beta}(\alpha d\overline{\alpha} + \overline{\alpha}d\alpha + \overline{\beta}d\beta).$$

When we multiply by $d\alpha \wedge d\beta$ the terms involving $d\alpha$ and $d\beta$ disappear. We thus get

$$-d\alpha \wedge d\beta \wedge (-\overline{\beta}d\overline{\alpha} + \overline{\alpha}d\overline{\beta}) = d\alpha \wedge d\beta \wedge (\overline{\beta}d\overline{\alpha} + \frac{\overline{\alpha}\alpha}{\beta}d\overline{\alpha}).$$

If we write

$$\overline{\beta}d\overline{\alpha} = \frac{\beta\overline{\beta}}{\beta}d\alpha$$

and use $|\alpha|^2 + |\beta|^2 = 1$ the above expression simplifies further to

$$\frac{1}{\beta}d\alpha \wedge d\beta \wedge d\overline{\alpha} \tag{8.5}$$

as a left invariant three form on $SU(2)$. You might think that this three form is complex valued, but we shall now give an alternative expression for it which will show that it is in fact real valued.

For this introduce polar coordinates in four dimensions as follows: Write

$$
\begin{aligned}
\alpha &= w + iz \\
\beta &= x + iy \ \text{ so } x^2 + y^2 + z^2 + w^2 = 1, \\
w &= \cos\theta \\
z &= \sin\theta \ \cos\psi \\
x &= \sin\theta \sin\psi \cos\phi \\
y &= \sin\theta \sin\psi \sin\phi
\end{aligned}
$$

$$0 \le \theta \le \pi, \ 0 \le \psi \le \pi, \ 0 \le \phi \le 2\pi.$$

Then

$$d\alpha \wedge d\overline{\alpha} = (dw + idz) \wedge (dw - idz) = -2idw \wedge dz$$

$$= -2id(\cos\theta) \wedge d(\sin\theta \cdot \cos\psi) = -2i\sin^2\theta d\theta \wedge d\psi.$$

Now

$$\beta = \sin\theta \sin\psi e^{i\phi}$$

so

$$d\beta = i\beta d\phi + \cdots$$

where the missing terms involve $d\theta$ and $d\psi$ and so will disappear when multiplied by $d\alpha \wedge d\overline{\alpha}$. Hence

$$d\alpha \wedge d\beta \wedge d\overline{\alpha} = -2\beta \sin^2\theta \sin\psi d\theta \wedge d\psi \wedge d\phi.$$

Finally, we see that the three form (8.5) when expressed in polar coordinates is

$$-2\sin^2\theta \sin\psi d\theta \wedge d\psi \wedge d\phi.$$

Of course we can multiply this by any constant. If we normalize so that $\mu(G) = 1$ the Haar measure is

$$\frac{1}{2\pi^2}\sin^2\theta \sin\psi d\theta d\psi d\phi.$$

8.2 Topological facts.

Since ℓ_a is a homeomorphism, if V is a neighborhood of the identity element e, then aV is a neighborhood of a, and if U is a neighborhood of U then $a^{-1}U$ is a neighborhood of e. Here we are using the obvious notation $aV = \ell_a(V)$ etc.

Suppose that U is a neighborhood of e. Then so is

$$U^{-1} := \{x^{-1} : \ x \in U\}$$

and hence so is

$$W = U \cap U^{-1}.$$

But

$$W^{-1} = W.$$

Proposition 8.2.1 *Every neighborhood of e contains a symmetric neighborhood, i.e. one that satisfies $W^{-1} = W$.*

Let U be a neighborhood of e. The inverse image of U under the multiplication map $G \times G \to G$ is a neighborhood of (e, e) in $G \times G$ and hence contains an open set of the form $V \times V$. Hence

Proposition 8.2.2 *Every neighborhood U of e contains a neighborhood V of e such that $V^2 = V \cdot V$.*

Here we are using the notation

$$A \cdot B = \{xy : \ x \in A, \ y \in B\}$$

where A and B are subsets of G.

If A and B are compact, so is $A \times B$ as a subset of $G \times G$, and since the image of a compact set under a continuous map is compact, we have

Proposition 8.2.3 *If A and B are compact, so is $A \cdot B$.*

Proposition 8.2.4 *If $A \subset G$ then $\overline{A}$, the closure of A, is given by*

$$\overline{A} = \bigcap_V AV$$

where V ranges over all neighborhoods of e.

Proof. If $a \in \overline{A}$ and V is a neighborhood of e, then aV^{-1} is an open set containing a, and hence containing a point of A. So $a \in AV$, and the left hand side of the equation in the proposition is contained in the right hand side. To show the reverse inclusion, suppose that x belongs to the right hand side. Then xV^{-1} intersects A for every V. But the sets xV^{-1} range over all neighborhoods of x. So $x \in \overline{A}$. QED

Recall that (following Loomis as we are) L denotes the space of continuous functions of compact support on G.

Proposition 8.2.5 *Suppose that G is locally compact. If $f \in L$ then f is uniformly left (and right) continuous. That is, given $\epsilon > 0$ there is a neighborhood V of e such that*

$$s \in V \ \Rightarrow \ |f(sx) - f(x)| < \epsilon.$$

Equivalently, this says that

$$xy^{-1} \in V \ \Rightarrow |f(x) - f(y)| < \epsilon.$$

Proof. Let

$$C := \mathrm{Supp}(f)$$

and let U be a symmetric compact neighborhood of e. Consider the set W of points s such that

$$|f(sx) - f(x)| < \epsilon \ \ \forall \ x \in UC.$$

I claim that this is an open neighborhood of e. Indeed, for each fixed $y \in UC$ the set of s satisfying this condition at y is an open neighborhood W_y of e, and this W_y works in some neighborhood O_y of y. Since UC is compact, finitely many of these O_y cover UC, and hence the intersection of the corresponding W_y form an open neighborhood W of e. Now take

$$V := U \cap W.$$

If $s \in V$ and $x \in UC$ then $|f(sx) - f(x)| < \epsilon$. If $x \notin UC$, then $sx \notin C$ (since we chose U to be symmetric) and $x \notin C$, so $f(sx) = 0$ and $f(x) = 0$, so $|f(sx) - f(x)| = 0 < \epsilon$. QED

In the construction of the Haar integral, we will need this proposition. So it is exactly at this point where the assumption that G is locally compact comes in.

8.3   Construction of the Haar integral.

Let f and g be non-zero elements of L^+ and let m_f and m_g be their respective maxima. At each

$$x \in \mathrm{Supp}(f),$$

we have

$$f(x) \leq \frac{m_f}{m_g} m_g$$

so if

$$c > \frac{m_f}{m_g}$$

and s is chosen so that g achieves its maximum at sx, then

$$f(y) \leq cg(sx)$$

in a neighborhood of x. Since $\text{Supp}(f)$ is compact, we can cover it by finitely many such neighborhoods, so that there exist finitely many c_i and s_i such that

$$f(x) \leq \sum c_i g(s_i x) \quad \forall \ x. \tag{8.6}$$

If we choose x so that $f(x) = m_f$, then the right hand side is at most $\sum_i c_i m_g$ and thus we see that

$$\sum_i c_i \geq m_f/m_g > 0.$$

So let us define the "size of f relative to g" by

$$(f;g) := \text{g.l.b.}\{\sum c_i : \ \exists s_i \text{ such that (8.6) holds}\}. \tag{8.7}$$

We have verified that

$$(f;g) \geq \frac{m_f}{m_g}. \tag{8.8}$$

It is clear that

$$
\begin{align}
(\ell_a^* f; g) &= (f;g) \ \ \forall \ a \in G \tag{8.9}\\
(f_1 + f_2; g) &\leq (f_1, g) + (f_2; g) \tag{8.10}\\
(cf; g) &= c(f;g) \ \ \forall \ c > 0 \tag{8.11}\\
f_1 \leq f_2 &\Rightarrow (f_1; g) \leq (f_2; g). \tag{8.12}
\end{align}
$$

If $f(x) \leq \sum c_i g(s_i x)$ for all x and $g(y) \leq \sum d_j g(t_j y)$ for all y then

$$f(x) \leq \sum_{ij} c_i d_j h(t_j s_i x) \ \forall \ x.$$

Taking greatest lower bounds gives

$$(f; h) \leq (f; g)(g; h). \tag{8.13}$$

To normalize our integral, fix some

$$f_0 \in L^+, \quad f_0 \neq 0.$$

Define

$$I_g(f) := \frac{(f;g)}{(f_0; g)}.$$

Since, according to (8.13) we have

$$(f_0; g) \leq (f_0; f)(f, g),$$

we see that

$$\frac{1}{(f_0; f)} \leq I_g(f).$$

Since (8.13) says that $(f;g) \leq (f;f_0)(f_0,g)$ we see that

$$I_g(f) \leq (f;f_0).$$

So for each non-zero $f \in L^+$ let S_f denote the closed interval

$$S_f := \left[\frac{1}{(f_0,f)}, (f,f_0) \right],$$

and let

$$S := \prod_{f \in L^+, f \neq 0} S_f.$$

This space is compact by Tychonoff. Each non-zero $g \in L^+$ determines a point $I_g \in S$ whose coordinate in S_f is $I_g(f)$.

For any neighborhood V of e, let C_V denote the closure in S of the set $I_g, g \in V$. We have

$$C_{V_1} \cap \cdots \cap C_{V_n} = C_{V_1 \cap \cdots \cap V_n} \neq \emptyset.$$

The C_V are all compact, and so there is a point I in the intersection of all the C_V:

$$I \in C := \bigcap_V C_V.$$

The idea is that I somehow is the limit of the I_g as we restrict the support of g to lie in smaller and smaller neighborhoods of the identity. We shall prove that as we make these neighborhoods smaller and smaller, the I_g are closer and closer to being additive, and so their limit I satisfies the conditions for being an invariant integral. Here are the details:

Lemma 8.3.1 *Given f_1 and f_2 in L^+ and $\epsilon > 0$ there exists a neighborhood V of e such that*

$$I_g(f_1) + I_g(f_2) \leq I_g(f_1 + f_2) + \epsilon$$

for all g with $\mathrm{Supp}(g) \subset V$.

Proof. Choose $\phi \in L$ such that $\phi = 1$ on $\mathrm{Supp}(f_1 + f_2)$. For a given $\delta > 0$ to be chosen later, let

$$f := f_1 + f_2 + \delta\phi, \quad h_1 := \frac{f_1}{f}, \quad h_2 := \frac{f_2}{f}.$$

Here h_1 and h_2 were defined on $\mathrm{Supp}(f)$ and vanish outside $\mathrm{Supp}(f_1 + f_2)$, so extend them to be zero outside $\mathrm{Supp}(\phi)$. For an $\eta > 0$ and $\delta > 0$ to be chosen later, find a neighborhood $V = V_{\delta,\eta}$ so that

$$|h_1(x) - h_1(y)| < \eta \quad \text{and} \quad |h_2(x) - h_2(y)| < \eta$$

when $x^{-1}y \in V$ which is possible by Prop. 8.2.5 if G is locally compact.

Let g be a non-zero element of L^+ with $\mathrm{Supp}(g) \subset V$. If

$$f(x) \leq \sum c_j g(s_j x)$$

then $g(s_j x) \neq 0$ implies that

$$|h_i(x) - h_i(s_j^{-1})| < \eta, \quad i = 1, 2$$

so

$$f_i(x) = f(x)h_i(x) \leq \sum c_j g(s_j x)h_i(x) \leq \sum c_j g(s_j x)[h_i(s_j^{-1}) + \epsilon], \quad i = 1, 2.$$

This implies that

$$(f_i; g) \leq \sum_j [h_i(s_j^{-1}) + \eta]$$

and since $0 \leq h_i \leq 1$ by definition,

$$(f_1; g) + (f_2; g) \leq \sum c_j [1 + 2\eta].$$

We can choose the c_j and s_j so that $\sum c_j$ is as close as we like to $(f; g)$. Hence

$$(f_1; g) + (f_2; g) \leq (f; g)[1 + 2\eta].$$

Dividing by $(f_0; g)$ gives

$$I_g(f_1) + I_g(f_2) \leq I_g(f)[1 + 2\eta]$$

$$\leq [I_g(f_1 + f_2) + \delta I_g(\phi)][1 + 2\eta],$$

where, in going from the second to the third inequality we have used the definition of f, (8.10) applied to $(f_1 + f_2)$ and $\delta\phi$ and (8.11). Now

$$I_g(f_1 + f_2) \leq (f_1 + f_2; f_0)$$

and $I_g(\phi) \leq (\phi, f_0)$. So choose δ and η so that

$$2\eta(f_1 + f_2; f_0) + \delta(1 + 2\eta)(\phi; f_0) < \epsilon.$$

This completes the proof of the lemma.

For any finite number of $f_i \in L^+$ and any neighborhood V of the identity, there is a non-zero g with $\mathrm{Supp}(g) \in V$ and

$$|I(f_i) - I_g(f_i)| < \epsilon, \quad i = 1, \dots, n.$$

Applying this to f_1, f_2 and $f_3 = f_1 + f_2$ and the V supplied by the lemma, we get

$$I(f_1 + f_2) - \epsilon \leq I_g(f_1 + f_2) \leq I_g(f_1) + I_g(f_2) \leq I(f_1) + I(f_2) + 2\epsilon$$

and

$$I(f_1) + I(f_2) \le I_g(f_1) + I_g(f_2) + 2\epsilon \le I_g(f_1 + f_2) + 3\epsilon \le I(f_1 + f_2) + 4\epsilon.$$

In short, I satisfies
$$I(f_1 + f_2) = I(f_1) + I(f_2)$$

for all f_1, f_2 in L^+, is left invariant, and $I(cf) = cI(f)$ for $c \ge 0$. As usual, extend I to all of L by

$$I(f_1 - f_2) = I(f_1) - I(f_2)$$

and this is well defined.

Since for $f \in L^+$ we have

$$I(f) \le (f; f_0) \le m_f/m_{f_0} = \|f\|_\infty/m_{f_0}$$

we see that I is bounded in the sup norm. So it is an integral (by Dini's lemma). Hence, by the Riesz representation theorem, if G is Hausdorff, we get a regular left invariant Borel measure. This completes the existence part of the main theorem.

From the fact that μ is regular, and not the zero measure, we conclude that there is some compact set K with $\mu(K) > 0$. Let U be any non-empty open set. The translates xU, $x \in K$ cover K, and since K is compact, a finite number, say n of them, cover K. But they all have the same measure, $\mu(U)$ since μ is left invariant. Thus
$$\mu(K) \le n\mu(U)$$

implying
$$\mu(U) > 0 \text{ for any non-empty open set } U \tag{8.14}$$

if μ is a left invariant regular Borel measure.

If $f \in L^+$ and $f \ne 0$, then $f > \epsilon > 0$ on some non-empty open set U, and hence its integral is $> \epsilon\mu(U)$. So

$$f \in L^+, \ f \ne 0 \ \Rightarrow \ \int f d\mu > 0 \tag{8.15}$$

for any left invariant regular Borel measure μ.

8.4   Uniqueness.

Let μ and ν be two left invariant regular Borel measures on G. Pick some $g \in L^+, g \ne 0$ so that both $\int g d\mu$ and $\int g d\nu$ are positive. We are going to use Fubini to prove that for any $f \in L$ we have

$$\frac{\int f d\nu}{\int g d\nu} = \frac{\int f d\mu}{\int g d\mu}. \tag{8.16}$$

This clearly implies that $\nu = c\mu$ where

$$c = \frac{\int g d\nu}{\int g d\mu}.$$

To prove (8.16), it is enough to show that the left hand side can be expressed in terms of any left invariant regular Borel measure (say ν) because this implies that both sides do not depend on the choice of Haar measure. Define

$$h(x, y) := \frac{f(x)g(yx)}{\int g(tx) d\nu(t)}.$$

The integral in the denominator is positive for all x and by the left uniform continuity the integral is a continuous function of x. Thus h is continuous function of compact support in (x, y) so by Fubini,

$$\int \int h(x, y) d\nu(y) d\mu(x) = \int \int h(x, y) d\mu(x) d\nu(y).$$

In the inner integral on the right replace x by $y^{-1}x$ using the left invariance of μ. The right hand side becomes

$$\int h(y^{-1}x, y) d\mu(x) d\nu(y).$$

Use Fubini again so that this becomes

$$\int h(y^{-1}x, y) d\nu(y) d\mu(x).$$

Now use the left invariance of ν to replace y by xy. This last iterated integral becomes

$$\int h(y^{-1}, xy) d\nu(y) d\mu(x).$$

So we have

$$\int \int h(x, y) d\nu(y) d\mu(x) = \int h(y^{-1}, xy) d\nu(y) d\mu(x).$$

From the definition of h the left hand side is $\int f(x) d\mu(x)$. For the right hand side

$$h(y^{-1}, xy) = f(y^{-1}) \frac{g(x)}{\int g(ty^{-1}) d\nu(t)}.$$

Integrating this first with respect to $d\nu(y)$ gives

$$kg(x)$$

where k is the constant

$$k = \int \frac{f(y^{-1})}{\int g(ty^{-1}) d\nu(t)} d\nu(y).$$

Now integrate with respect to μ. We get $\int f d\mu = k \int g d\mu$ so

$$\frac{\int f d\mu}{\int g d\mu},$$

the right hand side of (8.16), does not depend on μ, since it equals k which is expressed in terms of ν. QED

8.5 $\mu(G) < \infty$ if and only if G is compact.

Since μ is regular, the measure of any compact set is finite, so if G is compact then $\mu(G) < \infty$. We want to prove the converse. Let U be an open neighborhood of e with compact closure, K. So $\mu(K) > 0$. The fact that $\mu(G) < \infty$ implies that one can not have m disjoint sets of the form $x_i K$ if

$$m > \frac{\mu(G)}{\mu(K)}.$$

Let n be such that we can find n disjoint sets of the form $x_i K$ but no $n + 1$ disjoint sets of this form. This says that for any $x \in G$, xK can not be disjoint from all the $x_i K$. Thus

$$G = \left(\bigcup_i x_i K \right) \cdot K^{-1}$$

which is compact. QED

If G is compact, the Haar measure is usually normalized so that $\mu(G) = 1$.

8.6   The group algebra.

If $f, g \in L$ define their **convolution** by

$$(f \star g)(x) := \int f(xy)g(y^{-1})d\mu(y), \tag{8.17}$$

where we have fixed, once and for all, a (left) Haar measure μ. The left invariance (under left multiplication by x^{-1}) implies that

$$(f \star g)(x) = \int f(y)g(y^{-1}x)d\mu(y). \tag{8.18}$$

In what follows we will write dy instead of $d\mu(y)$ since we have chosen a fixed Haar measure μ.

If $A := \mathrm{Supp}(f)$ and $B := \mathrm{Supp}(g)$ then $f(y)g(y^{-1}x)$ is continuous as a function of y for each fixed x and vanishes unless $y \in A$ and $y^{-1}x \in B$. Thus $f \star g$ vanishes unless $x \in AB$. Also

$$|f \star g(x_1) - f \star g(x_2) \leq \|\ell_{x_1}^* f - \ell_{x_2}^* f\|_\infty \int |g(y^{-1})| dy.$$

Since $x \mapsto \ell_x^* f$ is continuous in the uniform norm, we conclude that

$$f, g \in L \Rightarrow f \star g \in L$$

and

$$\text{Supp}(f \star g) \subset (\text{Supp}(f)) \cdot (\text{Supp}(g)). \tag{8.19}$$

I claim that we have the associative law: If, $f, g, h \in L$ then

$$((f \star g) \star h) = (f \star (g \star h) \tag{8.20}$$

Indeed, using the left invariance of the Haar measure and Fubini we have

$$
\begin{aligned}
((f \star g) \star h)(x) \quad &:= \quad \int (f \star g)(xy)h(y^{-1})dy \\
&= \quad \int \int f(xyz)g(z^{-1})h(y^{-1})dzdy \\
&= \quad \int \int f(xz)g(z^{-1}y)h(y^{-1})dzdy \\
&= \quad \int \int f(xz)g(z^{-1}y)h(y^{-1})dydz \\
&= \quad \int f(xz)(g \star h)(z^{-1})dz \\
&= \quad (f \star (g \star h))(x).
\end{aligned}
$$

It is easy to check that $\star$ is commutative if and only if G is commutative.

I now want to extend the definition of $\star$ to all of L^1, and here I will follow Loomis and restrict our definition of L^1 so that our integrable functions belong to $\mathcal{B}$, the smallest monotone class containing L. When we were doing the Wiener integral, we needed all Borel sets. Here it is more convenient to operate with this smaller class, for technical reasons which will be practically invisible. For most groups one encounters in real life there is no difference between the Borel sets and the Baire sets. For example, if the Haar measure is σ-finite one can forget about these considerations.

If f and g are functions on G define the function $f \bullet g$ on $G \times G$ by

$$(f \bullet g)(x, y) := f(y)g(y^{-1}x).$$

Theorem 8.6.1 [31A in Loomis] *If $f, g \in \mathcal{B}^+$ then $f \bullet g \in \mathcal{B}^+(G \times G)$ and*

$$\|f \star g\|_p \le \|f\|_1 \|g\|_p \tag{8.21}$$

for any p with $1 \le p \le \infty$.

Proof. If $f \in L^+$ then the set of $g \in \mathcal{B}^+$ such that $f \bullet g \in \mathcal{B}^+(G \times G)$ is L-monotone and includes L^+, so includes $\mathcal{B}^+$. So if g is an L-bounded function in $\mathcal{B}^+$, the set of $f \in \mathcal{B}^+$ such that $f \bullet g \in \mathcal{B}^+(G \times G)$ includes L^+ and is L-monotone, and so includes $\mathcal{B}^+$. So $f \bullet g \in \mathcal{B}^+(G \times G)$ whenever f and g are

L-bounded elements of $\mathcal{B}^+$. But the most general element of $\mathcal{B}^+$ can be written as the limit of an increasing sequence of L bounded elements of $\mathcal{B}^+$, and so the first assertion in the theorem follows.

As f and g are non-negative, Fubini asserts that the function $y \mapsto f(y)g(y^{-1}x)$ is integrable for each fixed x, that $f \star g$ is integrable as a function of x and that

$$\|f \star g\|_1 = \int \int f(y)g(y^{-1}x)dxdy = \|f\|_1\|g\|_1.$$

This proves (8.21) (with equality) for $p = 1$. For $p = \infty$ (8.21) is obvious from the definitions.

For $1 < p < \infty$ we will use Hölder's inequality and the duality between L^p and L^q. We know that the product of two elements of $\mathcal{B}^+(G \times G)$ is an element of $\mathcal{B}^+(G \times G)$. So if $f, g, h \in \mathcal{B}^+(G)$ then the function $(x, y) \mapsto f(y)g(y^{-1}x)h(x)$ is an element of $\mathcal{B}^+(G \times G)$, and by Fubini

$$(f \star g, h) = \int f(y) \left[\int g(y^{-1}x)h(x)dx \right] dy.$$

We may apply Hölder's inequality to estimate the inner integral by $\|g\|_p\|h\|_q$. So for general $h \in L^q$ we have

$$|(f \star g, h)| \le \|f\|_1\|g\|_p\|h\|_q.$$

If $\|f\|_1$ or $\|g\|_p$ are infinite, then (8.21) is trivial. If both are finite, then

$$h \mapsto (f \star g, h)$$

is a bounded linear functional on L^q and so by the isomorphism $L^p = (L^q)^*$ we conclude that $f \star g \in L^p$ and that (8.21) holds. QED

We define a **Banach algebra** to be an associative algebra (possibly without a unit element) which is also a Banach space, and such that

$$\|fg\| \le \|f\|\|g\|.$$

So the special case $p = 1$ of Theorem 8.21 asserts that $L^1(G)$ is a Banach algebra.

8.7   The involution.

8.7.1   The modular function.

Instead of considering the action of G on itself by left multiplication, $a \mapsto \ell_a$ we can consider right multiplication, r_b where

$$r_b(x) = xb^{-1}.$$

It is one of the basic principles of mathematics, that on account of the associative law, right and left multiplication commute:

$$\ell_a \circ r_b = r_b \circ \ell_a. \tag{8.22}$$

Indeed, both sides send $x \in G$ to

$$axb^{-1}.$$

If μ is a choice of left Haar measure, then it follows from (8.22) that $r_{b*}\mu$ is another choice of Haar measure, and so must be some positive multiple of μ. The function Δ on G defined by

$$r_{b*}\mu = \Delta(b)\mu$$

is called the **modular function** of G. It is immediate that this definition does not depend on the choice of μ.

Example. In the case that we are dealing with manifolds and integration of n-forms,

$$I(f) = \int f\Omega,$$

then the push-forward of the measure associated to I under a diffeomorphism ϕ assigns to any function f the integral

$$I(\phi^* f) = \int (\phi^* f)\Omega = \int \phi^*(f(\phi^{-1})^*\Omega) = \int f(\phi^{-1})^*\Omega.$$

So the push forward measure corresponds to the form

$$(\phi^{-1})^*\Omega.$$

Thus in computing $\Delta(z)$ using differential forms, we have to compute the pull-back under right multiplication by z, not z^{-1}. For example, in the $ax + b$ group, we have

$$\begin{pmatrix} a & b \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x & y \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} ax & ay + b \\ 0 & 1 \end{pmatrix}$$

so

$$\frac{da \wedge db}{a^2} \mapsto \frac{xda \wedge db}{x^2 a^2}$$

and hence the modular function is given by

$$\Delta\left(\begin{pmatrix} x & y \\ 0 & 1 \end{pmatrix}\right) = \frac{1}{x}.$$

In all cases the modular function is continuous (as follows from the uniform right continuity, Proposition, 8.2.5), and from its definition, it follows that

$$\Delta(st) = \Delta(s)\Delta(t).$$

In other words, Δ is a continuous homomorphism from G to the multiplicative group of positive real numbers.

The group G is called **unimodular** if $\Delta \equiv 1$. For example, a commutative group is obviously unimodular. Also, a compact group is unimodular, because G has finite measure, and is carried into itself by right multiplication so

$$\Delta(s)\mu(G) = (r_{s*}(\mu)(G) = \mu(r_s^{-1}(G)) = \mu(G).$$

8.7.2   Definition of the involution.

For any complex valued continuous function f of compact support define $\tilde{f}$ by

$$\tilde{f}(x) := \overline{f(x^{-1})}\Delta(x^{-1}). \tag{8.23}$$

It follows immediately from the definition that

$$(\tilde{f})\tilde{} = f.$$

That is, applying $\tilde{}$ twice is the identity transformation. Also,

$$(r_s(\tilde{f}))(x) = \overline{f(sx^{-1})}\Delta(x^{-1})\Delta(s)$$

so

$$r_s\tilde{f} = \Delta(s)(\ell_s f)\tilde{}. \tag{8.24}$$

Similarly,

$$(r_s f)\tilde{}(x) = \overline{f(x^{-1}s^{-1})}\Delta(x^{-1}) = \Delta(s)\ell_s(\tilde{f})$$

or

$$(r_s f)\tilde{} = \Delta(s)\ell_s\tilde{f}. \tag{8.25}$$

Suppose that f is real valued, and consider the functional

$$J(f) := I(\tilde{f}).$$

Then from (8.24) and the definition of Δ we have

$$J(\ell_s f) = \Delta(s^{-1})I(r_s\tilde{f}) = \Delta(s^{-1})\Delta(s)I(\tilde{f}) = I(\tilde{f}) = J(f).$$

In other words, J is a left invariant integral on real valued functions, and hence must be some constant multiple of I,

$$J = cI.$$

Let V be a symmetric neighborhood of e chosen so small that $|1 - \Delta(s)| < \epsilon$ which is possible for any given $\epsilon > 0$, since $\Delta(e) = 1$ and Δ is continuous. If we take $f = \mathbf{1}_V$ then $f(x) = f(x^{-1})$ and

$$|J(f) - I(f)| \le \epsilon I(f).$$

Dividing by $I(f)$ shows that $|c - 1| < \epsilon$, and since ϵ is arbitrary, we have proved that

$$I(\tilde{f}) = \overline{I(f)}. \tag{8.26}$$

We can derive two immediate consequences:

Proposition 8.7.1 *Haar measure is inverse invariant if and only if G is unimodular,*

and

Proposition 8.7.2 *The involution $f \mapsto \tilde{f}$ extends to an anti-linear isometry of $L^1_{\mathbf{C}}$.*

8.7.3 Relation to convolution.

We claim that

$$(f \star g)^{\tilde{}} = \tilde{g} \star \tilde{f} \tag{8.27}$$

Proof.

$$
\begin{aligned}
(f \star g)^{\tilde{}}(x) &= \int \overline{f(x^{-1}y)g(y^{-1})}dy\Delta(x^{-1}) \\
&= \int \overline{g(y^{-1})}\Delta(y^{-1})\overline{f((y^{-1}x)^{-1})}\Delta((y^{-1}x)^{-1})dy \\
&= (\tilde{g} \star \tilde{f})(x). \text{ QED}
\end{aligned}
$$

8.7.4 Banach algebras with involutions.

For a general Banach algebra B (over the complex numbers) a map

$$x \mapsto x^\dagger$$

is called an involution if it is antilinear and anti-multiplicative, i.e. satisfies

$$(xy)^\dagger = y^\dagger x^\dagger.$$

Thus the map $f \mapsto \tilde{f}$ is an involution on $L^1(G)$.

8.8 The algebra of finite measures.

In general, the algebra $L^1(G)$ will not have an identity element, since the only candidate for the identity element would be the δ-"function"

$$\langle \delta, f \rangle = f(e),$$

and this will not be an honest function unless the topology of G is discrete.

So we need to introduce a different algebra if we want to have an algebra with identity. If G were a Lie group we could consider the algebra of all distributions. For a general locally compact Hausdorff group we can proceed as follows: Let $\mathcal{M}(G)$ denote the space of all finite complex measures on G: A non-negative measure ν is called **finite** if $\mu(G) < \infty$. A real valued measure is called finite if its positive and negative parts are finite, and a complex valued measure is called finite if its real and imaginary parts are finite.

Given two finite measures μ and ν on G we can form the product measure $\mu \otimes \nu$ on $G \times G$ and then push this measure forward under the multiplication map

$$m : G \times G \to G,$$

and so define their convolution by

$$\mu \star \nu := m_*(\mu \otimes \nu).$$

One checks that the convolution of two regular Borel measures is again a regular Borel measure, and that on measures which are absolutely continuous with respect to Haar measure, this coincides with the convolution as previously defined. One can also make the algebra of regular finite Borel measures under convolution into a Banach algebra (under the "total variation norm"). This algebra does include the δ-function (which is a measure!) and so has an identity. I will not go into this matter here except to make a number of vague but important points.

8.8.1   Algebras and coalgebras.

An algebra A is a vector space (over the complex numbers) together with a map

$$m : A \otimes A \to A$$

which is subject to various conditions (perhaps the associative law, perhaps the commutative law, perhaps the existence of the identity, etc.). The "dual" object would be a co-algebra, consisting of a vector space C and a map

$$c : C \to C \otimes C$$

subjects to a series of conditions dual to those listed above. If A is finite dimensional, then we have an identification of $(A \otimes A)^*$ with $A^\star \otimes A^\star$, and so the dual space of a finite dimensional algebra is a coalgebra and vice versa. For infinite dimensional algebras or coalgebras we have to pass to certain topological completions.

For example, consider the space $C_b(G)$ denote the space of continuous bounded functions on G endowed with the uniform norm

$$\|f\|_\infty = \mathrm{l.u.b.}_{x \in G}\{|f(x)|\}.$$

We have a bounded linear map

$$c : C_b(G) \to C_b(G \times G)$$

given by

$$c(f)(x, y) := f(xy).$$

In the case that G is finite, and endowed with the discrete topology, the space $C_b(G)$ is just the space of all functions on G, and $C_b(G \times G) = C_b(G) \otimes C_b(G)$ where $C_b(G) \otimes C_b(G)$ can be identified with the space of all functions on $G \times G$ of the form

$$(x, y) \mapsto \sum_i f_i(x) g_i(y)$$

where the sum is finite. In the general case, not every bounded continuous function on $G \times G$ can be written in the above form, but, by Stone-Weierstrass, the space of such functions is dense in $C_b(G \times G)$. So we can say that $C_b(G)$ is "almost" a co-algebra, or a "co-algebra in the topological sense", in that the

map c does not carry C into $C \otimes C$ but rather into the completion of $C \otimes C$. If A denotes the dual space of C, then A becomes an (honest) algebra. To make all this work in the case at hand, we need yet another version of the Riesz representation theorem. I will state and prove the appropriate theorem, but not go into the further details:

Let X be a topological space, let $C_b := C_b(X, \mathbf{R})$ be the space of bounded continuous real valued functions on X. For any $f \in C_b$ and any subset $A \subset X$ let

$$\|f\|_{\infty,A} := \text{l.u.b.}_{x \in A}\{|f(x)|\}.$$

So

$$\|f\|_\infty = \|f\|_{\infty,X}.$$

A continuous linear function ℓ is called **tight** if for every $\delta > 0$ there is a compact set K_δ and a positive number A_δ such that

$$|\ell(f)| \le A_\delta \|f\|_{\infty,K_\delta} + \delta \|f\|_\infty.$$

Theorem 8.8.1 [Yet another Riesz representation theorem.] *If $\ell \in C_b^*$ is a tight non-negative linear functional, then there is a finite non-negative measure μ on $(X, \mathcal{B}(\mathcal{X}))$ such that*

$$\langle \ell, f \rangle = \int_X f d\mu$$

for all $f \in C_b$.

Proof. We need to show that $f_n \searrow 0 \Rightarrow \langle \ell, f_n \rangle \searrow 0$. Given $\epsilon > 0$, choose

$$\delta := \frac{\epsilon}{1 + 2\|f_1\|_\infty}.$$

This same inequality then holds with f_1 replaced by f_n since the f_n are monotone decreasing. We have the K_δ as in the definition of tightness, and by Dini's lemma, we can choose N so that

$$\|f_n|_{K_\delta} \le \frac{\epsilon}{2A_\delta} \quad \forall\, n > N.$$

Then $|\langle \ell, f_n \rangle| \le \epsilon$ for all $n > N$. QED

8.9 Invariant and relatively invariant measures on homogeneous spaces.

Let G be a locally compact Hausdorff topological group, and let H be a closed subgroup. Then G acts on the quotient space G/H by left multiplication, the element a sending the coset xH into axH. By abuse of language, we will continue to denote this action by ℓ_a. So

$$\ell_a(xH) := (ax)H.$$

We can consider the corresponding action on measures

$$\kappa \mapsto \ell_{a*}\kappa.$$

The measure κ is said to be invariant if

$$\ell_{a*}\kappa = \kappa \quad \forall\, a \in G.$$

The measure κ on G/H is said to be **relatively invariant** with **modulus** D if D is a function on G such that

$$\ell_{a*}\kappa = D(a)\kappa \quad \forall\, a \in G.$$

From its definition it follows that

$$D(ab) = D(a)D(b),$$

and it is not hard to see from the ensuing discussion that D is continuous. We will only deal with positive measures here, so D is continuous homomorphism of G into the multiplicative group of real numbers. We call such an object a positive character. The questions we want to address in this section are what are the possible invariant measures or relatively invariant measures on G/H, and what are their modular functions.

For example, consider the $ax + b$ group acting on the real line. So G is the $ax+b$ group, and H is the subgroup consisting of those elements with $b = 0$, the "pure rescalings". So H is the subgroup fixing the origin in the real line, and we can identify G/H with the real line. Let $N \subset G$ be the subgroup consisting of pure translations, so N consists of those elements of G with $a = 1$. The group N acts as translations of the line, and (up to scalar multiple) the only measure on the real line invariant under all translations is Lebesgue measure, dx. But

$$h = \left(\begin{array}{cc} a & 0 \\ 0 & 1 \end{array} \right)$$

acts on the real line by sending $x \mapsto ax$ and hence

$$\ell_{h*}(dx) = a^{-1}dx.$$

(The push forward of the measure μ under the map ϕ assigns the measure $\mu(\phi^{-1}(A))$ to the set A.) So there is no measure on the real line invariant under G. On the other hand, the above formula shows that dx is relatively invariant with modular function

$$D\left(\left(\begin{array}{cc} a & b \\ 0 & 1 \end{array} \right) \right) = a^{-1}.$$

Notice that this is the same as the modular function Δ, of the group G, and that the modular function δ for the subgroup H is the trivial function $\delta \equiv 1$ since H is commutative.

We now turn to the general study, and will follow Loomis in dealing with the integrals rather than the measures, and so denote the Haar integral of G

by I, with Δ its modular function, denote the Haar integral of H by J and its modular function by χ. We will let K denote an integral on G/H, and D a positive character on G.

If $f \in C_0(G)$, then we will let $J_t f(xt)$ denote that function of x obtained by integrating the function $t \mapsto f(xt)$ on H with respect to J. By the left invariance of J, we see that

$$J_t(xst) = J_t(xt).$$

In other words, this function is constant on cosets and hence defines a function on G/H (which is easily seen to be continuous and of compact support). Thus J defines a map

$$\mathbf{J}: \ C_0(G) \to C_0(G/H).$$

We will prove below that this map is surjective.

The main result we are aiming for in this section (due to A. Weil) is

Theorem 8.9.1 *In order that D be the modular function of a relatively invariant integral K on G/H it is necessary and sufficient that*

$$D(s) = \frac{\Delta(s)}{\chi(s)} \quad \forall \ s \in H. \tag{8.28}$$

If this happens, then K is uniquely determined up to scalar multiple, in fact,

$$K(\mathbf{J}(fD)) = cI(f) \quad \forall \ f \in C_0(G) \tag{8.29}$$

where c is some positive constant.

We begin with some preliminaries. Let $\pi : G \to G/H$ denote the projection map which sends each element $x \in G$ into its right coset

$$\pi(x) = xH.$$

The topology on G/H is defined by declaring a set $U \subset (G/H)$ to be open if and only if $\pi^{-1}(U)$ is open. The map π is then not only continuous (by definition) but also open, i.e. sends open sets into open sets. Indeed, if $O \subset G$ is an open subset, then

$$\pi^{-1}(\pi(O)) = \bigcup_{h \in H} Oh$$

which is a union of open sets, hence open, hence $\pi(O)$ is open.

Lemma 8.9.1 *If B is a compact closed subset of G/H then there exists a compact set $A \subset B$ such that*

$$\pi(A) = B,$$

Proof. Since we are assuming that G is locally compact, we can find an open neighborhood O of e in G whose closure C is compact. The sets $\pi(xO)$, $x \in G$

are all open subsets of G/H since π is open, and there images cover all of G/H. In particular, since B is compact, finitely many of them cover B, so

$$B \subset \bigcup_i \pi(x_i U) \subset \bigcup_i \pi(x_i C) = \pi\left(\bigcup_i x_i C\right)$$

the unions being finite. The set

$$K = \bigcup_i x_i C$$

is compact, being the finite union of compact sets. The set $\pi^{-1}(B)$ is closed (since its complement is the inverse image of an open set, hence open). So

$$A := K \cap \pi^{-1}(B)$$

is compact, and its image is B. QED

Proposition 8.9.1 J *is surjective.*

Let $F \in C_0(G/H)$ and let $B = \mathrm{Supp}(F)$. Choose a compact set $A \subset G$ with $\pi(A) = B$ as in the lemma. Choose $\phi \in C_0(G)$ with $\phi \geq 0$ and $\phi > 0$ on A. If

$$x \in AH = \pi^{-1}(B)$$

then $\phi(xh) > 0$ for some $h \in H$, and so $\mathbf{J}(\phi) > 0$ on B. So we may extend the function

$$z \mapsto \frac{F(z)}{\mathbf{J}(h)(z)}$$

to a continuous function, call it γ, by defining it to be zero outside $B = \mathrm{Supp}(F)$. The function $g = \pi^* \gamma$, i.e.

$$g(x) = \gamma(\pi(x))$$

is hence a continuous function on G, and hence

$$f := g\phi$$

is a continuous function of compact support on G. Since g is constant on H cosets,

$$\mathbf{J}(f)(z) = \gamma(z) J(h)(z) = F(z). \text{ QED}$$

Now to the proof of the theorem. Suppose that K is an integral on $C_0(G/H)$ with modular function D. Define

$$M(f) = K(\mathbf{J}(fD))$$

for $f \in C_0(G)$. By applying the monotone convergence theorem for J and K we see that M is an integral. We must check that it is left invariant, and hence

determines a Haar measure which is a multiple of the given Haar measure.

$$
\begin{aligned}
M(\ell_a^*(f) &= K(\mathbf{J}((\ell_a^* f)D)) \\
&= D(a)^{-1} K((\mathbf{J}(\ell_a^*(fD)))) \\
&= D(a)^{-1} K((\ell_a^*(\mathbf{J}(fD)))) \\
&= D(a)^{-1} D(a) K(\mathbf{J}(fD))) \\
&= M(f).
\end{aligned}
$$

This shows that if K is relatively invariant with modular function D then it is unique up to scalar factor. Let us multiply K be a scalar if necessary (which does not change D) so that

$$
I(f) = K(\mathbf{J}(fD)).
$$

We now argue more or less as before: Let $h \in H$. Then

$$
\begin{aligned}
\Delta(h)I(f) &= I(r_h^* f) \\
&= K(\mathbf{J}((r_h^* f)D)) \\
&= D(h) K(\mathbf{J}(r_h^*(fD))) \\
&= D(h)\chi(h) K(\mathbf{J}(fd)) \\
&= D(h)\chi(h) I(f),
\end{aligned}
$$

proving that (8.28) holds.

Conversely, suppose that (8.28) holds, and try to define K by

$$
K(\mathbf{J}(f)) = I(fD^{-1}).
$$

Since $\mathbf{J}$ is surjective, this will define an integral on $C_0(G/H)$ once we show that it is well defined, i.e. once we show that

$$
\mathbf{J}(f) = 0 \Rightarrow I(fD^{-1}) = 0.
$$

Suppose that $\mathbf{J}(f) = 0$, and let $\phi \in C_0(G)$. Then

$$
\phi(x) D(x)^{-1} \pi^*(\mathbf{J}(f))(x) = 0
$$

for all $x \in G$, and so taking I of the above expression will also vanish. We will now use Fubini: We have

$$
0 = I_x\left(\phi(x)D^{-1}(x)J_h(f(xh))\right) = I_x J_h\left(\phi(x)D^{-1}(x)(f(xh))\right) =
$$

$$
J_h I_x\left(\phi(x)D^{-1}(x)(f(xh))\right)
$$

We can write the expression that is inside the last I_x integral as

$$
r_{h^{-1}}^*\left(\phi(xh^{-1})D^{-1}(xh^{-1})(f(x))\right)
$$

and hence

$$J_h I_x \left(\phi(x) D^{-1}(x)(f(xh)) \right) = J_h I_x \left(\phi(xh^{-1}) D^{-1}(xh^{-1})(f(x)\Delta(h^{-1})) \right)$$

by the defining properties of Δ. Now use the hypothesis that $\Delta = D\chi$ to get

$$J_h I_x \left(\chi(h^{-1})\phi(xh^{-1}) D^{-1}(x)(f(x)) \right)$$

and apply Fubini again to write this as

$$I_x \left(D^{-1}(x) f(x) J_h(\chi(h^{-1}\phi(xh^{-1}))) \right).$$

By equation (8.26) applied to the group H, we can replace the J integral above by $J_h(\phi(xh))$ so finally we conclude that

$$I_x \left(D^{-1}(x) f(x) J_h(\phi(xh)) \right) = 0$$

for any $\phi \in C_0(G)$. Now choose $\psi \in C_0(G/H)$ which is non-negative and identically one on $\pi(\mathrm{Supp}(f))$, and choose $\phi \in C_0(G)$ with $\mathbf{J}(\phi) = \psi$. Then the above expression is $I(D^{-1}f)$. So we have proved that

$$\mathbf{J}(f) = 0 \Rightarrow I(fD^{-1}) = 0,$$

and hence that $K : C_0(G/H) \to \mathbf{C}$ defined by

$$K(F) = I(D^{-1}f) \quad \text{if} \quad \mathbf{J}(f) = F$$

is well defined. We still must show that K defined this way is relatively invariant with modular function K. We compute

$$
\begin{aligned}
K(\ell_a^* F) &= I(D^{-1}\ell_a^*(f)) \\
&= D(a)I(\ell_a^*(D^{-1}f)) \\
&= D(a)I(D^{-1}f) \\
&= D(a)K(F). \ \ \text{QED}
\end{aligned}
$$

Of particular importance is the case where G and H are unimodular, for example compact. Then, up to scalar factor, there is a unique measure on G/H invariant under the action of G.

Chapter 9

Banach algebras and the spectral theorem.

In this chapter, all rings will be assumed to be associative and to have an identity element, usually denoted by e. If an element x in the ring is such that $(e - x)$ has a right inverse, then we may write this inverse as $(e - y)$, and the equation

$$(e - x)(e - y) = e$$

expands out to

$$x + y - xy = 0.$$

Following Loomis, we call y the right **adverse** of x and x the left adverse of y. Loomis introduces this term because he wants to consider algebras without identity elements. But it will be convenient to use it even under our assumption that all our algebras have an identity. If an element has both a right and left inverse then they must be equal by the associative law, so if x has a right and left adverse these must be equal. When we say that an element has (or does not have) an inverse, we will mean that it has (or does not have) a two sided inverse. Similarly for adverse.

All algebras will be over the complex numbers. The **spectrum** of an element x in an algebra is the set of all $\lambda \in \mathbf{C}$ such that $(x - \lambda e)$ has no inverse. We denote the spectrum of x by $\mathrm{Spec}(x)$.

Proposition 9.0.2 *If P is a polynomial then*

$$P(\mathrm{Spec}(x)) = \mathrm{Spec}(P(x)). \tag{9.1}$$

Proof. The product of invertible elements is invertible. For any $\lambda \in \mathbf{C}$ write $P(t) - \lambda$ as a product of linear factors:

$$P(t) - \lambda = c \prod (t - \mu_i).$$

Thus

$$P(x) - \lambda e = c \prod (x - \mu_i e)$$

in A and hence $(P(x) - \lambda e)^{-1}$ fails to exist if and only if $(x - \mu_i e)^{-1}$ fails to exist for some i, i.e. $\mu_i \in \mathrm{Spec}(x)$. But these μ_i are precisely the solutions of

$$P(\mu) = \lambda.$$

Thus $\lambda \in \mathrm{Spec}(P(x))$ if and only if $\lambda = P(\mu)$ for some $\mu \in \mathrm{Spec}(x)$ which is precisely the assertion of the proposition. QED

9.1 Maximal ideals.

9.1.1 Existence.

Theorem 9.1.1 *Every proper right ideal in a ring is contained in a maximal proper right ideal. Similarly for left ideals. Also any proper two sided ideal is contained in a maximal proper two sided ideal.*

Proof by Zorn's lemma. The proof is the same in all three cases: Let I be the ideal in question (right left or two sided) and $\mathcal{F}$ be the set of all proper ideals (of the appropriate type) containing I ordered by inclusion. Since e does not belong to any proper ideal, the union of any linearly ordered family of proper ideals is again proper, and so has an upper bound. Now Zorn guarantees the existence of a maximal element. QED

9.1.2 The maximal spectrum of a ring.

For any ring R we let $\mathrm{Mspec}(R)$ denote the set of maximal (proper) two sided ideals of R. For any two sided ideal I we let

$$\mathrm{Supp}(I) := \{M \in \mathrm{Mspec}(R) : I \subset M\}.$$

Notice that

$$\mathrm{Supp}(\{0\}) = \mathrm{Mspec}(R)$$

and

$$\mathrm{Supp}(R) = \emptyset.$$

For any family I_α of two sided ideals, a maximal ideal contains all of the I_α if and only if it contains the two sided ideal $\sum_\alpha I_\alpha$. In symbols

$$\bigcap_\alpha \mathrm{Supp}(I_\alpha) = \mathrm{Supp}\left(\sum_\alpha I_\alpha\right).$$

Thus the intersection of any collection of sets of the form $\mathrm{Supp}(I)$ is again of this form. Notice also that if

$$A = \mathrm{Supp}(I)$$

then

$$A = \mathrm{Supp}(J) \quad \text{where } J = \bigcap_{M \in A} M.$$

(Here $I \subset J$, but J might be a strictly larger ideal.) We claim that

$$A = \text{Supp}\left(\bigcap_{M \in A} M\right) \text{ and } B = \text{Supp}\left(\bigcap_{M \in B} M\right)$$

$$\Rightarrow A \cup B = \text{Supp}\left(\bigcap_{M \in A \cup B} M\right). \tag{9.2}$$

Indeed, if N is a maximal ideal belonging to $A \cup B$ then it contains the intersection on the right hand side of (9.2) so the left hand side contains the right. We must show the reverse inclusion. So suppose the contrary. This means that there is a maximal ideal N which contains the intersection on the right but does not belong to either A or B. Since N does not belong to A, the ideal $J(A) := \bigcap_{M \in A} M$ is not contained in N, so $J(A) + N = R$, and hence there exist $a \in J(A)$ and $m \in N$ such that $a + m = e$. Similarly, there exist $b \in J(B)$ and $n \in N$ such that $b + n = e$. But then

$$e = e^2 = (a + m)(b + n) = ab + an + mb + mn.$$

Each of the last three terms on the right belong to N since it is a two sided ideal, and so does ab since

$$ab \in \left(\bigcap_{M \in A} M\right) \cap \left(\bigcap_{M \in B} M\right) = \left(\bigcap_{M \in A \cup B} M\right) \subset N.$$

Thus $e \in N$ which is a contradiction.

The above facts show that the sets of the form $\text{Supp}(I)$ give the closed sets of a topology.

If $A \subset \text{Mspec}(R)$ is an arbitrary subset, its closure is given by

$$\overline{A} = \text{Supp}\left(\bigcap_{M \in A} M\right).$$

(For the case of commutative rings, a major advance was to replace maximal ideals by prime ideals in the preceding construction - giving rise to the notion of $\text{Spec}(R)$ - the prime spectrum of a commutative ring. But the motivation for this development in commutative algebra came from these constructions in the theory of Banach algebras.)

9.1.3   Maximal ideals in a commutative algebra.

Proposition 9.1.1 *An ideal M in a commutative algebra is maximal if and only if R/M is a field.*

Proof. If J is an ideal in R/M, its inverse image under the projection $R \to R/M$ is an ideal in R. If J is proper, so is this inverse image. Thus M is maximal if

and only if $F := R/M$ has no ideals other than 0 and F. Thus if $0 \neq X \in F$, the set of all multiples of X must be all of F if M is maximal. In particular every non-zero element has an inverse. Conversely, if every non-zero element of F has an inverse, then F has no proper ideals. QED

9.1.4   Maximal ideals in the ring of continuous functions.

Let S be a compact Hausdorff space, and let $\mathcal{C}(S)$ denote the ring of continuous complex valued functions on S. For each $p \in S$, the map of $\mathcal{C}(S) \to \mathbf{C}$ given by

$$f \mapsto f(p)$$

is a surjective homomorphism. The kernel of this map consists of all f which vanish at p. By the preceding proposition, this is then a maximal ideal, which we shall denote by M_p.

Theorem 9.1.2 *If I is a proper ideal of $\mathcal{C}(S)$, then there is a point $p \in S$ such that*

$$I \subset M_p.$$

In particular every maximal ideal in $\mathcal{C}(S)$ is of the form M_p so we may identify $\mathrm{Mspec}(\mathcal{C}(S))$ *with S as a set. This identification is a homeomorphism between the original topology of S and the topology given above on* $\mathrm{Mspec}(\mathcal{C}(S))$.

Proof. Suppose that for every $p \in S$ there is an $f \in I$ such that $f(p) \neq 0$. Then $|f|^2 = f\overline{f} \in I$ and $|f(p)|^2 > 0$ and $|f|^2 \geq 0$ everywhere. Thus each point of S is contained in a neighborhood U for which there exists a $g \in I$ with $g \geq 0$ everywhere, and $g > 0$ on U. Since S is compact, we can cover S with finitely many such neighborhoods. If we take h to be the sum of the corresponding g's, then $h \in I$ and $h > 0$ everywhere. So $h^{-1} \in \mathcal{C}(S)$ and $e = 1 = hh^{-1} \in I$ so $I = \mathcal{C}(S)$, a contradiction. This proves the first part of the the theorem.

To prove the last statement, we must show that the closure of any subset $A \subset S$ in the original topology coincides with its closure in the topology derived from the maximal ideal structure. That is, we must show that

$$\text{closure of } A \text{ in the topology of } S = \mathrm{Supp}\left(\bigcap_{M \in A} M\right).$$

Now

$$\bigcap_{M \in A} M$$

consists exactly of all continuous functions which vanish at all points of A. Any such function must vanish on the closure of A in the topology of S. So the left hand side of the above equation is contained in the right hand side. We must show the reverse inclusion. Suppose $p \in S$ does not belong to the closure of A in the topology of S. Then Urysohn's Lemma asserts that there is an $f \in \mathcal{C}(S)$ which vanishes on A and $f(p) \neq 0$. Thus $p \notin \mathrm{Supp}(\bigcap_{M \in A} M)$. QED

Theorem 9.1.3 *Let I be an ideal in $\mathcal{C}(S)$ which is closed in the uniform topology on $\mathcal{C}(S)$. Then*

$$I = \bigcap_{M \in \text{Supp}(I)} M.$$

Proof. $\text{Supp}(I)$ consists of all points p such that $f(p) = 0$ for all $f \in I$. Since f is continuous, the set of zeros of f is closed, and hence $\text{Supp}(I)$ being the intersection of such sets is closed. Let O be the complement of $\text{Supp}(I)$ in S. Then O is a locally compact space, and the elements of $\bigcap_{M \in \text{Supp}(I)} M$ when restricted to O consist of all functions which vanish at infinity. I, when restricted to O is a uniformly closed subalgebra of this algebra. If we could show that the elements of I separate points in O then the Stone-Weierstrass theorem would tell us that I consists of all continuous functions on O which "vanish at infinity", i.e. all continuous functions which vanish on $\text{Supp}(I)$, which is the assertion of the theorem. So let p and q be distinct points of O, and let $f \in \mathcal{C}(S)$ vanish on $\text{Supp}(I)$ and at q with $f(p) = 1$. Such a function exists by Urysohn's Lemma, again. Let $g \in I$ be such that $g(p) \neq 0$. Such a g exists by the definition of $\text{Supp}(I)$. Then $gf \in I$, $(gf)(q) = 0$, and $(gf)(p) \neq 0$. QED

9.2 Normed algebras.

A **normed algebra** is an algebra (over the complex numbers) which has a norm as a vector space which satisfies

$$\|xy\| \leq \|x\|\|y\|. \tag{9.3}$$

Since $e = ee$ this implies that

$$\|e\| \leq \|e\|^2$$

so

$$\|e\| \geq 1.$$

Consider the new norm

$$\|y\|_N := \text{lub}_{\|x\| \neq 0} \|yx\|/\|x\|.$$

This still satisfies (9.3) and from (9.3) we have

$$\|y\|_N \leq \|y\|.$$

Under this norm we have

$$\|e\|_N = 1.$$

On the other hand, from its definition

$$\|y\|/\|e\| \leq \|y\|_N.$$

Combining this with the previous inequality gives

$$\|y\|/\|e\| \leq \|y\|_N \leq \|y\|.$$

In other words the norms $\| \ \|$ and $\| \ \|_N$ are equivalent. So with no loss of generality we can add the requirement

$$\|e\| = 1 \tag{9.4}$$

to our axioms for a normed algebra.

Suppose we weaken our condition and allow $\| \ \|$ to be only a pseudo-norm. this means that we allow the possible existence of non-zero elements x with $\|x\| = 0$. Then (9.3) implies that the set of all such elements is an ideal, call it I. Then $\| \ \|$ descends to A/I . Furthermore, any continuous (i.e. bounded) linear function must vanish on I so also descends to A/I with no change in norm. In other words, A^* can be identified with $(A/I)^*$.

If A is a normed algebra which is complete (i.e. A is a Banach space as a normed space) then we say that A is a **Banach algebra**.

9.3 The Gelfand representation.

Let A be a normed vector space. The space A^* of continuous linear functions on A becomes a normed vector space under the norm

$$\|\ell\| := \sup_{\|x\|\neq 0} |\ell(x)|/\|x\|.$$

Each $x \in A$ defines a linear function on A^* by

$$x(\ell) := \ell(x)$$

and

$$|x(\ell)| \leq \|\ell\|\|x\|$$

so x is a continuous function of ℓ (relative to the norm introduced above on A^*).

Let $B = B_1(A^*)$ denote the unit ball in A^*. In other words $B = \{\ell : \|\ell\| \leq 1\}$. The functions $x(\cdot)$ on B induce a topology on B called the **weak topology**.

Proposition 9.3.1 *B is compact under the weak topology.*

Proof. For each $x \in A$, the values assumed by the set of $\ell \in B$ at x lie in the closed disk $D_{\|x\|}$ of radius $\|x\|$ in **C**. Thus

$$B \subset \prod_{x \in A} D_{\|x\|}$$

which is compact by Tychonoff's theorem - being the product of compact spaces. To prove that B is compact, if is sufficient to show that B is a closed subset of this product space. Suppose that f is in the closure of B. For any x and y in A and any $\epsilon > 0$, we can find an $\ell \in B$ such that

$$|f(x) - \ell(x)| < \epsilon, \quad |f(y) - \ell(y)| < \epsilon, \text{ and } |f(x + y) - \ell(x + y)| < \epsilon.$$

Since $\ell(x + y) = \ell(x) + \ell(y)$ this implies that

$$|f(x + y) - f(x) - f(y)| < 3\epsilon.$$

Since ϵ is arbitrary, we conclude that

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

Similarly, $f(\lambda x) = \lambda f(x)$. In other words, $f \in B$. QED

Now let A be a normed algebra. Let $\Delta \subset A^*$ denote the set of all continuous homomorphisms of A onto the complex numbers. In other words, in addition to being linear, we demand of $h \in \Delta$ that

$$h(xy) = h(x)h(y) \quad \text{and} \quad h(e) = 1.$$

Let $E := h^{-1}(1)$. Then E is closed under multiplication. In particular, if $x \in E$ we can not have $\|x\| < 1$ for otherwise x^n is a sequence of elements in E tending to 0, and so by the continuity of h we would have $h(0) = 1$ which is impossible. So $\|x\| \geq 1$ for all $x \in E$. If y is such that $h(y) = \lambda \neq 0$, then $x := y/\lambda \in E$ so

$$\|h(y)| \leq \|y\|,$$

and this clearly also holds if $h(y) = 0$. In other words,

$$\Delta \subset B.$$

Since the conditions for being a homomorphism will hold for any weak limit of homomorphisms (the same proof as given above for the compactness of B), we conclude that Δ is compact.

Once again we can turn the tables and think of $y \in A$ as a function $\hat{y}$ on Δ via

$$\hat{y}(h) := h(y).$$

This map from A into an algebra of functions on Δ is called the **Gelfand representation**.

The inequality $h(y) \leq \|y\|$ for all h translates into

$$\|\hat{y}\|_\infty \leq \|y\|. \tag{9.5}$$

Putting it all together we get

Theorem 9.3.1 Δ *is a compact subset of A^* and the Gelfand representation $y \mapsto \hat{y}$ is a norm decreasing homomorphism of A onto a subalgebra $\hat{A}$ of $\mathcal{C}(\Delta)$.*

The above theorem is true for any normed algebra - we have not used any completeness condition. For Banach algebras, i.e. complete normed algebras, we can proceed further and relate Δ to $\mathrm{Mspec}(A)$. Recall that an element of $\mathrm{Mspec}(A)$ corresponds to a homomorphism of A onto some field. In the commutative Banach algebra case we will show that this field is $\mathbf{C}$ and that any such homomorphism is automatically continuous. So for commutative Banach algebras we can identify Δ with $\mathrm{Mspec}(A)$.

9.3.1 Invertible elements in a Banach algebra form an open set.

In this section A will be a Banach algebra.

Proposition 9.3.2 *If $\|x\| < 1$ then x has an adverse x' given by*

$$x' = -\sum_{n=1}^{\infty} x^n$$

so that $e - x$ has an inverse given by

$$e - x' = e + \sum_{n=1}^{\infty} x^n.$$

Both are continuous functions of x

Proof. Let

$$s_n := -\sum_{1}^{n} x^i.$$

Then if $m < n$

$$\|s_m - s_n\| \leq \sum_{m+1}^{n} \|x\|^i < \|x\|^m \frac{1}{1 - \|x\|} \to 0$$

as $m \to \infty$. Thus s_n is a Cauchy sequence and

$$x + s_n - x s_n = x^{n+1} \to 0.$$

Thus the series $-\sum_{1}^{\infty} x^i$ as stated in the theorem converges and gives the adverse of x and is continuous function of x. The corresponding statements for $(e - x)^{-1}$ now follow. QED

The proof shows that the adverse x' of x satisfies

$$\|x'\| \leq \frac{\|x\|}{1 - \|x\|}. \tag{9.6}$$

Theorem 9.3.2 *Let y be an invertible element of A and set*

$$a := \frac{1}{\|y^{-1}\|}.$$

Then $y + x$ is invertible whenever

$$\|x\| < a.$$

Furthermore

$$\|(x + y)^{-1} - y^{-1}\| \leq \frac{\|x\|}{(a - \|x\|)a}. \tag{9.7}$$

Thus the set of elements having inverses is open and the map $x \to x^{-1}$ is continuous on its domain of definition.

Proof. If $\|x\| < \|y^{-1}\|^{-1}$ then

$$\|y^{-1}x\| \le \|y^{-1}\|\|x\| < 1.$$

Hence $e + y^{-1}x$ has an inverse by the previous proposition. Hence $y + x = y(e + y^{-1}x)$ has an inverse. Also

$$(y + x)^{-1} - y^{-1} = \left((e + y^{-1}x)^{-1} - e\right)y^{-1} = -(-y^{-1}x)'y^{-1}$$

where $(-y^{-1}x)'$ is the adverse of $-y^{-1}x$.

From (9.6) and the above expression for $(x + y)^{-1} - y^{-1}$ we see that

$$\|(x + y)^{-1} - y^{-1}\| \le \|(-y^{-1}x)'\|\|y^{-1}\| \le \frac{\|x\|\|y^{-1}\|^2}{1 - \|x\|\|y^{-1}\|} = \frac{\|x\|}{a(a - \|x\|)}.$$

QED

Proposition 9.3.3 *If I is a proper ideal then $\|e - x\| \ge 1$ for all $x \in I$.*

Proof. Otherwise there would be some $x \in I$ such that $e - x$ has an adverse, i.e. x has an inverse which contradicts the hypothesis that I is proper.

Proposition 9.3.4 *The closure of a proper ideal is proper. In particular, every maximal ideal is closed.*

Proof. The closure of an ideal I is clearly an ideal, and all elements in the closure still satisfy $\|e - x\| \ge 1$ and so the closure is proper. QED

Proposition 9.3.5 *If I is a closed ideal in A then A/I is again a Banach algebra.*

Proof. The quotient of a Banach space by a closed subspace is again a Banach space. The norm on A/I is given by

$$\|X\| = \min_{x \in X} \|x\|$$

where X is a coset of I in A. The product of two cosets X and Y is the coset containing xy for any $x \in X$, $y \in Y$. Thus

$$\|XY\| = \min_{x \in X,\ y \in Y} \|xy\| \le \min_{x \in X,\ y \in Y} \|x\|\|y\| = \|X\|\|Y\|.$$

Also, if E is the coset containing e then E is the identity element for A/I and so

$$\|E\| \le 1.$$

But we know that this implies that $\|E\| = 1$. QED

Suppose that A is commutative and M is a maximal ideal of A. We know that A/M is a field, and the preceding proposition implies that A/M is a normed field containing the complex numbers. The following famous result implies that A/M is in fact norm isomorphic to $\mathbf{C}$. It deserves a subsection of its own:

The Gelfand-Mazur theorem.

A division algebra is a (possibly not commutative) algebra in which every non-zero element has an inverse.

Theorem 9.3.3 *Every normed division algebra over the complex numbers is isometrically isomorphic to the field of complex numbers.*

Let A be the normed division algebra and $x \in A$. We must show that $x = \lambda e$ for some complex number λ. Suppose not. Then by the definition of a division algebra, $(x - \lambda e)^{-1}$ exists for all $\lambda \in \mathbf{C}$ and all these elements commute. Thus

$$(x - (\lambda + h)e)^{-1} - (x - \lambda e)^{-1} = h(x - (\lambda + h)e)^{-1}(x - \lambda e)^{-1}$$

as can be checked by multiplying both sides of this equation on the right by $x - \lambda e$ and on the left by $x - (\lambda + h)e$. Thus the strong derivative of the function

$$\lambda \mapsto (x - \lambda e)^{-1}$$

exists and is given by the usual formula $(x - \lambda e)^{-2}$. In particular, for any $\ell \in A^*$ the function

$$\lambda \mapsto \ell((x - \lambda e)^{-1})$$

is analytic on the entire complex plane. On the other hand for $\lambda \neq 0$ we have

$$(x - \lambda e)^{-1} = \lambda^{-1}(\frac{1}{\lambda}x - e)^{-1}$$

and this approaches zero as $\lambda \to \infty$. Hence for any $\ell \in A^*$ the function $\lambda \mapsto \ell((x - \lambda e)^{-1})$ is an everywhere analytic function which vanishes at infinity, and hence is identically zero by Liouville's theorem. But this implies that $(x - \lambda e)^{-1} \equiv 0$ by the Hahn Banach theorem, a contradiction. QED

9.3.2 The Gelfand representation for commutative Banach algebras.

Let A be a commutative Banach algebra. We know that every maximal ideal is the kernel of a homomorphism h of A onto the complex numbers. Conversely, suppose that h is such a homomorphism. We claim that

$$|h(x)| \leq \|x\|$$

for any $x \in A$. Indeed, suppose that $|h(x)| > \|x\|$ for some x. Then

$$\|x/h(x)\| < 1$$

so $e - x/h(x)$ is invertible; in particular $h(e - x/h(x)) \neq 0$ which implies that $1 = h(e) \neq h(x)/h(x)$, a contradiction.

In short, we can identify $\mathrm{Mspec}(A)$ with Δ and the map $x \mapsto \hat{x}$ is a norm decreasing map of A onto a subalgebra $\hat{A}$ of $\mathcal{C}(\mathrm{Mspec}(A))$ where we use the uniform norm $\| \ \|_\infty$ on $\mathcal{C}(\mathrm{Mspec}(A))$. A complex number is in the spectrum of an $x \in A$ if and only if $(x - \lambda e)$ belongs to some maximal ideal M, in which case $\hat{x}(M) = \lambda$. Thus

$$\|\hat{x}\|_\infty = \text{l.u.b. } \{|\lambda| : \lambda \in \mathrm{Spec}(x)\}. \tag{9.8}$$

9.3.3 The spectral radius.

The right hand side of (9.8) makes sense in any algebra, and is called the **spectral radius** of x and is denoted by $|x|_{sp}$. We claim that

Theorem 9.3.4 *In any Banach algebra we have*

$$|x|_{sp} = \lim_{n \to \infty} \|x^n\|^{\frac{1}{n}}. \tag{9.9}$$

Proof. If $|\lambda| > \|x\|$ then $e - x/\lambda$ is invertible, and therefore so is $x - \lambda e$ so $\lambda \notin \mathrm{Spec}(x)$. Thus

$$|x|_{sp} \leq \|x\|.$$

We know from (9.1) that $\lambda \in \mathrm{Spec}(x) \Rightarrow \lambda^n \in \mathrm{Spec}(x^n)$, so the previous inequality applied to x^n gives

$$|x|_{sp} \leq \|x^n\|^{\frac{1}{n}}$$

and so

$$|x|_{sp} \leq \liminf \|x^n\|^{\frac{1}{n}}.$$

We must prove the reverse inequality with lim sup. Suppose that $|\mu| < 1/|x|_{sp}$ so that $\mu := 1/\lambda$ satisfies $|\lambda| > |x|_{sp}$ and hence $e - \mu x$ is invertible. The formula for the adverse gives

$$(\mu x)' = -\sum_1^{\infty} (\mu x)^n$$

where we know that this converges in the open disk of radius $1/\|x\|$. However, we know that $(e - \mu x)^{-1}$ exists for $|\mu| < 1/|x|_{sp}$. In particular, for any $\ell \in A^*$ the function $\lambda \mapsto \ell((\mu x)')$ is analytic and hence its Taylor series

$$-\sum \ell(x^n)\mu^n$$

converges on this disk. Here we use the fact that the Taylor series of a function of a complex variable converges on any disk contained in the region where it is analytic. Thus

$$|\ell(\mu^n x^n)| \to 0$$

for each fixed $\ell \in A^*$ if $|\mu| < 1/|x|_{sp}$. Considered as a family of linear functions of ℓ, we see that

$$\ell \mapsto \ell(\mu^n x^n)$$

is bounded for each fixed ℓ, and hence by the *uniform boundedness principle*, there exists a constant K such that

$$\|\mu^n x^n\| < K$$

for each μ in this disk, in other words

$$\|x^n\|^{\frac{1}{n}} \leq K^{\frac{1}{n}}(1/|\mu|)$$

so

$$\limsup \|x^n\|^{\frac{1}{n}} \leq 1/|\mu| \quad \text{if } 1/|\mu| > |x|_{sp}.$$

QED

In a commutative Banach algebra we can combine (9.9) with (9.8) to conclude that

$$\|\hat{x}\|_\infty = \lim_{n\to\infty} \|x^n\|^{\frac{1}{n}}. \tag{9.10}$$

We say that x is a **generalized nilpotent element** if $\lim \|x^n\|^{\frac{1}{n}} = 0$. From (9.9) we see that x is a generalized nilpotent element if and only if $\hat{x} \equiv 0$. This means that x belongs to all maximal ideals. The intersection of all maximal ideals is called the **radical** of the algebra. A Banach algebra is called **semi-simple** if its radical consists only of the 0 element.

9.3.4   The generalized Wiener theorem.

Theorem 9.3.5 *Let A be a commutative Banach algebra. Then $x \in A$ has an inverse if and only if $\hat{x}$ never vanishes.*

Proof. If $xy = e$ then $\hat{x}\hat{y} \equiv 1$. So if x has an inverse, then $\hat{x}$ can not vanish anywhere. Conversely, suppose x does not have an inverse. Then Ax is a proper ideal. So x is contained in some maximal ideal M (by Zorn's lemma). So $\hat{x}(M) = 0$.

Example. Let G be a countable commutative group given the discrete topology. Then we may choose its Haar measure to be the counting measure. Thus $L^1(G)$ consists of all complex valued functions on G which are absolutely summable, i.e. such that

$$\sum_{a\in G} |f(a)| < \infty.$$

Recall that $L^1(G)$ is a Banach algebra under convolution:

$$(f \star g)(x) := \sum_{y\in G} f(y^{-1}x)g(y).$$

We repeat the proof: Since $L^1(G) \subset L^2(G)$ this sum converges and

$$
\begin{aligned}
\sum_{x\in G} |(f \star g)(x)| \quad &\leq \quad \sum_{x,y\in G} |f(xy^{-1})| \cdot |g(y)| \\
&= \quad \sum_{y\in G} |g(y)| \sum_{x\in G} |f(xy^{-1})| \\
&= \quad \sum_{y\in G} |g(y)|)(\sum_{w\in G} |f(w)| \quad \text{i.e.} \\
\|f \star g\| \quad &\leq \quad \|f\|\|g\|.
\end{aligned}
$$

If $\delta_x \in L^1(G)$ is defined by

$$\delta_x(t) = \begin{cases} 1 & \text{if } t = x \\ 0 & \text{otherwise} \end{cases}$$

then

$$\delta_x \star \delta_y = \delta_{xy}.$$

We know that the most general continuous linear function on $L^1(G)$ is obtained from multiplication by an element of $L^\infty(G)$ and then integrating $=$ summing. That is it is given by

$$f \mapsto \sum_{x \in G} f(x)\rho(x)$$

where ρ is some bounded function. Under this linear function we have

$$\delta_x \mapsto \rho(x)$$

and so, if this linear function is to be multiplicative, we must have

$$\rho(xy) = \rho(x)\rho(y).$$

Since $\rho(x^n) = \rho(x)^n$ and $|\rho(x)|$ is to be bounded, we must have $|\rho(x)| \equiv 1$.

A function ρ satisfying these two conditions:

$$\rho(xy) = \rho(x)\rho(y) \quad \text{and} \quad |\rho(x)| \equiv 1$$

is called a **character** of the commutative group G. The space of characters is itself a group, denoted by $\hat{G}$.

We have shown that $\mathrm{Mspec}(L^1(G)) = \hat{G}$. In particular we have a topology on $\hat{G}$ and the Gelfand transform $f \mapsto \hat{f}$ sends every element of $L^1(G)$ to a continuous function on $\hat{G}$.

For example, if $G = \mathbf{Z}$ under addition, the condition to be a character says that

$$\rho(m + n) = \rho(m)\rho(n), \quad |\rho| \equiv 1.$$

So

$$\rho(n) = \rho(1)^n$$

where

$$\rho(1) = e^{i\theta}$$

for some $\theta \in \mathbf{R}/(2\pi\mathbf{Z})$. Thus

$$\hat{f}(\theta) = \sum_{n \in \mathbf{Z}} f(n)e^{in\theta}$$

is just the Fourier series with coefficients $f(n)$. The image of the Gelfand transform is just the set of Fourier series which converge absolutely. We conclude from Theorem 9.3.5 that if F is an absolutely convergent Fourier series which

vanishes nowhere, then $1/F$ has an absolutely convergent Fourier series. Before Gelfand, this was a deep theorem of Wiener.

To deal with the version of this theorem which treats the Fourier transform rather than Fourier series, we would have to consider algebras which do not have an identity element. Most of what we did goes through with only mild modifications, but I do not to go into this, as my goals are elsewhere.

9.4 Self-adjoint algebras.

Let A be a semi-simple commutative Banach algebra. Since "semi-simple" means that the radical is $\{0\}$, we know that the Gelfand transform is injective. A is called **self adjoint** if for every $x \in A$ there exists an $x^* \in A$ such that

$$(x^*)\hat{} = \overline{\hat{x}}.$$

By the injectivity of the Gelfand transform, the element x^* is uniquely specified by this equation.

In general, for any Banach algebra, a map $f \mapsto f^\dagger$ is called an **involutory anti-automorphism** if

- $(fg)^\dagger = g^\dagger f^\dagger$

- $(f + g)^\dagger = f^\dagger + g^\dagger$

- $(\lambda f)^\dagger = \overline{\lambda} f^\dagger$ and

- $(f^\dagger)^\dagger = f$.

For example, if A is the algebra of bounded operators on a Hilbert space, then the map $T \mapsto T^*$ sending every operator to its adjoint is an example of an involutory anti-automorphism. Another example is $L^1(G)$ under convolution, for a locally compact Hausdorff group G where the involution was the map $f \mapsto \tilde{f}$.

If A is a semi-simple self-adjoint commutative Banach algebra, the map $x \mapsto x^*$ is an involutory anti-automorphism. It has this further property:

$$f = f^* \Rightarrow 1 + f^2 \quad \text{is invertible.}$$

indeed, if $f = f^*$ then $\hat{f}$ is real valued, so $1 + \hat{f}^2$ vanishes nowhere, and so $1 + f^2$ is invertible by Theorem 9.3.5. Conversely

Theorem 9.4.1 *Let A be a commutative semi-simple Banach algebra with an involutory anti-automorphism $f \mapsto f^\dagger$ such that $1 + f^2$ is invertible whenever $f = f^\dagger$. Then A is self-adjoint and $\dagger = *$.*

Proof. We must show that $(f^\dagger)\hat{} = \overline{\hat{f}}$. We first prove that if we set

$$g := f + f^\dagger$$

then $\hat{g}$ is real valued. Suppose the contrary, that

$$\hat{g}(M) = a + ib, \quad b \neq 0 \quad \text{for some } M \in \text{Mspec}(A).$$

Now $g^\dagger = f^\dagger + (f^\dagger)^\dagger = g$ and hence $(g^2)^\dagger = g^2$ so

$$h := \frac{ag^2 - (a^2 - b^2)g}{b(a^2 + b^2)}$$

satisfies

$$h^\dagger = h.$$

We have

$$h(M) = \frac{a(a + ib)^2 - (a^2 - b^2)(a + ib)}{b(a^2 + b^2)} = i.$$

So

$$1 + h(M)^2 = 0$$

contradicting the hypothesis that $1 + h^2$ is invertible. Now let us apply this result to $\frac{1}{2}f$ and to $\frac{1}{2i}f$. We have

$$f = g + ih \quad \text{where} \quad g = \frac{1}{2}(f + f^\dagger), \quad h = \frac{1}{2i}(f - f^\dagger)$$

and we know that $\hat{g}$ and $\hat{h}$ are real and satisfy $g^\dagger = g$ and $h^\dagger = h$. So

$$\overline{\hat{f}} = \overline{\hat{g} + i\hat{h}} = \hat{g} - i\hat{h} = (f^\dagger)\hat{\ }.$$

QED

Theorem 9.4.2 *Let A be a commutative Banach algebra with an involutory anti-automorphism $\dagger$ which satisfies the condition*

$$\|ff^\dagger\| = \|f\|^2 \quad \forall f \in A. \tag{9.11}$$

Then the Gelfand transform $f \mapsto \hat{f}$ is a norm preserving surjective isomorphism which satisfies

$$(f^\dagger)\hat{\ } = \overline{\hat{f}}.$$

In particular, A is semi-simple and self-adjoint.

Proof. $\|f\|^2 = \|ff^\dagger\| \leq \|f\|\|f^\dagger\|$ so $\|f\| \leq \|f^\dagger\|$. Replacing f by $f^\dagger$ gives $\|f^\dagger\| \leq \|f\|$. so

$$\|f\| = \|f^\dagger\|$$

and

$$\|ff^\dagger\| = \|f\|^2 = \|f\|\|f^\dagger\|. \tag{9.12}$$

Now since A is commutative,

$$ff^\dagger = f^\dagger f$$

and

$$f^2(f^2)^\dagger = f^2(f^\dagger)^2 = (ff^\dagger)(ff^\dagger)^\dagger \tag{9.13}$$

and so applying (9.12) to f^2 and then applying it once again to f we get

$$\|f^2\|\|(f^\dagger)^2\| = \|ff^\dagger(ff^\dagger)^\dagger\| = \|ff^\dagger\|\|ff^\dagger\| = \|f\|^2\|f^\dagger\|^2$$

or

$$\|f^2\|^2 = \|f\|^4.$$

Thus

$$\|f^2\| = \|f\|^2,$$

and therefore

$$\|f^4\| = \|f^2\|^2 = \|f\|^4$$

and by induction

$$\|f^{2^k}\| = \|f\|^{2^k}$$

for all non-negative integers k.

Hence letting $n = 2^k$ in the right hand side of (9.10) we see that $\|\hat{f}\|_\infty = \|f\|$ so the Gelfand transform is norm preserving, and hence injective. To show that $\dagger = *$ it is enough to show that if $f = f^\dagger$ then $\hat{f}$ is real valued, as in the proof of the preceding theorem. Suppose not, so $\hat{f}(M) = a + ib$, $b \neq 0$. For any real number c we have

$$(f + ice)\hat{}(M) = a + i(b + c)$$

so

$$|(f + ice)\hat{}(M)|^2 = a^2 + (b + c)^2 \leq \|f + ice\|^2 = \|(f + ice)(f - ice)\|$$

$$= \|f^2 + c^2 e\| \leq \|f\|^2 + c^2.$$

This says that

$$a^2 + b^2 + 2bc + c^2 \leq \|f\|^2 + c^2$$

which is impossible if we choose c so that $2bc > \|f\|^2$.

So we have proved that $\dagger = *$. Now by definition, if $f(M) = f(N)$ for all $f \in A$, the maximal ideals M and N coincide. So the image of elements of A under the Gelfand transform separate points of $\mathrm{Mspec}(A)$. But every $f \in A$ can be written as

$$f = \frac{1}{2}(f + f^*) + i\frac{1}{2i}(f - f^*)$$

i.e. as a sum $g + ih$ where $\hat{g}$ and $\hat{f}$ are real valued. Hence the real valued functions of the form $\hat{g}$ separate points of $\mathrm{Mspec}(A)$. Hence by the *Stone Weierstrass theorem* we know that the image of the Gelfand transform is dense in $\mathcal{C}(\mathrm{Mspec}(A))$. Since A is complete and the Gelfand transform is norm preserving, we conclude that the Gelfand transform is surjective. QED

9.4.1 An important generalization.

A Banach algebra with an involution † such that (9.11) holds is called a C^*-**algebra**. Notice that we are *not* assuming that this Banach algebra is commutative. But an element x of such an algebra is called **normal** if

$$xx^\dagger = x^\dagger x,$$

in other words if x *does* commute with $x^\dagger$. Then we can repeat the argument at the beginning of the proof of Theorem 9.4.2 to conclude that if x is a normal element of a C^* algebra, then

$$\|x^{2^k}\| = \|x\|^{2^k}$$

and hence by (9.9)

$$|x|_{sp} = \|x\|. \tag{9.14}$$

An element x of an algebra with involution is called **self-adjoint** if $x^\dagger = x$. In particular, every self-adjoint element is normal.

Again, a rerun of a previous argument shows that if x is self-adjoint, meaning that $x^* = x$ then

$$\mathrm{Spec}(x) \subset \mathbf{R} \tag{9.15}$$

Indeed, suppose that $a + ib \in \mathrm{Spec}(x)$ with $b \neq 0$, and let

$$y := \frac{1}{b}(x - ae)$$

so that $y = y^*$ and $i \in \mathrm{Spec}(y)$. So $e + iy$ is not invertible. So for any real number r,

$$(r + 1)e - (re - iy) = e + iy$$

is not invertible. This implies that

$$|r + 1| \leq \|re - iy\|$$

and so

$$(r + 1)^2 \leq \|re - iy\|^2 = \|(re - iy)(re + iy)\|^2.$$

by (9.11). Thus

$$(r + 1)^2 \leq \|r^2 e + y^2\|^2 \leq r^2 + \|y\|^2$$

which is not possible if $2r - 1 > \|y\|^2$.

So we have proved:

Theorem 9.4.3 *Let A be a C^* algebra. If $x \in A$ is normal, then*

$$|x|_{sp} = \|x\|.$$

If $x \in A$ is self-adjoint, then

$$\mathrm{Spec}(x) \subset \mathbf{R}.$$

9.4.2 An important application.

Proposition 9.4.1 *If T is a bounded linear operator on a Hilbert space, then*

$$\|TT^*\| = \|T\|^2. \tag{9.16}$$

In other words, the algebra of all bounded operators on a Hilbert space is a C^*-algebra under the involution $T \mapsto T^*$.

Proof.

$$
\begin{aligned}
\|TT^*\| &= \sup_{\|\phi\|=1} \|TT^*\phi\| \\
&= \sup_{\|\phi\|=1,\|\psi\|=1} |(TT^*\phi,\psi)| \\
&= \sup_{\|\phi\|=1,\|\psi\|=1} |(T^*\phi,T^*\psi)| \\
&\geq \sup_{\|\phi\|=1} (T^*\phi,T^*\phi) \\
&= \|T^*\|^2
\end{aligned}
$$

so

$$\|T^*\|^2 \leq \|TT^*\| \leq \|T\|\|T^*\|$$

so

$$\|T^*\| \leq \|T\|.$$

Reversing the role of T and T^* gives the reverse inequality so $\|T\| = \|T^*\|$. Inserting into the preceding inequality gives

$$\|T^2\| \leq \|TT^*\| \leq \|T\|^2$$

so we have the equality (9.16). QED

Thus the map $T \mapsto T^*$ sending every bounded operator on a Hilbert space into its adjoint is an anti-involution on the Banach algebra of all bounded operators, and it satisfies (9.11). We can thus apply Theorem 9.4.2 to conclude:

Theorem 9.4.4 *Let B be any commutative subalgebra of the algebra of bounded operators on a Hilbert space which is closed in the strong topology and with the property that $T \in B \Rightarrow T^* \in B$. Then the Gelfand transform $T \mapsto \hat{T}$ gives a norm preserving isomorphism of B with $\mathcal{C}(\mathcal{M})$ where $\mathcal{M} = \mathrm{Mspec}(B)$. Furthermore, $(T^*)\hat{} = \overline{\hat{T}}$ for all $T \in B$. In particular, if T is self-adjoint, then $\hat{T}$ is real valued.*

9.5 The Spectral Theorem for Bounded Normal Operators, Functional Calculus Form.

As a special case of the definition we gave earlier, a (bounded) operator T on a Hilbert space **H** is called **normal** if

$$TT^* = T^*T.$$

We can then consider the subalgebra of the algebra of all bounded operators which is generated by e, the identity operator, T and T^*. Take the closure, B, of this algebra in the strong topology. We can apply the preceding theorem to B to conclude that B is isometrically isomorphic to the algebra of all continuous functions on the compact Hausdorff space

$$\mathcal{M} = \mathrm{Mspec}(B).$$

Remember that a point of $\mathcal{M}$ is a homomorphism $h : B \to \mathbf{C}$ and that

$$h(T^*) = (T^*)\hat{}(h) = \overline{\hat{T}(h)} = \overline{h(T)}.$$

Since h is a homomorphism, we see that h is determined on the algebra generated by e, T and T^* by the value $h(T)$, and since it is continuous, it is determined on all of B by the knowledge of $h(T)$. We thus get a map

$$\mathcal{M} \to \mathbf{C}, \qquad \mathcal{M} \ni h \mapsto h(T), \tag{9.17}$$

and we know that this map is injective. Now

$$h(T - h(T)e) = h(T) - h(T) = 0$$

so $T - h(T)e$ belongs to the maximal ideal h, and hence is not invertible in the algebra B. Thus the image of our map lies in $\mathrm{Spec}_B(T)$. Here I have added the subscript B, because our general definition of the spectrum of an element T in an algebra B consists of those complex numbers z such that $T - ze$ does not have an inverse in B. If we let A denote the algebra of *all* bounded operators on $\mathbf{H}$, it is logically conceivable that $T - ze$ has an inverse in A which does not lie in B. In fact, this can not happen. But this requires a proof, which I will give later. So for the time being I will stick with the temporary notation Spec_B.

So the map $h \mapsto h(T)$ actually maps $\mathcal{M}$ to the subset $\mathrm{Spec}_B(T)$ of the complex plane. If $\lambda \in \mathrm{Spec}_B(T)$, then by definition, $T - \lambda e$ is not invertible in B, so lies in some maximal ideal h, so $\lambda \in \mathrm{Spec}_B(T)$. So the map (9.17) maps $\mathcal{M}$ onto $\mathrm{Spec}_B(T)$. Since the topology on $\mathcal{M}$ is inherited from the weak topology, the map $h \mapsto h(T)$ is continuous. Since $\mathcal{M}$ is compact and $\mathbf{C}$ is Hausdorff, the fact that h is bijective and continuous implies that h^{-1} is also continuous. Indeed we must show that if U is an open subset of $\mathcal{M}$, then $h(U)$ is open in $\mathrm{Spec}_B(T)$. But $\mathcal{M} \setminus U$ is compact, hence its image under the continuous map h is compact, hence closed, and so the complement of this image which is $h(U)$ is open.

Thus h is a homeomorphism, and hence we have a norm-preserving $*$-isomorphism $T \mapsto \hat{T}$ of B with the algebra of continuous functions on $\mathrm{Spec}_B(T)$. Furthermore the element T corresponds the function $z \mapsto z$ (restricted to $\mathrm{Spec}_B(T)$).

Since B is determined by T, let me use the notation $\sigma(T)$ for $\mathrm{Spec}_B(T)$, postponing until later the proof that $\sigma(T) = \mathrm{Spec}_A(T)$. Let me set

$$\phi = h^{-1}$$

and now follow the customary notation in Hilbert space theory and use I to denote the identity operator.

9.5.1   Statement of the theorem.

Theorem 9.5.1 *Let T be a bounded normal operator on a Hilbert space $\mathbf{H}$. Let $B(T)$ denote the closure in the strong topology of the algebra generated by I, T and T^*. Then there is a unique continuous $*$ isomorphism*

$$\phi : \sigma(T) \to B(T)$$

such that
$$\phi(\mathbf{1}) = I$$

and
$$\phi(z \mapsto z) = T.$$

Furthermore,
$$T\psi = \lambda\psi, \quad \psi \in \mathbf{H} \;\Rightarrow\; \phi(f)\psi = f(\lambda)\psi. \tag{9.18}$$

If $f \in C(\sigma(T))$ is real valued then $\phi(f)$ is self-adjoint, and if $f \geq 0$ then $\phi(f) \geq 0$ as an operator.

The only facts that we have not yet proved (aside from the big issue of proving that $\sigma(T) = \mathrm{Spec}_A(T)$) are (9.18) and the assertions which follow it. Now (9.18) is clearly true if we take f to be a polynomial, in which case $\phi(f) = f(T)$. Then just apply the Stone-Weierstrass theorem to conclude (9.18) for all f. If f is real then $f = \bar{f}$ and therefore $\phi(f) = \phi(f)^*$. If $f \geq 0$ then we can find a real valued $g \in C(\sigma(T)$ such that $f = g^2$ and the square of a self-adjoint operator is non-negative. QED

In view of this theorem, there is a more suggestive notation for the map ϕ. Since the image of the monomial z is T, and since the image of any polynomial P (thought of as a function on $\sigma(T)$) is $P(T)$, we are safe in using the notation

$$f(T) := \phi(f)$$

for any $f \in C(\sigma(T))$.

9.5.2 $\mathrm{Spec}_B(T) = \mathrm{Spec}_A(T)$.

Here is the main result of this section:

Theorem 9.5.2 *Let A be a C^* algebra and let B be a subalgebra of A which is closed under the involution. Then for any $x \in B$ we have*

$$\mathrm{Spec}_B(x) = \mathrm{Spec}_A(x). \tag{9.19}$$

Remarks:

1.   Applied to the case where A is the algebra of all bounded operators on a Hilbert space, and where B is the closed subalgebra by I, T and T^* we get the spectral theorem for normal operators as promised.

2. If $T - ze$ has no inverse in A it has no inverse in B. So

$$\mathrm{Spec}_A(x) \subset \mathrm{Spec}_B(x).$$

We must show the reverse inclusion. We begin by formulating some general results and introducing some notation.

For any associative algebra A we let $G(A)$ denote the set of elements of A which are invertible (the "group-like" elements).

Proposition 9.5.1 *Let B be a Banach algebra, and let $x_n \in G(B)$ be such that $x_n \to x$ and $x \notin G(B)$. Then*

$$\|x_n^{-1}\| \to \infty.$$

Proof. Suppose not. Then there is some $C > 0$ and a subsequence of elements (which we will relabel as x_n) such that

$$\|x_n^{-1}\| < C.$$

Then

$$x = x_n(e + x_n^{-1}(x - x_n))$$

with $x - x_n \to 0$. In particular, for n large enough

$$\|x_n^{-1}\| \cdot \|x - x_n\| < 1,$$

so $(e + x_n^{-1}(x - x_n))$ is invertible as is x_n and so x is invertible contrary to hypothesis. QED

Proposition 9.5.2 *Let B be a closed subalgebra of a Banach algebra A containing the unit e. Then*

- *$G(B)$ is the union of some of the components of $B \cap G(A)$.*

- *If $x \in B$ then $\mathrm{Spec}_B(x)$ is the union of $\mathrm{Spec}_A(x)$ and a (possibly empty) collection of bounded components of $\mathbf{C} \setminus \mathrm{Spec}_A(x)$.*

- *If $\mathrm{Spec}_A(x)$ does not separate the complex plane then*

$$\mathrm{Spec}_B(x) = \mathrm{Spec}_A(x).$$

Proof. We know that $G(B) \subset G(A) \cap B$ and both are open subsets of B. We claim that $G(A) \cap B$ contains no boundary points of $G(B)$. If x were such a boundary point, then $x = \lim x_n$ for $x_n \in G(B)$, and by the continuity of the map $y \mapsto y^{-1}$ in A we conclude that $x_n^{-1} \to x^{-1}$ in A, so in particular the $\|x_n\|^{-1}$ are bounded which is impossible by Proposition 9.5.1. Let O be a component of $B \cap G(A)$ which intersects $G(B)$, and let U be the complement of $\overline{G(B)}$. Then $O \cap G(B)$ and $O \cap U$ are open subsets of B and since O does not intersect $\partial G(B)$, the union of these two disjoint open sets is O. So $O \cap U$

is empty since we assumed that O is connected. Hence $O \subset G(B)$. This proves the first assertion.

For the second assertion, let us fix the element x, and let

$$G_A(x) = \mathbf{C} \setminus \mathrm{Spec}_A(x)$$

so that $G_A(x)$ consists of those complex numbers z for which $x - ze$ is invertible in A, with a similar notation for $G_B(x)$. Both of these are open subsets of $\mathbf{C}$ and $G_B(x) \subset G_A(x)$. Furthermore, as before, $G_B(x) \subset G_A(x)$ and $G_A(x)$ can not contain any boundary points of $G_B(x)$. So again, $G_B(x)$ is a union of some of the connected components of $G_A(x)$. Therefore $\mathrm{Spec}_B(x)$ is the union of $\mathrm{Spec}_A(x)$ and the remaining components. Since $\mathrm{Spec}_B(x)$ is bounded, it will not contain any unbounded components.

The third assertion follows immediately from the second. QED

But now we can prove Theorem 9.5.2. We need to show that if $x \in B$ is invertible in A then it is invertible in B. If x is invertible in A then so are x^* and xx^*. But xx^* is self-adjoint, hence its spectrum is a bounded subset of $\mathbf{R}$, so does not separate $\mathbf{C}$. Since $0 \notin \mathrm{Spec}_A(xx^*)$ we conclude from the last assertion of the proposition that $0 \notin \mathrm{Spec}_B(xx^*)$ so xx^* has an inverse in B. But then

$$x^*(xx^*)^{-1} \in B$$

and

$$x\left(x^*(xx^*)^{-1}\right) = e.$$

QED.

9.5.3 A direct proof of the spectral theorem.

I started out this chaper with the general theory of Banach algebras, went to the Gelfand representation theorem, the special properties of C^* algebras, and then some general facts about how the spectrum of an element can vary with the algebra containing it. I took this route because of the impact the Gelfand representation theorem had on the course of mathematics, especially in algebraic geometry. But the key ideas are

- (9.9), which, for a bounded operator T on a Banach space says that

$$\max_{\lambda \in \mathrm{Spec}(T)} |\lambda| = \lim_{n \to \infty} \|T^n\|^{\frac{1}{n}}$$

,

- (9.16) which says that if T is a bounded operator on a Hilbert space then $\|TT^*\| = \|T\|^2$, and

- If T is a bounded operator on a Hilbert space and $TT^* = T^*T$ then it follows from (9.16) that

$$\|T^{2^k}\| = \|T\|^{2^k}.$$

We could prove these facts by the arguments given above and conclude that if T is a normal bounded operator on a Hilbert space then

$$\max_{\lambda \in \operatorname{Spec}(T)} |\lambda| = \|T\|. \tag{9.20}$$

Suppose (for simplicity) that T is self-adjoint: $T = T^*$. Then the argument given several times above shows that $\operatorname{Spec}(T) \subset \mathbf{R}$. Let P be a polynomial. Then (9.1) combined with the preceding equation says that

$$\|P(T)\| = \max_{\lambda \in \operatorname{Spec}(T)} |P(\lambda)|. \tag{9.21}$$

The norm on the right is the restriction to polynomials of the uniform norm $\|\cdot\|_\infty$ on the space $C(\operatorname{Spec}(T))$.

Now the map

$$P \mapsto P(T)$$

is a homomorphism of the ring of polynomials into bounded normal operators on our Hilbert space satisfying

$$\overline{P} \mapsto P(T)^*$$

and

$$\|P(T)\| = \|P\|_{\infty, \operatorname{Spec}(T)}.$$

The Weierstrass approximation theorem then allows us to conclude that this homomorphism extends to the ring of continuous functions on $\operatorname{Spec}(T)$ with all the properties stated in Theorem 9.5.1 .

Chapter 10

The spectral theorem.

The purpose of this chapter is to give a more thorough discussion of the spectral theorem, especially for unbounded self-adjoint operators,

We begin by giving a slightly different discussion showing how the Gelfand representation theorem, especially for algebras with involution, implies the spectral theorem for bounded self-adjoint (or, more generally normal) operators on a Hilbert space. Recall that an operator T is called **normal** if it commutes with T^*. More generally, an element T of a Banach algebra A with involution is called normal if $TT^* = T^*T$.

Let B be the closed commutative subalgebra generated by e, T and T^*. We know that B is isometrically isomorphic to the ring of continuous functions on $\mathcal{M}$, the space of maximal ideals of B, which is the same as the space of continuous homomorphisms of B into the complex numbers. We shall show again that $\mathcal{M}$ can also be identified with $\mathrm{Spec}_A(T)$, the set of all $\lambda \in \mathbf{C}$ such that $(T - \lambda e)$ is not invertible. This means that every continuous continuous function $\hat{f}$ on $\mathcal{M} = \mathrm{Spec}_A(T)$ corresponds to an element f of B. In the case where A is the algebra of bounded operators on a Hilbert space, we will show that this homomorphism extends to the space of Borel functions on $\mathrm{Spec}_A(T)$.(In general the image of the extended homomorphism will lie in A, but not necessarily in B.) We now restrict attention to the case where A is the algebra of bounded operators on a Hilbert space.

If U is a Borel subset of $\mathrm{Spec}_A(T)$, let us denote the element of A corresponding to $\mathbf{1}_U$ by $P(U)$. Then

$$P(U)^2 = P(U) \quad \text{and} \quad P(U)^* = P(U)$$

so $P(U)$ is a self adjoint (i.e. "orthogonal") projection. Also, if $U \cap V = \emptyset$ then $P(U)P(V) = 0$ and

$$P(U \cup V) = P(U) + P(V).$$

Thus $U \mapsto P(U)$ is finitely additive. In fact, it is completely additive in the weak sense that for any pair of vectors x, y in our Hilbert space H the map

$$\mu_{x,y} : \quad U \mapsto (P(U)x, y)$$

is a complex valued measure on $\mathcal{M}$. We shall prove these results in partially reversed order, in that we first prove the existence of the complex valued measure $\mu_{x,y}$ using the Riesz representation theorem describing all continuous linear functions on $C(\mathcal{M})$, and then deduce the existence of the **resolution of the identity** or **projection valued measure**

$$U \mapsto P(U),$$

(more precise definition below) from which we can recover T. The key tool, in addition to the Gelfand representation theorem and the Riesz theorem describing all continuous linear functions on $C(\mathcal{M})$ as being signed measures is our old friend, the Riesz representation theorem for continuous linear functions on a Hilbert space.

10.1   Resolutions of the identity.

In this section B denotes a closed commutative self-adjoint subalgebra of the algebra of all bounded linear operators on a Hilbert space H. (Self-adjoint means that $T \in B \;\Rightarrow T^* \in B$.) By the Gelfand representation theorem we know that B is isometrically isomorphic to $C(\mathcal{M})$ under a map we denote by

$$T \mapsto \hat{T}$$

and we know that

$$T^* \mapsto \overline{\hat{T}}.$$

Fix

$$x, y \in H.$$

The map

$$\hat{T} \mapsto (Tx, y)$$

is a linear function on $C(\mathcal{M})$ with

$$|(Tx, y)| \leq \|T\|\|x\|\|y\| = \|\hat{T}\|_\infty \|x\|\|y\|.$$

In particular it is a continuous linear function on $C(\mathcal{M})$. Hence, by the Riesz representation theorem, there exists a unique complex valued bounded measure

$$\mu_{x,y}$$

such that

$$(Tx, y) = \int_{\mathcal{M}} \hat{T} d\mu_{x,y} \quad \forall \hat{T} \in C(\mathcal{M}).$$

When $\hat{T}$ is real, $T = T^*$ so $(Tx, y) = (x, Ty) = \overline{(Ty, x)}$. The uniqueness of the measure implies that

$$\mu_{y,x} = \overline{\mu_{x,y}}.$$

Thus, for each fixed Borel set $U \subset \mathcal{M}$ its measure $\mu_{x,y}(U)$ depends linearly on x and anti-linearly on y. We have

$$\mu(\mathcal{M}) = \int_{\mathcal{M}} \mathbf{1} d\mu_{x,y} = (ex, y) = (x, y)$$

so

$$|\mu_{x,y}(\mathcal{M})| \leq \|x\| \|y\|.$$

So if f is any bounded Borel function on $\mathcal{M}$, the integral

$$\int_{\mathcal{M}} f d\mu_{x,y}$$

is well defined, and is bounded in absolute value by $\|f\|_\infty \|x\| \|y\|$. If we hold f and x fixed, this integral is a bounded anti-linear function of y, and hence by the Riesz representation theorem there exists a $w \in H$ such that this integral is given by (w, y). The w in question depends linearly on f and on x because the integral does, and so we have defined a linear map O from bounded Borel functions on $\mathcal{M}$ to bounded operators on H such that

$$(O(f)x, y) = \int_{\mathcal{M}} f d\mu_{x,y}$$

and

$$\|O(f)\| \leq \|f\|_\infty.$$

On continuous functions we have

$$O(\hat{T}) = T$$

so O is an extension of the inverse of the Gelfand transform from continuous functions to bounded Borel functions. So we know that O is multiplicative and takes complex conjugation into adjoint when restricted to continuous functions. Let us prove these facts for all Borel functions. If f is real we know that $(O(f)y, x)$ is the complex conjugate of $(O(f)x, y)$ since $\mu_{y,x} = \overline{\mu_{x,y}}$. Hence $O(f)$ is self-adjoint if f is real from which we deduce that

$$O(\overline{f}) = O(f)^*.$$

Now to the multiplicativity: For $S, T \in B$ we have

$$\int_{\mathcal{M}} \hat{S}\hat{T} d\mu_{x,y} = (STx, y) = \int_{\mathcal{M}} \hat{S} d\mu_{Tx,y}.$$

Since this holds for all $\hat{S} \in C(\mathcal{M})$ (for fixed T, x, y) we conclude by the uniqueness of the measure that

$$\mu_{Tx,y} = \hat{T}\mu_{x,y}.$$

Therefore, for any bounded Borel function f we have

$$(Tx, O(f)^*y) = (O(f)Tx, y) = \int_{\mathcal{M}} f d\mu_{Tx,y} = \int_{\mathcal{M}} \hat{T} f d\mu_{x,y}.$$

This holds for all $\hat{T} \in C(\mathcal{M})$ and so by the uniqueness of the measure again, we conclude that

$$\mu_{x,O(f)^*y} = f\mu_{x,y}$$

and hence

$$(O(fg)x, y) = \int_{\mathcal{M}} gf d\mu_{x,y} = \int_{\mathcal{M}} g d\mu_{x,O(f)^*y} = (O(g)x, O(f)^*y) = (O(f)O(g)x, y)$$

or

$$O(fg) = O(f)O(g)$$

as desired.

We have now extended the homomorphism from $C(\mathcal{M})$ to A to a homomorphism from the bounded Borel functions on $\mathcal{M}$ to bounded operators on H.

Now define:

$$P(U) := O(\mathbf{1}_U)$$

for any Borel set U. The following facts are immediate:

1. $P(\emptyset) = 0$

2. $P(\mathcal{M}) = e$ the identity

3. $P(U \cap V) = P(U)P(V)$ and $P(U)^* = P(U)$. In particular, $P(U)$ is a self-adjoint projection operator.

4. If $U \cap V = \emptyset$ then $P(U \cup V) = P(U) + P(V)$.

5. For each fixed $x, y \in H$ the set function $P_{x,y} : U \mapsto (P(U)x, y)$ is a complex valued measure.

Such a P is called a **resolution of the identity**. It follows from the last item that for any fixed $x \in H$, the map $U \mapsto P(U)x$ is an H valued measure.

We have shown that any commutative closed self-adjoint subalgebra B of the algebra of bounded operators on a Hilbert space H gives rise to a unique resolution of the identity on $\mathcal{M} = \text{Mspec}(B)$ such that

$$T = \int_{\mathcal{M}} \hat{T} dP \tag{10.1}$$

in the "weak" sense that

$$(Tx, y) = \int_{\mathcal{M}} \hat{T} d\mu_{x,y} \quad \mu_{x,y}(U) = (P(U)x, y).$$

Actually, given any resolution of the identity we can give a meaning to the integral

$$\int_{\mathcal{M}} f dP$$

for any bounded Borel function f in the strong sense as follows: if

$$s = \sum \alpha_i \mathbf{1}_{U_i}$$

is a simple function where

$$\mathcal{M} = U_1 \cup \cdots \cup U_n, \quad U_i \cap U_j = \emptyset, \;\; i \neq j$$

and $\alpha_1, \ldots, \alpha_n \in \mathbf{C}$, define

$$O(s) := \sum \alpha_i P(U_i) =: \int_{\mathcal{M}} s dP.$$

This is well defined on simple functions (is independent of the expression) and is multiplicative

$$O(st) = O(s)O(t).$$

Also, since the $P(U)$ are self adjoint,

$$O(\bar{s}) = O(s)^*.$$

It is also clear that O is linear and

$$(O(s)x, y) = \int_{\mathcal{M}} s dP_{x,y}.$$

As a consequence, we get

$$\|O(s)x\|^2 = (O(s)^*O(s)x, x) = \int_{\mathcal{M}} |s|^2 dP_{x,x}$$

so

$$\|O(s)x)\|^2 \leq |s\|_\infty \|x\|^2.$$

If we choose i such that $|\alpha_i| = \|s\|_\infty$ and take $x = P(U_i)y \neq 0$, then we see that

$$\|O(s)\| = \|s\|_\infty$$

provided we now take $\|f\|_\infty$ to denote the **essential supremum** which means the following:

It follows from the properties of a resolution of the identity that if U_n is a sequence of Borel sets such that $P(U_n) = 0$, then $P(U) = 0$ if $U = \bigcup U_n$. So if f is any complex valued Borel function on $\mathcal{M}$, there will exist a largest open subset $V \subset \mathbf{C}$ such that $P(f^{-1}(V)) = 0$. We define the **essential range** of f to be the complement of V, say that f is **essentially bounded** if its essential range is compact, and then define its essential supremum $\|f\|_\infty$ to be the supremum of $|\lambda|$ for λ in the essential range of f. Furthermore we identify two essentially bounded functions f and g if $\|f - g\|_\infty = 0$ and call the corresponding space $L^\infty(P)$.

Every element of $L^\infty(P)$ can be approximated in the $\| \ \|_\infty$ norm by simple functions, and hence the integral

$$O(f) = \int_{\mathcal{M}} f dp$$

is defined as the strong limit of the integrals of the corresponding simple functions. The map $f \mapsto O(f)$ is linear, multiplicative, and satisfies

$$O(\overline{f}) = O(f)^*$$

and

$$\|O(f)\| = \|f\|_\infty$$

as before.

If S is a bounded operator on H which commutes with all the $O(f)$ then it commutes with all the $P(U) = O(\mathbf{1}_U)$. Conversely, if S commutes with all the $P(U)$ it commutes with all the $O(s)$ for s simple and hence with all the $O(f)$.

Putting it all together we have:

Theorem 10.1.1 *Let B be a commutative closed self adjoint subalgebra of the algebra of all bounded operators on a Hilbert space H. Then there exists a resolution of the identity P defined on $\mathcal{M} = \mathrm{Mspec}(B)$ such that (10.1) holds. The map $\hat{T} \mapsto T$ of $C(\mathcal{M}) \to B$ given by the inverse of the Gelfand transform extends to a map O from $L^\infty(P)$ to the space of bounded operators on H*

$$O(f) = \int_{\mathcal{M}} f dP.$$

Furthermore, $P(U) \neq 0$ for any non-empty open set U and an operator S commutes with every element of B if and only if it commutes with all the $P(U)$ in which case it commutes with all the $O(f)$.

We must prove the last two statements. If U is open, we may choose $T \neq 0$ such that $\hat{T}$ is supported in U (by Urysohn's lemma). But then (10.1) implies that $T = 0$, a contradiction.

For any bounded operator S and any $x, y \in H$ and $T \in B$ we have

$$(STx, y) = (Tx, S^*y) = \int \hat{T} dP_{x, S^*y}$$

while

$$(TSx, y) = \int \hat{T} dP_{Sx, y}.$$

If $ST = TS$ for all $T \in B$ this means that the measures $P_{Sx, y}$ and P_{x, S^*y} are the same, which means that

$$(P(U)Sx, y) = (P(U)x, S^*y) = (SP(U)x, y)$$

for all x and y which means that

$$SP(U) = P(U)S$$

for all U. We already know that $SP(U) = P(U)S$ for all S implies that $SO(f) = O(f)S$ for all $f \in L^\infty(P)$. QED

10.2 The spectral theorem for bounded normal operators.

Let T be a bounded operator on a Hilbert space H satisfying

$$TT^* = T^*T.$$

Recall that such an operator is called normal. Let B be the closure of the algebra generated by e, T and T^*. We can apply the theorem of the preceding section to this algebra. The one useful additional fact is that *we may identify* $\mathcal{M}$ *with* $\mathrm{Spec}(T)$. Indeed, define the map

$$\mathcal{M} \to \mathrm{Spec}(T)$$

by

$$h \mapsto h(T).$$

We know that $h(T - h(T)e) = 0$ so $T - h(T)e$ lies in the maximal ideal corresponding to h and so is not invertible, consequently $h(T) \in \mathrm{Spec}(T)$. So the map is indeed into $\mathrm{Spec}(T)$.

If $\lambda \in \mathrm{Spec}(T)$ then by definition $T - \lambda e$ is not invertible, hence lie in some maximal ideal, hence $\lambda = h(T)$ for some h so this map is surjective. If $h_1(T) = h_2(T)$ then $h_1(T^*) = \overline{h_1(T)} = h_2(T^*)$. Since h_1 agrees with h_2 on T and T^* they agree on all of B, hence $h_1 = h_2$. In other words the map $h \mapsto h(T)$ is injective. From the definition of the topology on $\mathcal{M}$ it is continuous. Since $\mathcal{M}$ is compact, this implies that it is a homeomorphism. QED

Thus in the theorem of the preceding section, we may replace $\mathcal{M}$ by $\mathrm{Spec}\,T$ when B is the closed algebra generated by T and T^* where T is a normal operator.

In the case that T is a self-adjoint operator, we know that $\mathrm{Spec}\,T \subset \mathbf{R}$, so our resolution of the identity P is defined as a projection valued measure on $\mathbf{R}$ and (10.1) gives a bounded selfadjoint operator as

$$T = \int \lambda dP$$

relative to a resolution of the identity defined on $\mathbf{R}$.

10.3 Stone's formula.

Let T be a bounded self-adjoint operator. We know that

$$T = \int_{\mathbf{R}} \lambda dP(\lambda)$$

for some projection valued measure P on $\mathbf{R}$. We also know that every bounded Borel function on $\mathbf{R}$ gives rise to an operator. In particular, if z is a complex number which is not real, the function

$$\lambda \mapsto \frac{1}{\lambda - z}$$

is bounded, and hence corresponds to a bounded operator

$$R(z, T) = \int_{\mathbf{R}} (z - \lambda)^{-1} dP(\lambda).$$

Since

$$(ze - T) = \int_{\mathbf{R}} (z - \lambda) dP(\lambda)$$

and our homomorphism is multiplicative, we have

$$R(z, T) = (z - T)^{-1}.$$

A conclusion of the above argument is that this inverse does indeed exist for all non-real z. The operator (valued function) $R(z, T)$ is called the **resolvent** of T. **Stone's formula** gives an expression for the projection valued measure in terms of the resolvent. It says that for any real numbers $a < b$ we have

$$\text{s-}\lim_{\epsilon \to 0} \frac{1}{2\pi i} \int_a^b [R(\lambda - i\epsilon, T) - R(\lambda + i\epsilon, T)] \, d\lambda = \frac{1}{2} \left(P((a, b)) + P([a, b]) \right).$$
$$(10.2)$$

Although this formula cries out for a "complex variables" proof, and I plan to give one later, we can give a direct "real variables" proof in terms of what we already know. Indeed, let

$$f_\epsilon(x) := \frac{1}{2\pi i} \int_a^b \left(\frac{1}{x - \lambda - i\epsilon} - \frac{1}{x - \lambda + i\epsilon} \right) d\lambda.$$

We have

$$f_\epsilon(x) = \frac{1}{\pi} \int_a^b \frac{\epsilon}{(x - \lambda)^2 + \epsilon^2} d\lambda = \frac{1}{\pi} \left(\arctan \left[\frac{x - a}{\epsilon} \right] - \arctan \left[\frac{x - b}{\epsilon} \right] \right).$$

The expression on the right is uniformly bounded, and approaches zero if $x \notin [a, b]$, approaches $\frac{1}{2}$ if $x = a$ or $x = b$, and approaches 1 if $x \in (a, b)$. In short,

$$\lim_{\epsilon \to 0} f_\epsilon(x) = \frac{1}{2} (\mathbf{1}_{(a,b)} + \mathbf{1}_{[a,b]}).$$

We may apply the dominated convergence theorem to conclude Stone's formula. QED

10.4 Unbounded operators.

Many important operators in Hilbert space that arise in physics and mathematics are "unbounded". For example the operator $D = \frac{1}{i} \frac{d}{dx}$ on $L_2(\mathbf{R})$. This operator is not defined on all of L_2, and where it is defined it is not bounded as an operator. One of the great achievements of Wintner in the late 1920's,

followed by Stone and von Neumann was to prove a version of the spectral theorem for unbounded self-adjoint operators.

There are two (or more) approaches we could take to the proof of this theorem. Both involve the resolvent

$$R_z = R(z, T) = (zI - T)^{-1}. \tag{10.3}$$

After spending some time explaining what an unbounded operator is and giving the very subtle definition of what an unbounded self-adjoint operator is, we will prove that the resolvent of a self-adjoint operator exists and is a bounded normal operator for all non-real z.

We could then apply the spectral theorem for bounded normal operators to derive the spectral theorem for unbounded self-adjoint operators. This is the fastest approach, but depends on the whole machinery of the Gelfand representation theorem that we have developed so far. Or, we could could prove the spectral theorem for unbounded self-adjoint operators directly using (a mild modification of) Stone's formula. We will present both methods. In the second method we will follow the treatment by Lorch.

10.5 Operators and their domains.

Let B and C be Banach spaces. We make $B \oplus C$ into a Banach space via

$$\|\{x, y\}\| = \|x\| + \|y\|.$$

Here we are using $\{x, y\}$ to denote the ordered pair of elements $x \in B$ and $y \in C$ so as to avoid any conflict with our notation for scalar product in a Hilbert space. So $\{x, y\}$ is just another way of writing $x \oplus y$.

A subspace

$$\Gamma \subset B \oplus C$$

will be called a **graph** (more precisely a graph of a linear transformation) if

$$\{0, y\} \in \Gamma \quad \Rightarrow \quad y = 0.$$

Another way of saying the same thing is

$$\{x, y_1\} \in \Gamma \quad \text{and} \quad \{x, y_2\} \in \Gamma \quad \Rightarrow \quad y_1 = y_2.$$

In other words, if $\{x, y\} \in \Gamma$ then y is determined by x. So let

$D(\Gamma)$ denote the set of all $x \in B$ such that there is a $y \in C$ with $\{x, y\} \in \Gamma$.

Then $D(\Gamma)$ is a linear subspace of B, but, and this is very important, $D(\Gamma)$ is *not* necessarily a closed subspace. We have a linear map

$$T(\Gamma): \ D(\Gamma) \to C, \quad Tx = y \quad \text{where} \quad \{x, y\} \in \Gamma.$$

Equally well, we could start with the linear transformation: Suppose we are given a (not necessarily closed) subspace $D(T) \subset B$ and a linear transformation

$$T : D(T) \to C.$$

We can then consider its graph $\Gamma(T) \subset B \oplus C$ which consists of all

$$\{x, Tx\}.$$

Thus the notion of a graph, and the notion of a linear transformation defined only on a subspace of B are logically equivalent. When we start with T (as usually will be the case) we will write $D(T)$ for the domain of T and $\Gamma(T)$ for the corresponding graph. There is a certain amount of abuse of language here, in that when we write T, we mean to include $D(T)$ and hence $\Gamma(T)$ as part of the definition.

A linear transformation is said to be **closed** if its graph is a closed subspace of $B \oplus C$. Let us disentangle what this says for the operator T. It says that if $f_n \in D(T)$ then

$$f_n \to f \;\; \text{and} \;\; Tf_n \to g \;\; \Rightarrow \;\; f \in D(T) \;\; \text{and} \;\; Tf = g.$$

This is a much weaker requirement than continuity. Continuity of T would say that $f_n \to f$ alone would imply that Tf_n converges to Tf. Closedness says that if we know that *both* f_n converges and $g_n = Tf_n$ converges then we can conclude that $f = \lim f_n$ lies in $D(T)$ and that $Tf = g$.

An important theorem, known as the *closed graph theorem* says that if T is closed and $D(T)$ is all of B then T is bounded. As we will not need to use this theorem in this lecture, we will not present its proof here.

10.6   The adjoint.

Suppose that we have a linear operator $T : D(T) \to C$ and let us make the hypothesis that

$$D(T) \; \textit{is dense in } B.$$

Any element of B^* is then completely determined by its restriction to $D(T)$. Now consider

$$\Gamma(T)^* \subset C^* \oplus B^*$$

defined by

$$\{\ell, m\} \in \Gamma(T)^* \;\; \Leftrightarrow \;\; \langle \ell, Tx \rangle = \langle m, x \rangle \;\; \forall \, x \in D(T). \qquad (10.4)$$

Since m is determined by its restriction to $D(T)$, we see that $\Gamma^* = \Gamma(T^*)$ is indeed a graph. (It is easy to check that it is a linear subspace of $C^* \oplus B^*$.) In other words we have defined a linear transformation

$$T^* := T(\Gamma(T)^*)$$

whose domain consists of all $\ell \in C^*$ such that there exists an $m \in B^*$ for which
$\langle \ell, Tx \rangle = \langle m, x \rangle \quad \forall \, x \in D(T)$.

If $\ell_n \to \ell$ and $m_n \to m$ then the definition of convergence in these spaces implies that for any $x \in D(T)$ we have

$$\langle \ell, Tx \rangle = \lim \langle \ell_n, Tx \rangle = \lim \langle m_n, x \rangle = \langle m, x \rangle.$$

If we let x range over all of $D(T)$ we conclude that Γ^* is a closed subspace of $C^* \oplus B^*$. In other words we have proved

Theorem 10.6.1 *If $T : D(T) \to C$ is a linear transformation whose domain $D(T)$ is dense in B, it has a well defined adjoint T^* whose graph is given by (10.4). Furthermore T^* is a closed operator.*

10.7 Self-adjoint operators.

Now let us restrict to the case where $B = C = H$ is a Hilbert space, so we may identify $B^* = C^* = H^*$ with H via the Riesz representation theorem. If $T : D(T) \to H$ is an operator with $D(T)$ dense in H we may identify the domain of T^* as consisting of all $\{g, h\} \in H \oplus H$ such that

$$(Tx, g) = (x, h) \quad \forall x \in D(T)$$

and then write

$$(Tx, g) = (x, T^*g) \quad \forall \, x \in D(T), \ \ g \in D(T^*).$$

We now come to the central definition: An operator A defined on a domain $D(A) \subset H$ is called **self-adjoint** if

- $D(A)$ is dense in H,

- $D(A) = D(A^*)$, and

- $Ax = A^*x \; \forall x \in D(A)$.

The conditions about the domain $D(A)$ are rather subtle, and we shall go into some of their subtleties in a later lecture. For the moment we record one immediate consequence of the theorem of the preceding section:

Proposition 10.7.1 *Any self adjoint operator is closed.*

If we combine this proposition with the closed graph theorem which asserts that a closed operator defined on the whole space must be bounded, we derive a famous theorem of Hellinger and Toeplitz which asserts that any self adjoint operator defined on the whole Hilbert space must be bounded. This shows that for self-adjoint operators, being globally defined and being bounded amount to the same thing. At the time of its appearance in the second decade of the twentieth century, this theorem of Hellinger and Toeplitz was considered an astounding result. It was only after the work of Banach, in particular the proof of the closed graph theorem, that this result could be put in proper perspective.

10.8   The resolvent.

The following theorem will be central for us.

Theorem 10.8.1 *Let A be a self-adjoint operator on a Hilbert space H with domain D = D(A). Let*

$$c = \lambda + i\mu, \quad \mu \neq 0$$

be a complex number with non-zero imaginary part. Then

$$(cI - A) : D(A) \to H$$

is bijective. Furthermore the inverse transformation

$$(cI - A)^{-1} : H \to D(A)$$

is bounded and in fact

$$\|(cI - A)^{-1}\| \leq \frac{1}{|\mu|}. \tag{10.5}$$

Remark. In the case of a *bounded* self adjoint operator this is an immediate consequence of the spectral theorem, more precisely of the fact that Gelfand transform is an isometric isomorphism of the closed algebra generated by A with the algebra $C(\mathrm{Spec}\, A)$. Indeed, the function $\lambda \mapsto 1/(c - \lambda)$ is bounded on the whole real axis with supremum $1/|\mu|$. Since $\mathrm{Spec}(A) \subset \mathbf{R}$ we conclude that $(cI - A)^{-1}$ exists and its norm satisfies (10.5). We will now give a direct proof of this theorem valid for general self-adjoint operators, and will use this theorem for the proof of the spectral theorem in the general case.

Proof.   Let $g \in D(A)$ and set

$$f = (cI - A)g = [\lambda I - A]g + i\mu g.$$

Then $\|f\|^2 = (f, f) =$

$$\|[cI - A]g\|^2 + \mu^2\|g\|^2 + ([\lambda I - A]g, i\mu g) + (i\mu g, [\lambda I - A]g).$$

I claim that these last two terms cancel. Indeed, since $g \in D(A)$ and A is self adjoint we have

$$(\mu g, [\lambda I - A]g) = (\mu[\lambda I - A]g, g) = ([\lambda I - A]g, \mu g)$$

since μ is real. Hence

$$([\lambda I - A]g, i\mu g) = -i(\mu g, [\lambda I - A]g).$$

We have thus proved that

$$\|f\|^2 = \|(\lambda I - A)g\|^2 + \mu^2\|g\|^2. \tag{10.6}$$

In particular

$$\|f\|^2 \geq \mu^2 \|g\|^2$$

for all $g \in D(A)$. Since $|\mu| > 0$, we see that $f = 0 \Rightarrow g = 0$ so $(cI - A)$ is injective on $D(A)$, and furthermore that that $(cI - A)^{-1}$ (which is defined on im $(cI - A)$)satisfies (10.5). We must show that this image is all of H.

First we show that the image is dense. For this it is enough to show that there is no $h \neq 0 \in H$ which is orthogonal to im $(cI - A)$. So suppose that

$$([cI - A]g, h) = 0 \quad \forall g \in D(A).$$

Then

$$(g, \bar{c}h) = (cg, h) = (Ag, h) \quad \forall g \in D(A)$$

which says that $h \in D(A^*)$ and $A^*h = \bar{c}h$. But A is self adjoint so $h \in D(A)$ and $Ah = \bar{c}h$. Thus

$$\bar{c}(h, h) = (\bar{c}h, h) = (Ah, h) = (h, Ah) = (h, \bar{c}h) = c(h, h).$$

Since $c \neq \bar{c}$ this is impossible unless $h = 0$. We have now established that the image of $cI - A$ is dense in H.

We now prove that it is all of H. So let $f \in H$. We know that we can find

$$f_n = (cI - A)g_n, \quad g_n \in D(A) \quad \text{with } f_n \to f.$$

The sequence f_n is convergent, hence Cauchy, and from (10.5) applied to elements of $D(A)$ we know that

$$\|g_m - g_n\| \leq |\mu|^{-1}\|f_n - f_m\|.$$

Hence the sequence $\{g_n\}$ is Cauchy, so $g_n \to g$ for some $g \in H$. But we know that A is a closed operator. Hence $g \in D(A)$ and $(cI - A)g = f$. QED

The operator
$$R_z = R_z(A) = (zI - A)^{-1}$$

is called the **resolvent** of A when it exists as a bounded operator. The set of $z \in \mathbf{C}$ for which the resolvent exists is called the **resolvent set** and the complement of the resolvent set is called the **spectrum** of the operator. The preceding theorem asserts that the spectrum of a self-adjoint operator is a subset of the real numbers.

Let z and w both belong to the resolvent set. We have

$$wI - A = (w - z)I + (zI - A).$$

Multiplying this equation on the left by R_w gives

$$I = (w - z)R_w + R_w(zI - A),$$

and multiplying this on the right by R_z gives

$$R_z - R_w = (w - z)R_w R_z.$$

From this it follows (interchanging z and w) that $R_z R_w = R_w R_z$, in other words all resolvents R_z commute with one another and we can also write the preceding equation as

$$R_z - R_w = (w - z)R_z R_w. \tag{10.7}$$

This equation, which is known as the **resolvent equation** dates back to the theory of integral equations in the nineteenth century.

It follows from the resolvent equation that $z \mapsto R_z$ (for fixed A) is a continuous function of z. Once we know that the resolvent is a continuous function of z, we may divide the resolvent equation by $(z - w)$ if $z \neq w$ and, if w is interior to the resolvent set, conclude that

$$\lim_{z \to w} \frac{R_z - R_w}{z - w} = -R_w^2.$$

This says that the "derivative in the complex sense" of the resolvent exists and is given by $-R_z^2$. In other words, the resolvent is a "holomorphic operator valued" function of z.

To emphasize this holomorphic character of the resolvent, we have

Proposition 10.8.1 *Let z belong to the resolvent set. The the open disk of radius $\|R_z\|^{-1}$ about z belongs to the resolvent set and on this disk we have*

$$R_w = R_z(I + (z - w)R_z + (z - w)^2 R_z^2 + \cdots). \tag{10.8}$$

Proof.   The series on the right converges in the uniform topology since $|z - w| < \|R_z\|^{-1}$. Multiplying this series by $(zI - A) - (z - w)I$ gives I. But $zI - A - (z - w)I = wI - A$. So the right hand side is indeed R_w. QED

This suggests that we can develop a "Cauchy theory" of integration of functions such as the resolvent, and we shall do so, eventually leading to a proof of the spectral theorem for unbounded self-adjoint operators.

However we first give a proof (following the treatment in Reed-Simon) in which we derive the spectral theorem for unbounded operators from the Gelfand representation theorem applied to the closed algebra generated by the *bounded* normal operators $(\pm iI - A)^{-1}$.

10.9   The multiplication operator form of the spectral theorem.

We first state this theorem for closed commutative self-adjoint algebras of (bounded) operators. Recall that "self-adjoint" in this context means that if $T \in B$ then $T^* \in B$.

Theorem 10.9.1 *Let B be a commutative closed self-adjoint subalgebra of the algebra of all bounded operators on a separable Hilbert space H. Then there exists a measure space $(M, \mathcal{F}, \mu)$ with $\mu(M) < \infty$, a unitary isomorphism*

$$W : H \to L_2(M, \mu),$$

and a map

$$B \to \quad \text{bounded measurable functions on } M, \quad T \mapsto \tilde{T}$$

such that

$$[(WTW^{-1})f](m) = \tilde{T}(m)f(m).$$

In fact, M can be taken to be a finite or countable disjoint union of $\mathcal{M} = \mathrm{Mspec}(B)$

$$M = \bigcup_1^N \mathcal{M}_i, \quad \mathcal{M}_i = \mathcal{M}$$

$N \in \mathbf{Z}_+ \cup \infty$ and

$$\tilde{T}(m) = \hat{T}(m) \quad \text{if } m \in \mathcal{M}_i = \mathcal{M}.$$

In short, the theorem says that any such B is isomorphic to an algebra of multiplication operators on an L_2 space. We prove the theorem in two stages.

10.9.1 Cyclic vectors.

An element $x \in H$ is called a **cyclic vector** for B if $Bx = H$. In more mundane terms this says that the space of linear combinations of the vectors Tx, $T \in B$ are dense in H.

For example, if B consists of all multiples of the identity operator, then Bx consists of all multiples of x, so B can not have a cyclic vector unless H is one dimensional. More generally, if H is finite dimensional and B is the algebra generated by a self-adjoint operator, then B can not have a cyclic vector if A has a repeated eigenvalue.

Proposition 10.9.1 *Suppose that x is a cyclic vector for B. Then it is a cyclic vector for the projection valued measure P on $\mathcal{M}$ associated to B in the sense that the linear combinations of the vectors $P(U)x$ are dense in H as U ranges over the Borel sets on $\mathcal{M}$.*

Proof. Suppose not. Then there exists a non-zero $y \in H$ such that

$$(P(U)x, y) = 0$$

for all Borel subset U of $\mathcal{M}$. Then

$$(Tx, y) = \int_M \hat{T} d(P(U)x, y) = 0$$

which contradicts the assumption that the linear combinations of the Tx are dense in H. QED

Let us continue with the assumption that x is a cyclic vector for B. Let

$$\mu = \mu_{x,x}$$

so

$$\mu(U) = (P(U)x, x).$$

This is a finite measure on $\mathcal{M}$, in fact

$$\mu(\mathcal{M}) = \|x\|^2. \tag{10.9}$$

We will construct a unitary isomorphism of H with $L_2(\mathcal{M}, \mu)$ starting with the assignment

$$Wx = \mathbf{1} = \mathbf{1}_{\mathcal{M}}.$$

We would like this to be a B morphism, even a morphism for the action of multiplication by bounded Borel functions. This forces the definition

$$WP(U)x = \mathbf{1}_U\mathbf{1} = \mathbf{1}_U.$$

This then forces

$$W[c_1 P(U_1)x + \cdots c_n P(U_n)x] = s$$

for any simple function

$$s = c_1 \mathbf{1}_{U_1} + \cdots c_n \mathbf{1}_{U_n}.$$

A direct check shows that this is well defined for simple functions. We can write this map as

$$W[O(s)x] = s,$$

and another direct check shows that

$$\|W[O(s)x]\| = \|s\|_2$$

where the norm on the right is the L_2 norm relative to the measure μ. Since the simple functions are dense in $L_2(\mathcal{M}, \mu)$ and the vectors $O(s)x$ are dense in H this extends to a unitary isomorphism of H onto $L_2(M, \mu)$. Furthermore,

$$W^{-1}(f) = O(f)x$$

for any $f \in L_2(\mathcal{M}, \mu)$. For simple functions, and therefore for all $f \in L_2(\mathcal{M}, \mu)$ we have

$$W^{-1}(\hat{T}f) = O(\hat{T}f)x = TO(f)x = TW^{-1}(f)$$

or

$$(WTW^{-1})f = \hat{T}f$$

which is the assertion of the theorem. In other words we have proved the theorem under the assumption of the existence of a cyclic vector.

10.9.2   The general case.

Start with any non-zero vector x_1 and consider $H_1 = Bx_1 =$ the closure of linear combinations of Tx_1, $T \in B$. The space H_1 is a closed subspace of H which is invariant under B, i.e. $TH_1 \subset H_1 \ \forall T \in B$. Therefore the space $H_1^\perp$ is also invariant under B since if $(x_1, y) = 0$ then

$$(x_1, Ty) = (T^* x_1, y) = 0 \ \text{ since } T^* \in B.$$

Now if $H_1 = H$ we are done, since x_1 is a cyclic vector for B acting on H_1. If not choose a non-zero $x_2 \in H_2$ and repeat the process. We can choose a collection of non-zero vectors z_i whose linear combinations are dense in H - this is the separability assumption. So we may choose our x_i to be obtained from orthogonal projections applied to the z_i. In other words we have

$$H = H_1 \oplus H_2 \oplus H_3 \oplus \cdots$$

where this is either a finite or a countable Hilbert space (completed) direct sum.

Let us also take care to choose our x_n so that

$$\sum \|x_n\|^2 < \infty$$

which we can do, since $c_n x_n$ is just as good as x_n for any $c_n \neq 0$. We have a unitary isomorphism of H_n with $L_2(\mathcal{M}, \mu_n)$ where $\mu_n(U) = (P(U)x_n, x_n)$. In particular,

$$\mu_n(\mathcal{M}) = \|x_n\|^2.$$

So if we take M to be the disjoint union of copies $\mathcal{M}_n$ of $\mathcal{M}$ each with measure μ_n then the total measure of M is finite and

$$L_2(M) = \bigoplus L_2(\mathcal{M}_n, \mu_n)$$

where this is either a finite direct sum or a (Hilbert space completion of) a countable direct sum. Thus the theorem for the cyclic case implies the theorem for the general case. QED

10.9.3   The spectral theorem for unbounded self-adjoint operators, multiplication operator form.

We now let A be a (possibly unbounded) self-adjoint operator, and we apply the previous theorem to the algebra generated by the bounded operators $(\pm iI - A)^{-1}$ which are the adjoints of one another. Observe that there is no non-zero vector $y \in H$ such that

$$(A + iI)^{-1} y = 0.$$

Indeed if such a $y \in H$ existed, we would have

$$0 = (x, (A + iI)^{-1} y) = ((A - iI)^{-1} x, y) = -((iI - A)^{-1} x, y) \quad \forall \, x \in H$$

and we know that the image of $(iI - A)^{-1}$ is $D(A)$ which is dense in H.

Now consider the function $\big((A + iI)^{-1}\big)^{\tilde{}}$ on M given by Theorem 10.9.1. It can not vanish on any set of positive measure, since any function supported on such a set would be in the kernel of the operator consisting of multiplication by $\big((A + iI)^{-1}\big)^{\tilde{}}$.

Thus the function

$$\tilde{A} := \big[\big((A + iI)^{-1}\big)^{\tilde{}}\,\big]^{-1} - i$$

is finite almost everywhere on M relative to the measure μ although it might (and generally will) be unbounded. Our plan is to show that under the unitary isomorphism W the operator A goes over into multiplication by $\tilde{A}$.

First we show

Proposition 10.9.2 $x \in D(A)$ if and only if $\tilde{A}Wx \in L_2(M, \mu)$.

Proof. Suppose $x \in D(A)$. Then $x = (A + iI)^{-1}y$ for some $y \in H$ and so

$$Wx = \big((A + iI)^{-1}\big)^{\tilde{}}f, \quad f = Wy.$$

But

$$\tilde{A}\big((A + iI)^{-1}\big)^{\tilde{}} = 1 - ih$$

where

$$h = \big((A + iI)^{-1}\big)^{\tilde{}}$$

is a bounded function. Thus $\tilde{A}Wx \in L_2(M, \mu)$.

Conversely, if $\tilde{A}Wx \in L_2(M, \mu)$, then $(\tilde{A} + iI)Wx \in L_2(M, \mu)$, which means that there is a $y \in H$ such that $Wy = (\tilde{A} + iI)Wx$. Therefore

$$\big((A + iI)^{-1}\big)^{\tilde{}}(\tilde{A} + iI)Wx = Wx$$

and hence

$$x = (A + iI)^{-1}y \in D(A).$$

QED

Proposition 10.9.3 If $h \in W(D(A))$ then $\tilde{A}h = WAW^{-1}h$.

Proof. Let $x = W^{-1}h$ which we know belongs to $D(A)$ so we may write $x = (A + iI)^{-1}y$ for some $y \in H$, and hence

$$Ax = y - ix \quad \text{and} \quad Wy = \big[\big((A + iI)^{-1}\big)^{\tilde{}}\,\big]^{-1} h.$$

So

$$
\begin{aligned}
WAx &= Wy - iWx \\
&= \big[\big((A + iI)^{-1}\big)^{\tilde{}}\,\big]^{-1} h - ih \\
&= \tilde{A}h \quad \text{QED}
\end{aligned}
$$

The function $\tilde{A}$ must be real valued almost everywhere since if its imaginary part were positive (or negative) on a set U of positive measure, then $(\tilde{A}\mathbf{1}_U, \mathbf{1}_U)_2$ would have non-zero imaginary part contradicting the fact that multiplication by $\tilde{A}$ is a self adjoint operator, being unitarily equivalent to the self adjoint operator A.

Putting all this together we get

Theorem 10.9.2 *Let A be a self adjoint operator on a separable Hilbert space H. Then there exists a finite measure space (M, μ), a unitary isomorphism $W : H \rightarrow L_2(M, \mu)$ and a real valued measurable function $\tilde{A}$ which is finite almost everywhere such that $x \in D(A)$ if and only if $\tilde{A}Wx \in L_2(M, \mu)$ and if $h \in W(D(A))$ then $\tilde{A}h = WAW^{-1}h$.*

10.9.4 The functional calculus.

Let f be any bounded Borel function defined on $\mathbf{R}$. Then

$$f \circ \tilde{A}$$

is a bounded function defined on M. Multiplication by this function is a bounded operator on $L_2(M, \mu)$ and hence corresponds to a bounded self-adjoint operator on H. With a slight abuse of language we might denote this operator by $O(f \circ \tilde{A})$. However we will use the more suggestive notation

$$f(A).$$

The map

$$f \mapsto f(A)$$

- is an algebraic homomorphism,

- $\overline{f}(A) = f(A)^*$,

- $\|f(A)\| \leq \|f\|_\infty$ where the norm on the left is the uniform operator norm and the norm on the right is the sup norm on $\mathbf{R}$

- if $Ax = \lambda x$ then $f(A)x = f(\lambda)x$,

- if $f \geq 0$ then $f(A) \geq 0$ in the operator sense,

- if $f_n \rightarrow f$ pointwise and if $\|f_n\|_\infty$ is bounded, then $f_n(A) \rightarrow f(A)$ strongly, and

- if f_n is a sequence of Borel functions on the line such that $|f_n(\lambda)| \leq |\lambda|$ for all n and for all $\lambda \in \mathbf{R}$, and if $f_n(\lambda) \rightarrow \lambda$ for each fixed $\lambda \in \mathbf{R}$ then for each $x \in D(A)$

$$f_n(A)x \rightarrow Ax.$$

All of the above statements are obvious except perhaps for the last two which follow from the dominated convergence theorem. It is also clear from the preceding discussion that the map $f \mapsto f(A)$ is uniquely determined by the above properties.

Multiplication by the function $e^{it\tilde{A}}$ is a unitary operator on $L_2(M, \mu)$ and

$$e^{is\tilde{A}}e^{it\tilde{A}} = e^{i(s+t)\tilde{A}}.$$

Hence from the above we conclude

Theorem 10.9.3 [Half of Stone's theorem.] *For an self adjoint operator A the operator e^{itA} given by the functional calculus as above is a unitary operator and*

$$t \mapsto e^{itA}$$

is a one parameter group of unitary transformations.

The full Stone's theorem asserts that any unitary one parameter groups is of this form. We will discuss this later.

10.9.5   Resolutions of the identity.

For each measurable subset X of the real line we can consider its indicator function $\mathbf{1}_X$ and hence $\mathbf{1}_X(A)$ which we shall denote by $P(X)$. In other words

$$P(X) := \mathbf{1}_X(A).$$

It follows from the above that

$$
\begin{aligned}
P(X)^* &= P(X) \\
P(X)P(Y) &= P(X \cap Y) \\
P(X \cup Y) &= P(X) + P(Y) \text{ if } X \cap Y = 0 \\
P(X) &= s - \lim \sum_1^N P(X_i) \text{ if } X_i \cap X_j = \emptyset \text{ if } i \neq j \text{ and } X = \bigcup X_i \\
P(\emptyset) &= 0 \\
P(\mathbf{R}) &= I.
\end{aligned}
$$

For each $x, y \in H$ we have the complex valued measure

$$P_{x,y}(X) = (P(X)x, y)$$

and for any bounded Borel function f we have

$$(f(A)x, y) = \int_{\mathbf{R}} f(\lambda)dP_{x,y}.$$

If g is an unbounded (complex valued) Borel function on $\mathbf{R}$ we define $D(g(A))$ to consist of those $x \in H$ for which

$$\int_{\mathbf{R}} |g(\lambda)|^2 dP_{x,x} < \infty.$$

The set of such x is dense in H and we define $g(A)$ on $D(A)$ by

$$(g(A)x, y) = \int_{\mathbf{R}} g(\lambda) dP_{x,y}$$

for $x, y \in D(g(A))$ (and the Riesz representation theorem). This is written symbolically as

$$g(A) = \int_{\mathbf{R}} g(\lambda) dP.$$

In the special case $g(\lambda) = \lambda$ we write

$$A = \int_{\mathbf{R}} \lambda dP.$$

this is the spectral theorem for self adjoint operators.

In the older literature one often sees the notation

$$E_\lambda := P(-\infty, \lambda).$$

A translation of the properties of P into properties of E is

$$
\begin{align}
E_\lambda^2 &= E_\lambda & (10.10)\\
E_\lambda^* &= E_\lambda & (10.11)\\
\lambda < \mu &\Rightarrow E_\lambda E_\mu = E_\lambda & (10.12)\\
\lambda_{n \to} -\infty &\Rightarrow E_{\lambda_n} \to 0 \text{ strongly} & (10.13)\\
\lambda_n \to +\infty &\Rightarrow E_{\lambda_n} \to I \text{ strongly} & (10.14)\\
\lambda_n \nearrow \lambda &\Rightarrow E_n \to E_\lambda \text{ strongly.} & (10.15)
\end{align}
$$

One then writes the spectral theorem as

$$A = \int_{-\infty}^{\infty} \lambda dE_\lambda. \qquad (10.16)$$

We shall now give an alternative proof of this formula which does not depend on either the Gelfand representation theorem or any of the limit theorems of Lebesgue integration. Instead, it depends on the Riesz-Dunford extension of the Cauchy theory of integration of holomorphic functions along curves to operator valued holomorphic functions.

10.10   The Riesz-Dunford calculus.

Suppose that we have a continuous map $z \mapsto S_z$ defined on some open set of complex numbers, where S_z is a bounded operator on some fixed Banach space and by continuity, we mean continuity relative to the uniform metric on operators. If C is a continuous piecewise differentiable (or more generally any rectifiable) curve lying in this open set, and if $t \mapsto z(t)$ is a piecewise smooth (or rectifiable) parametrization of this curve, then the map $t \mapsto S_{z(t)}$ is continuous.

For any partition $0 = t_0 \leq t_1 \leq \cdots \leq t_n = 1$ of the unit interval we can form the Cauchy approximating sum

$$\sum_{i=1}^{n} S_{z(t_i}(z(t_i) - z(t_{i-1})),$$

and the usual proof of the existence of the Riemann integral shows that this tends to a limit as the mesh becomes more and more refined and the mesh distance tends to zero. The limit is denoted by

$$\int_C S_z dz$$

and this notation is justifies because the change of variables formula for an ordinary integral shows that this value does not depend on the parametrization, but only on the orientation of the curve C.

We are going to apply this to $S_z = R_z$, the resolvent of an operator, and the main equations we shall use are the resolvent equation (10.7) and the power series for the resolvent (10.8) which we repeat here:

$$R_z - R_w = (w - z)R_z R_w$$

and

$$R_w = R_z(I + (z - w)R_z + (z - w)^2 R_z^2 + \cdots).$$

We proved that the resolvent of a self-adjoint operator exists for all non-real values of z.

But a lot of the theory goes over for the resolvent

$$R_z = R(z, T) = (zI - T)^{-1}$$

where T is an arbitrary operator on a Banach space, so long as we restrict ourselves to the resolvent set, i.e. the set where the resolvent exists as a bounded operator. So, following Lorch *Spectral Theory* we first develop some facts about integrating the resolvent in the more general Banach space setting (where our principal application will be to the case where T is a bounded operator).

For example, suppose that C is a simple closed curve contained in the disk of convergence about z of (10.8) i.e. of the above power series for R_w. Then we can integrate the series term by term. But

$$\int_C (z - w)^n dw = 0$$

for all $n \neq -1$ so

$$\int_C R_w dw = 0.$$

By the usual method of breaking any any deformation up into a succession of small deformations and then breaking any small deformation up into a sequence of small "rectangles" we conclude

Theorem 10.10.1 *If two curves C_0 and C_1 lie in the resolvent set and are homotopic by a family C_t of curves lying entirely in the resolvent set then*

$$\int_{C_0} R_z dz = \int_{C_1} R_z dz.$$

Here are some immediate consequences of this elementary result.

Suppose that T is a bounded operator and $|z| > \|T\|$. Then

$$(zI - T)^{-1} = z^{-1}(I - z^{-1}T)^{-1} = z^{-1}(I + z^{-1}T + z^{-2}T^2 + \cdots)$$

exists because the series in parentheses converges in the uniform metric. In other words, all points in the complex plane outside the disk of radius $\|T\|$ lie in the resolvent set of T. From this it follows that the spectrum of any bounded operator can not be empty (if the Banach space is not $\{0\}$). (Recall the the spectrum is the complement of the resolvent set.) Indeed, if the resolvent set were the whole plane, then the circle of radius zero about the origin would be homotopic to a circle of radius $> \|T\|$ via a homotopy lying entirely in the resolvent set. Integrating R_z around the circle of radius zero gives 0. We can integrate around a large circle using the above power series. In performing this integration, all terms vanish except the first which give $2\pi i I$ by the usual Cauchy integral (or by direct computation). Thus $2\pi I = 0$ which is impossible in a non-zero vector space.

Here is another very important (and easy) consequence of the preceding theorem:

Theorem 10.10.2 *Let C be a simple closed rectifiable curve lying entirely in the resolvent set of T. Then*

$$P := \frac{1}{2\pi i} \int_C R_z dz \tag{10.17}$$

is a projection which commutes with T, i.e.

$$P^2 = P \quad and \quad PT = TP.$$

Proof.   Choose a simple closed curve disjoint from C but sufficiently close to C so as to be homotopic to C via a homotopy lying in the resolvent set. Thus

$$P = \frac{1}{2\pi i} \int_{C'} R_w dw$$

and so

$$(2\pi i)^2 P^2 = \int_C R_z dz \int_{C'} R_w dw = \int_C \int_{C'} (R_w - R_z)(z - w)^{-1} dw dz$$

where we have used the resolvent equation (10.7). We write this last expression as a sum of two terms,

$$\int_{C'} R_w \int_C \frac{1}{z - w} dz dw - \int_C R_z \int_{C'} \frac{1}{z - w} dw dz.$$

Suppose that we choose C' to lie entirely inside C. Then the first expression above is just $(2\pi i) \int_{C'} R_w dw$ while the second expression vanishes, all by the elementary Cauchy integral of $1/(z - w)$. Thus we get

$$(2\pi i)^2 P^2 = (2\pi i)^2 P$$

or $P^2 = P$. This proves that P is a projection. It commutes with T because it is an integral whose integrand R_z commutes with T for all z. QED

The same argument proves

Theorem 10.10.3 *Let C and C' be simple closed curves each lying in the resolvent set, and let P and P' be the corresponding projections given by (10.17). Then $PP' = 0$ if the curves lie exterior to one another while $PP' = P'$ if C' is interior to C.*

Let us write
$$B' := PB, \quad B'' = (I - P)B$$

for the images of the projections P and $I - P$ where P is given by (10.17). Each of these spaces is invariant under T and hence under R_z because $PT = TP$ and hence $PR_z = R_z P$.

For any transformation C commuting with P let us write

$$C' := PC = CP = PCP \quad \text{and } C'' = (I - P)C = C(I - P) = (I - P)C(I - P)$$

so that C' and C'' are the restrictions of C to B' and B'' respectively.

For example, we may consider $R'_z = PR_z = R_z P$. For $x' \in B'$ we have $R'_z(zI - T')x' = R_z P(zI - TP)x' = R_z(zI - T)Px' = x'$. In other words R'_z is the resolvent of T' (on B') and similarly for R''_z. So if z is in the resolvent set for T it is in the resolvent set for T' and T''.

Conversely, suppose that z is in the resolvent set for both T' and T''. Then there exists an inverse A_1 for $zI' - T'$ on B' and an inverse A_2 for $zI'' - T''$ on B'' and so $A_1 \oplus A_2$ is the inverse of $zI - T$ on $B = B' \oplus B''$.

So a point belongs to the resolvent set of T if and only if it belongs to the resolvent set of T' and of T''. Since the spectrum is the complement of the resolvent set, we can say that a point belongs to the spectrum of T if and only if it belongs either to the spectrum of T' or of T'':

$$\mathrm{Spec}(T) = \mathrm{Spec}(T') \cup \mathrm{Spec}(T'').$$

We now show that this decomposition is in fact the decomposition of $\mathrm{Spec}(T)$ into those points which lie inside C and outside C.

So we must show that if z lies exterior to C then if lies in the resolvent set of T'. This will certainly be true if we can find a transformation A on B which commutes with T and such that

$$A(zI - T) = P \tag{10.18}$$

for then A' will be the resolvent at z of T'. Now

$$(zI - T)R_w = (wI - T)R_w + (z - w)R_w = I + (z - w)R_w$$

so

$$(zI - T) \cdot \frac{1}{2\pi i} \int_C R_w \cdot \frac{1}{z - w} dw =$$

$$= \frac{1}{2\pi i} \int_C \frac{1}{z - w} dw \cdot I + \frac{1}{2\pi i} \int_C R_w dw = 0 + P = P.$$

We have thus proved

Theorem 10.10.4 *Let T be a bounded linear transformation on a Banach space and C a simple closed curve lying in its resolvent set. Let P be the projection given by (10.17) and*

$$B = B' \oplus B'', \qquad T = T' \oplus T''$$

the corresponding decomposition of B and of T. Then $\mathrm{Spec}(T')$ consists of those points of $\mathrm{Spec}(T)$ which lie inside C and $\mathrm{Spec}(T'')$ consists of those points of $\mathrm{Spec}(T)$ which lie exterior to C.

We now begin to have a better understanding of Stone's formula: Suppose A is a self-adjoint operator. We know that its spectrum lies on the real axis. If we draw a rectangle whose upper and lower sides are parallel to the axis, and if its vertical sides do not intersect $\mathrm{Spec}(A)$, we would get a projection onto a subspace M of our Hilbert space which is invariant under A, and such that the spectrum of A when restricted to M lies in the interval cut out on the real axis by our rectangle. The problem is how to make sense of this procedure when the vertical edges of the rectangle might cut through the spectrum, in which case the integral (10.17) might not even be defined. This is resolved by the method of Lorch (the exposition is taken from his book) which we explain in the next section.

10.11 Lorch's proof of the spectral theorem.

10.11.1 Positive operators.

Recall that if A is a bounded self-adjoint operator on a Hilbert space H then we write $A \geq 0$ if $(Ax, x) \geq 0$ for all $x \in H$ and (by a slight abuse of language) call

such an operator positive. Clearly the sum of two positive operators is positive as is the multiple of a positive operator by a non-negative number. Also we write $A_1 \geq A_2$ for two self adjoint operators if $A_1 - A_2$ is positive.

Proposition 10.11.1 *If A is a bounded self-adjoint operator and $A \geq I$ then A^{-1} exists and*

$$\|A^{-1}\| \leq 1.$$

Proof.   We have

$$\|Ax\| \, \|x\| \geq (Ax, x) \geq (x, x) = \|x\|^2$$

so

$$\|Ax\| \geq \|x\| \quad \forall \, x \in H.$$

So A is injective, and hence A^{-1} is defined on im A and is bounded by 1 there. We must show that this image is all of H.

If y is orthogonal to im A we have

$$(x, Ay) = (Ax, y) = 0 \ \ \forall \, x \in H$$

so $Ay = 0$ so $(y, y) \leq (Ay, y) = 0$ and hence $y = 0$. Thus im A is dense in H.

Suppose that $Ax_n \to z$. Then the x_n form a Cauchy sequence by the estimate above on $\|A^{-1}\|$ and so $x_n \to x$ and the continuity of A implies that $Ax = z$. QED

Suppose that $A \geq 0$. Then for any $\lambda > 0$ we have $A + \lambda I \geq \lambda I$, and by the proposition $(A + \lambda I)^{-1}$ exists, i.e. $-\lambda$ belongs to the resolvent set of A. So we have proved.

Proposition 10.11.2 *If $A \geq 0$ then* $\mathrm{Spec}(A) \subset [0, \infty)$.

Theorem 10.11.1 *If A is a self-adjoint transformation then*

$$\|A\| \leq 1   \Leftrightarrow   -I \leq A \leq I. \tag{10.19}$$

Proof.   Suppose $\|A\| \leq 1$. Then using Cauchy-Schwarz and then the definition of $\|A\|$ we get

$$([I - A]x, x) = (x, x) - (Ax, x) \geq \|x\|^2 - \|Ax\| \, \|x\| \geq \|x\|^2 - \|A\|\|x\|^2 \geq 0$$

so $(I - A) \geq 0$ and applied to $-A$ gives $I + A \geq 0$ or $-I \leq A \leq I$.

Conversely, suppose that $-I \leq A \leq I$. Since $I - A \geq 0$ we know that $\mathrm{Spec}(A) \subset (-\infty, 1]$ and since $I + A \geq 0$ we have $\mathrm{Spec}(A) \subset (-1, \infty]$. So

$$\mathrm{Spec}(A) \subset [-1, 1]$$

so that the spectral radius of A is ≤ 1. But for self adjoint operators we have $\|A^2\| = \|A\|^2$ and hence the formula for the spectral radius gives $\|A\| \leq 1$. QED

An immediate corollary of the theorem is the following: Suppose that μ is a real number. Then $\|A - \mu I\| \leq \epsilon$ is equivalent to $(\mu - \epsilon)I \leq A \leq (\mu_+ \epsilon)I$. So one way of interpreting the spectral theorem

$$A = \int_\infty \lambda dE_\lambda$$

is to say that for any doubly infinite sequence

$$\cdot < \lambda_{-2} < \lambda_{-1} < \lambda_0 < \lambda_1 < \lambda_2 < \cdots$$

with $\lambda_{-n} \to -\infty$ and $\lambda_n \to \infty$ there is a corresponding Hilbert space direct sum decomposition

$$H = \bigoplus H_i$$

invariant under A and such that the restriction of A to H_i satisfies

$$\lambda_i I \leq A_{|H_i} \leq \lambda_{i+1} I.$$

If $\mu_i := \frac{1}{2}(\lambda_i + \lambda_{i+1})$ then another way of writing the preceding inequality is

$$\|A_{|H_i} - \mu_i I\| \leq \frac{1}{2}(\lambda_{i+1} - \lambda_i).$$

10.11.2   The point spectrum.

We now let A denote an arbitrary (not necessarily bounded) self adjoint transformation. We say that λ belongs to the **point spectrum** of A if there exists an $x \in D(A)$ such that $x \neq 0$ and $Ax = \lambda x$. In other words if λ is an eigenvalue of A. Notice that eigenvectors corresponding to distinct eigenvalues are orthogonal: if $Ax = \lambda x$ and $Ay = \mu y$ then

$$\lambda(x, y) = (\lambda x, y) = (Ax, y) = (x, Ay) = (x, \mu y) = \mu(x, y)$$

implying that $(x, y) = 0$ if $\lambda \neq \mu$.

Also, the fact that a self-adjoint operator is closed implies that the space of eigenvectors corresponding to a fixed eigenvalue is a closed subspace of H. We let N_λ denote the space of eigenvectors corresponding to an eigenvalue λ.

We say that A has **pure point spectrum** if its eigenvectors span H, in other words if

$$H = \bigoplus N_{\lambda_i}$$

where the λ_i range over the set of eigenvalues of A. Suppose that this is the case. Then let

$$M_\lambda := \bigoplus_{\mu < \lambda} N_\mu$$

where this denotes the Hilbert space direct sum, i.e. the closure of the algebraic direct sum. Let E_λ denote projection onto M_λ. Then it is immediate that the E_λ satisfy (10.10)-(10.15) and that (10.16) holds with the interpretation given in the preceding section. We thus have a proof of the spectral theorem for operators with pure point spectrum.

10.11.3   Partition into pure types.

Now consider a general self-adjoint operator A, and let

$$H_1 := \bigoplus N_\lambda$$

(Hilbert space direct sum) and set

$$H_2 := H_1^\perp.$$

The space H_1 and hence the space H_2 are invariant under A in the sense that A maps $D(A) \cap H_1$ to H_1 and similarly for H_2.

We let P denote orthogonal projection onto H_1 so $I - P$ is orthogonal projection onto H_2. We claim that

$$P[D(A)] = D(A) \cap H_1 \quad \text{and} \quad (I - P)[D(A)] = D(A) \cap H_2. \qquad (10.20)$$

Suppose that $x \in D(A)$. We must show that $Px \in D(A)$ for then $x = Px + (I - P)x$ is a decomposition of every element of $D(A)$ into a sum of elements of $D(A) \cap H_1$ and $D(A) \cap H_2$.

By definition, we can find an orthonormal basis of H_1 consisting of eigenvectors u_i of A, and then

$$Px = \sum a_i u_i \quad a_i := (x, u_i).$$

The sum on the right is (in general) infinite. Let y denote any finite partial sum. Since eigenvectors belong to $D(A)$ we know that $y \in D(A)$. We have

$$(A[x - y], Ay) - ([x - y], A^2 y) = 0$$

since $x - y$ is orthogonal to all the eigenvectors occurring in the expression for y. We thus have

$$\|Ax\|^2 = \|A(x - y)\|^2 + \|Ay\|^2$$

From this we see (as we let the number of terms in y increase) that both y converges to Px and the Ay converge. Hence $Px \in D(A)$ proving (10.20).

Let A_1 denote the operator A restricted to $P[D(A)] = D(A) \cap H_1$ with similar notation for A_2. We claim that A_1 is self adjoint (as is A_2). Clearly $D(A_1) := P(D(A))$ is dense in H_1, for if there were a vector $y \in H_1$ orthogonal to $D(A_1)$ it would be orthogonal to $D(A)$ in H which is impossible. Similarly $D(A_2) := D(A) \cap H_2$ is dense in H_2.

Now suppose that y_1 and z_1 are elements of H_1 such that

$$(A_1 x_1, y_1) = (x_1, z_1) \ \ \forall \, x_1 \in D(A_1).$$

Since $A_1 x_1 = A x_1$ and $x_1 = x - x_2$ for some $x \in D(A)$, and since y_1 and z_1 are orthogonal to x_2, we can write the above equation as

$$(A x, y_1) = (x, z_1) \ \ \forall \, x \in D(A)$$

which implies that $y_1 \in D(A) \cap H_1 = D(A_1)$ and $A_1 y_1 = A y_1 = z_1$.

In other words, A_1 is self-adjoint. Similarly, so is A_2. We have thus proved

Theorem 10.11.2 *Let A be a self-adjoint transformation on a Hilbert space H. Then*

$$H = H_1 \oplus H_2$$

with self-adjoint transformations A_1 on H_1 having pure point spectrum and A_2 on H_2 having no point spectrum such that

$$D(A) = D(A_1) \oplus D(A_2)$$

and

$$A = A_1 \oplus A_2.$$

We have proved the spectral theorem for a self adjoint operator with pure point spectrum. Our proof of the full spectral theorem will be complete once we prove it for operators with no point spectrum.

10.11.4 Completion of the proof.

In this subsection we will assume that A is a self-adjoint operator with no point spectrum, i.e. no eigenvalues.

Let $\lambda < \mu$ be real numbers and let C be a closed piecewise smooth curve in the complex plane which is symmetrical about the real axis and cuts the real axis at non-zero angle at the two points λ and μ (only). Let $m > 0$ and $n > 0$ be positive integers, and let

$$K_{\lambda\mu}(m,n) := \frac{1}{2\pi i} \int_C (z-\lambda)^m (z-\mu)^n R_z dz. \tag{10.21}$$

In fact, we would like to be able to consider the above integral when $m = n = 0$, in which case it should give us a projection onto a subspace where $\lambda I \leq A \leq \mu I$. But unfortunately if λ or μ belong to $\mathrm{Spec}(A)$ the above integral need not converge with $m = n = 0$. However we do know that $\|R_z\| \leq (|\mathrm{im}\, z|)^{-1}$ so that the blow up in the integrand at λ and μ is killed by $(z-\lambda)^m$ and $(\mu-z)^n$ since the curve makes non-zero angle with the real axis. Since the curve is symmetric about the real axis, the (bounded) operator $K_{\lambda\mu}(m,n)$ is self-adjoint. Furthermore, modifying the curve C to a curve C' lying inside C, again intersecting the real axis only at the points λ and μ and having these intersections at non-zero angles does not change the value: $K_{\lambda\mu}(m,n)$.

We will now prove a succession of facts about $K_{\lambda\mu}(m,n)$:

$$K_{\lambda\mu}(m,n) \cdot K_{\lambda\mu}(m',n') = K_{\lambda\mu}(m+m',n+n'). \tag{10.22}$$

Proof. Calculate the product using a curve C' for $K_{\lambda\mu}(m',n')$ as indicated above. Then use the functional equation for the resolvent and Cauchy's integral formula exactly as in the proof of Theorem 10.10.2: $(2\pi i)^2 K_{\lambda\mu}(m,n) \cdot K_{\lambda\mu}(m',n') =$

$$\int_C \int_{C'} (z-\lambda)^m (\mu-z)^n (w-\lambda)^{m'} (\mu-w)^{n'} \frac{1}{z-w} [R_w - R_z] dz dw$$

which we write as a sum of two integrals, the first giving $(2\pi i)^2 K_{\lambda\mu}(m+m', n+n')$ and the second giving zero. QED

A similar argument (similar to the proof of Theorem 10.10.3) shows that

$$K_{\lambda\mu}(m, n) \cdot K_{\lambda'\mu'}(m', n') = 0 \ \ \text{if } (\lambda, \mu) \cap (\lambda', \mu') = \emptyset. \tag{10.23}$$

Proposition 10.11.3 *There exists a bounded self-adjoint operator* $L_{\lambda\mu}(m, n)$ *such that*

$$L_{\lambda\mu}(m, n)^2 = K_{\lambda\mu}(m, n).$$

Proof.   The function $z \mapsto (z - \lambda)^{m/2}(\mu - z)^{n/2}$ is defined and holomorphic on the complex plane with the closed intervals $(-\infty, \lambda]$ and $[\mu, \infty)$ removed. The integral

$$L_{\lambda\mu}(m, n) = \frac{1}{2\pi i} \int_C (z - \lambda)^{m/2}(\mu - z)^{n/2} R_z dz$$

is well defined since, if $m = 1$ or $n = 1$ the singularity is of the form $|\text{im } z|^{-\frac{1}{2}}$ at worst which is integrable. Then the proof of (10.22) applies to prove the proposition. QED

For each complex z we know that $R_z x \in D(A)$. Hence

$$(A - \lambda I)R_z x = (A - zI)R_z x + (z - \lambda)R_z x = x + (z - \lambda)R_z x.$$

By writing the integral defining $K_{\lambda\mu}(m, n)$ as a limit of approximating sums, we see that $(A - \lambda I)K_{\lambda\mu}(m, n)$ is defined and that it is given by the sum of two integrals, the first of which vanishes and the second gives $K_{\lambda\mu}(m + 1, n)$.

We have thus shown that $K_{\lambda\mu}(m, n)$ maps H into $D(A)$ and

$$(A - \lambda I)K_{\lambda\mu}(m, n) = K_{\lambda\mu}(m + 1, n). \tag{10.24}$$

Similarly

$$(\mu I - A)K_{\lambda\mu}(m, n) = K_{\lambda\mu}(m, n + 1). \tag{10.25}$$

We also have

$$\lambda(x, x) \le (Ax, x) \le \mu(x, x) \ \ \text{for } x \in \ \text{im } K_{\lambda\mu}(m, n). \tag{10.26}$$

Proof.   We have

$$([A - \lambda I]K_{\lambda\mu}(m, n)y, K_{\lambda\mu}(m, n)y)$$

$$\begin{aligned}
&= & (K_{\lambda\mu}(m + 1, n)y, K_{\lambda\mu}(m, n)y) \\
&= & (K_{\lambda\mu}(m, n)K_{\lambda\mu}(m + 1, n)y, y) \\
&= & (K_{\lambda\mu}(2m + 1, 2n)y, y) \\
&= & (L_{\lambda\mu}(2m + 1, 2n)^2 y, y) \\
&= & (L_{\lambda\mu}(2m + 1, 2n)y, L_{\lambda\mu}(2m + 1, 2n)y) \ge 0. \\
&\Rightarrow & A \ge \lambda I \ \ \text{on im } K_{\lambda\mu}(m, n).
\end{aligned}$$

A similar argument shows that $A \leq \mu I$ there. QED

Thus if we define $M_{\lambda\mu}(m,n)$ to be the closure of im $K_{\lambda\mu}(m,n)$ we see that A is bounded when restricted to $M_{\lambda\mu}(m,n)$ and

$$\lambda I \leq A \leq \mu I$$

there.

We let $N_{\lambda\mu}(m,n)$ denote the kernel of $K_{\lambda\mu}(m,n)$ so that $M_{\lambda\mu}(m,n)$ and $N_{\lambda\mu}(m,n)$ are the orthogonal complements of one another.

So far we have not made use of the assumption that A has no point spectrum. Here is where we will use this assumption: Since

$$(A - \lambda I)K_{\lambda\mu}(m,n) = K_{\lambda\mu}(m+1,n)$$

we see that if $K_{\lambda\mu}(m+1,n)x = 0$ we must have $(A-\lambda I)K_{\lambda\mu}(m,n)x = 0$ which, by our assumption implies that $K_{\lambda\mu}(m,n)x = 0$. In other words,

Proposition 10.11.4   *The space $N_{\lambda\mu}(m,n)$, and hence its orthogonal complement $M_{\lambda\mu}(m,n)$ is independent of m and n.*

We will denote the common space $M_{\lambda\mu}(m,n)$ by $M_{\lambda\mu}$. We have proved that A is a bounded operator when restricted to $M_{\lambda\mu}$ and satisfies

$$\lambda I \leq A \leq \mu I \quad \text{on } M_{\lambda\mu}$$

there.

We now claim that

$$\text{If } \lambda < \nu < \mu \quad \text{then} \quad M_{\lambda\mu} = M_{\lambda\nu} \oplus M_{\nu\mu}. \tag{10.27}$$

Proof.   Let $C_{\lambda\mu}$ denote the rectangle of height one parallel to the real axis and cutting the real axis at the points λ and μ. Use similar notation to define the rectangles $C_{\lambda\nu}$ and $C_{\nu\mu}$. Consider the integrand

$$S_z := (z - \lambda)(z - \mu)(z - \nu)R_z$$

and let

$$T_{\lambda\mu} := \frac{1}{2\pi i} \int_{C_{\lambda\mu}} S_z dz$$

with similar notation for the integrals over the other two rectangles of the same integrand. Then clearly

$$T_{\lambda\mu} = T_{\lambda\nu} + T_{\nu\mu} \quad \text{and } T_{\lambda\nu} \cdot T_{\nu\mu} = 0. \tag{10.28}$$

Also, writing $zI - A = (z - \nu)I + (\nu I - A)$ we see that

$$(\nu I - A)K_{\lambda\nu}(1,1) = T_{\lambda\mu}$$

Since A has no point spectrum, the closure of the image of $T_{\lambda\mu}$ is the same as the closure of the image of $K_{\lambda\mu}(1,1)$, namely $M_{\lambda\mu}$. The proposition now follows from (10.28).

If we now have a doubly infinite sequence as in our reformulation of the spectral theorem, and we set $M_i := M_{\lambda_i \lambda_{i+1}}$ we have proved the spectral theorem (in the no point spectrum case - and hence in the general case) if we show that

$$\bigoplus M_i = H.$$

In view of (10.27) it is enough to prove that the closure of the limit of M_{-rr} is all of H as $r \to \infty$, or, what amounts to the same thing, if y is perpendicular to all $K_{-rr}(1,1)x$ then y must be zero. Now

$$(K_{-rr}(1,1)x, y) = (x, K_{-rr}(1,1)y)$$

so we must show that if $K_{-rr}y = 0$ for all r then $y = 0$. Now

$$K_{-rr} = \frac{1}{2\pi i} \int_C (z+r)(r-z)R_z dz = -\frac{1}{2\pi i} \int (z^2 - r^2)R_z$$

where we may take C to be the circle of radius r centered at the origin. We also have

$$1 = \frac{1}{2\pi i r^2} \int_C \frac{r^2 - z^2}{z} dz.$$

So

$$y = \frac{1}{2\pi i r^2} \int_C (r^2 - z^2)[z^{-1}I - R_z]dz \cdot y.$$

Now $(zI - A)R_z = I$ so $-AR_z = I - zR_z$ or

$$z^{-1}I - R_z = -z^{-1}AR_z$$

so (pulling the A out from under the integral sign) we can write the above equation as

$$y = Ag_r \quad \text{where } g_r = \frac{1}{2\pi i r^2} \int_C (r^2 - z^2)z^{-1}R_z dz \cdot y.$$

Now on C we have $z = re^{i\theta}$ so $z^2 = r^2 e^{2i\theta} = r^2(\cos 2\theta + i \sin 2\theta)$ and hence

$$z^2 - r^2 = r^2(\cos 2\theta - 1 + i \sin 2\theta) = 2r^2(-\sin^2 \theta + i \sin \theta \cos \theta).$$

Now $\|R_z\| \le |r \sin \theta|^{-1}$ so we see that

$$\|(z^2 - r^2)R_z\| \le 4r.$$

Since $|z^{-1}| = r^{-1}$ on C, we can bound $\|g_r\|$ by

$$\|g_r| \le (2\pi r^2)^{-1} \cdot r^{-1} \cdot 4r \cdot 2\pi r \|y\| = 4r^{-1}\|y\| \to 0$$

as $r \to \infty$. Since $y = Ag_r$ and A is closed (being self-adjoint) we conclude that $y = 0$. This concludes Lorch's proof of the spectral theorem.

10.12 Appendix. The closed graph theorem.

Lurking in the background of our entire discussion is the closed graph theorem which says that if a closed linear transformation from one Banach space to another is everywhere defined, it is in fact bounded. We did not actually use this theorem, but its statement and proof by Banach greatly clarified the notion of a what an unbounded self-adjoint operator is, and explained the Hellinger Toeplitz theorem as I mentioned earlier. So here I will give the standard proof of this theorem (essentially a Baire category style argument) taken from Loomis.

In what follows X and Y will denote Banach spaces,

$$B_n := B_n(X) = \{x \in X; \ \|x\| \le n\}$$

denotes the ball of radius n about the origin in X and

$$U_r = B_r(Y) = \{y \in Y : \ \|y\| \le r\}$$

the ball of radius r about the origin in Y.

Lemma 10.12.1 *Let*

$$T : X \to Y$$

be a bounded (everywhere defined) linear transformation. If $T[B_1] \cap U_r$ is dense in U_r then

$$U_r \subset T[B_1].$$

Proof. The set $T[B_1]$ is closed, so it will be enough to show that

$$U_{r(1-\delta)} \subset T[B_1]$$

for any $\delta > 0$, or, what is the same thing, that

$$U_r \subset \frac{1}{1-\delta} T[B_1] = T[B_{\frac{1}{1-\delta}}].$$

So fix $\delta > 0$. Let $z \in U_r$. Set $y_0 := 0$, and choose $y_1 \in T[B_1] \cap U_r$ such that $\|z - y_1\| < \delta r$. Since $\delta \left(T[B_1] \cap U_r\right)$ is dense in δU_r we can find $y_2 - y_1 \in \delta \left(T[B_1] \cap U_r\right)$ within distance $\delta^2 r$ of $z - y_1$ which implies that $\|y_2 - z\| < \delta^2 r$. Proceeding inductively we find a sequence $\{y_n \in Y$ such that

$$y_{n+1} - y_n \in \delta^n \left(T[B_1] \cap U_r\right)$$

and

$$\|y_{n+1} - z\|, \delta^{n+1} r.$$

We can thus find $x_n \in X$ such that

$$T(x_{n+1} = y_{n+1} - y_n \quad \text{and} \quad \|x_{n+1}\| < \delta^n.$$

If

$$x := \sum_1^\infty x_n$$

then $\|x\| < 1/(1-\delta)$ and $Tx = z$. QED

Lemma 10.12.2 *If $T[B_1]$ is dense in no ball of positive radius in Y, then $T[X]$ contains no ball of positive radius in Y.*

Proof. Under the hypotheses of the lemma, $T[B_n]$ is also dense in no ball of positive radius of Y. So given any ball $U \subset Y$, we can find a (closed) ball U_{r_1,y_1} of radius r_1 about y_1 such that $U_{r_1,y_1} \subset U$ and is disjoint from $T[B_1]$. By induction, we can find a nested sequence of balls $U_{r_n,y_n} \subset U_{r_{n-1}y_{n-1}}$ such that U_{r_n,y_n} is disjoint form $T[B_n]$ and can also arrange that $r_n \to 0$. Choosing a point in each of these balls we get a Cauchy sequence which converges to a point $y \in U$ which lies in none of the $T[B_n]$, i.e. $y \notin T[X]$. So $u \not\subset T[X]$. QED

Theorem 10.12.1 [The bounded inverse theorem.] *If $T : X \to Y$ is bounded and bijective, then T^{-1} is bounded.*

Proof. By Lemma 10.12.2, $T[B_1]$ is dense in some ball U_{r,y_1} and hence

$$T[B_1 + B_1] = T[B_1 - B_1]$$

is dense in a ball of radius r about the origin. Since $B_1 + B_1 \subset B_2$ so $T[B_2] \cap U_r$ is dense in U_r. By Lemma 10.12.1, this implies that

$$T[B_2] \supset U_r$$

i.e. that

$$T^{-1}[U_r] \subset B_2$$

which says that

$$\|T^{-1}\| \leq \frac{2}{r}.$$

QED

Theorem 10.12.2 *If $T : X \to Y$ is defined on all of X and is such that graph(T) is a closed subspace of $X \oplus Y$, then T is bounded.*

Proof. Let $\Gamma \subset X \oplus Y$ denote the graph of T. By assumption, it is a closed subspace of the Banach space $X \oplus Y$ under the norm $\|\{x,y\}\| = \|x\| + \|y\|$. So Γ is a Banach space and the projection

$$\Gamma \to X, \quad \{x,y\} \mapsto x$$

is bijective by the definition of a graph, and has norm ≤ 1. So its inverse is bounded. Similarly the projection onto the second factor is bounded. So the composite map

$$X \to Y \quad x \mapsto \{x,y\} \mapsto y = Tx$$

is bounded. QED

Chapter 11

Stone's theorem

Recall that if A is a self-adjoint operator on a Hilbert space $\mathbf{H}$ we can form the one parameter group of unitary operators

$$U(t) = e^{iAt}$$

by virtue of a functional calculus which allows us to construct $f(A)$ for any bounded Borel function defined on $\mathbf{R}$ (if we use our first proof of the spectral theorem using the Gelfand representation theorem) or for any function holomorphic on $\mathrm{Spec}(A)$ if we use our second proof. In any event, the spectral theorem allows us to write

$$U(t) = \int_{-\infty}^{\infty} e^{it\lambda} dE_\lambda$$

and to verify that

$$U(0) = I, \quad U(s+t) = U(s)U(t)$$

and that U depends continuously on t. We called this assertion the first half of Stone's theorem. The second half (to be stated more precisely below) asserts the converse: that any one parameter group of unitary transformations can be written in either, hence both, of the above forms.

The idea that we will follow hinges on the following elementary computation

$$\int_0^{\infty} e^{(-z+ix)t} dt = \left. \frac{e^{(-z+ix)t}}{-z+ix} \right|_{t=0}^{\infty} = \frac{1}{z-ix} \text{ if } \mathrm{Re}\ z > 0$$

valid for any real number x. If we substitute A for x and write $U(t)$ instead of e^{iAt} this suggests that

$$R(z, iA) = (zI - iA)^{-1} = \int_0^{\infty} e^{-zt} U(t) dt \text{ if } \mathrm{Re}\ z > 0.$$

Since A is self-adjoint, its spectrum is real. So the spectrum of iA is purely imaginary, and hence any z not on the imaginary axis is in the resolvent set of iA. The above formula gives us an expression for the resolvent in terms of $U(t)$

283

for z lying in the right half plane. We can obtain a similar formula for the left half plane.

Our previous studies encourage us to believe that once we have found all these putative resolvents, it should not be so hard to reconstruct A and then the one-parameter group $U(t) = e^{iAt}$.

This program works! But because of some of the subtleties involved in the definition of a self-adjoint operator, we will begin with an important theorem of von-Neumann which we will need, and which will also greatly clarify exactly what it means to be self-adjoint.

A second matter which will lengthen these proceedings is that while we are at it, we will prove a more general version of Stone's theorem valid in an arbitrary Frechet space **F** and for "uniformly bounded semigroups" rather than unitary groups. Stone proved his theorem to meet the needs of quantum mechanics, where a unitary one parameter group corresponds, via *Wigner's theorem* to a one parameter group of symmetries of the logic of quantum mechanics. In more pedestrian terms, unitary one parameter groups arise from solutions of Schrodinger's equation. But many other important equations, for example the heat equations in various settings, require the more general result.

The treatment here will essentially follow that of Yosida, *Functional Analysis* especially Chapter IX, Nelson, *Topics in dynamics I: Flows*, and Reed and Simon *Methods of Mathematical Physics, II. Fourier Analysis, Self-Adjointness.*

11.1   von Neumann's Cayley transform.

The group $Gl(2, \mathbf{C})$ of all invertible complex two by two matrices acts as "fractional linear transformations" on the plane: the matrix

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \text{ sends } \quad z \mapsto \frac{az + b}{cz + d}.$$

Two different matrices M_1 and M_2 give the same fractional linear transformation if and only if $M_1 = \lambda M_2$ for some (non-zero complex) number λ as is clear from the definition. Since

$$\begin{pmatrix} 1 & -i \\ 1 & i \end{pmatrix} \begin{pmatrix} i & i \\ -1 & 1 \end{pmatrix} = 2i \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},$$

the fractional linear transformations corresponding to $\begin{pmatrix} 1 & -i \\ 1 & i \end{pmatrix}$ and $\begin{pmatrix} i & i \\ -1 & 1 \end{pmatrix}$ are inverse to one another.

It is a theorem in the elementary theory of complex variables that fractional linear transformations are the only orientation preserving transformations of the plane which carry circles and lines into circles and lines. Even without this general theory, an immediate computation shows that $\begin{pmatrix} 1 & -i \\ 1 & i \end{pmatrix}$ carries the (extended) real axis onto the unit circle, and hence its inverse carries the unit

circle onto the extended real axis. ("Extended" means with the point ∞ added.) Indeed in the expression

$$z = \frac{x - i}{x + i}$$

when x is real, the numerator is the complex conjugate of the denominator and hence $|z| = 1$. Under this transformation, the cardinal points $0, 1, \infty$ of the extended real axis are mapped as follows:

$$0 \mapsto -1, \quad 1 \mapsto -i, \quad \text{and} \quad \infty \mapsto 1.$$

We might think of (multiplication by) a real number as a self-adjoint transformation on a one dimensional Hilbert space, and (multiplication by) a number of absolute value one as a unitary operator on a one dimensional Hilbert space. This suggests in general that if A is a self adjoint operator, then

$$(A - iI)(A + iI)^{-1}$$

should be unitary. In fact, we can be much more precise. First some definitions:

An operator U, possibly defined only on a subspace of a Hilbert space $\mathbf{H}$ is called **isometric** if

$$\|Ux\| = \|x\|$$

for all x in its domain of definition.

Recall that in order to define the adjoint T^* of an operator T it is necessary that its domain $D(T)$ be dense in $\mathbf{H}$. Otherwise the equation

$$(Tx, y) = (x, T^*y) \,\forall\, x \in D(T)$$

does not determine T^*y. A transformation T (in a Hilbert space $\mathbf{H}$) is called **symmetric** if $D(T)$ is dense in $\mathbf{H}$ so that T^* is defined and

$$D(T) \subset D(T^*) \text{ and } Tx = T^*x \quad \forall\, x \in D(T).$$

Another way of saying the same thing is T is symmetric if $D(T)$ is dense and

$$(Tx, y) = (x, Ty) \quad \forall\, x, y \in D(T).$$

A self-adjoint transformation is symmetric since $D(T) = D(T^*)$ is one of the requirements of being self-adjoint. Exactly how and why a symmetric operator can fail to be self-adjoint will be clarified in the ensuing discussion. All of the results of this section are due to von Neumann.

Theorem 11.1.1 *Let T be a closed symmetric operator. Then $(T + iI)x = 0$ implies that $x = 0$ for any $x \in D(T)$ so $(T + iI)^{-1}$ exists as an operator on its domain*

$$D\left[(T + iI)^{-1}\right] = \operatorname{im}(T + iI).$$

This operator is bounded on its domain and the operator

$$U_T := (T - iI)(T + iI)^{-1} \quad \text{with } D(U_T) = D\left[(T + iI)^{-1}\right] = \operatorname{im}(T + iI)$$

is isometric and closed. The operator $(I - U_T)^{-1}$ exists and

$$T = i(U_T + I)(U_T - I)^{-1}.$$

In particular, $D(T) = \operatorname{im}(I - U_T)$ is dense in H.

Conversely, if U is a closed isometric operator such that $\operatorname{im}(I - U)$ is dense in $\mathbf{H}$ then $T = i(U + I)(I - U)^{-1}$ is a symmetric operator with $U = U_T$.

Proof.   For any $x \in D(T)$ we have

$$([T \pm iI]x, [T \pm iI]x) = (Tx, Tx) \pm (Tx, ix) \pm (ix, Tx) + (x, x).$$

The middle terms cancel because T is symmetric. Hence

$$\|[T \pm iI]x\|^2 = \|Tx\|^2 + \|x\|^2. \tag{11.1}$$

Taking the plus sign shows that $(T + iI)x = 0 \ \Rightarrow \ x = 0$ and also shows that $\|[T + iI]x\| \geq \|x\|$ so

$$\|[T + iI]^{-1}y\| \leq \|y\| \quad \text{for} \ \ y \in [T + iI](D(T)).$$

If we write $x = [T + iI]^{-1}y$ then (11.1) shows that

$$\|U_T y\|^2 = \|Tx\|^2 + \|x\|^2 = \|y\|^2$$

so U_T is an isometry with domain consisting of all $y = (T + iI)x$, i.e. with domain $D([T + iI]^{-1}) = \operatorname{im}[T + iI]$.

We now show that U_T is closed. So we must show that if $y_n \to y$ and $z_n \to z$ where $z_n = U_T y_n$ then $y \in D(U_T)$ and $U_T y = z$. The y_n form a Cauchy sequence and $y_n = [T + iI]x_n$ since $y_n \in \operatorname{im}(T + iI)$. From (11.1) we see that the x_n and the Tx_n form a Cauchy sequence, so $x_n \to x$ and $Tx_n \to w$ which implies that $x \in D(T)$ and $Tx = w$ since T is assumed to be closed. But then $(T + iI)x = w + ix = y$ so $y \in D(U_T)$ and $w - ix = z = U_T y$. So we have shown that U_T is closed.

Subtract and add the equations

$$
\begin{aligned}
y &= (T + iI)x \\
U_T y &= (T - iI)x \quad \text{to get} \\
\frac{1}{2}(I - U_T)y &= ix \quad \text{and} \\
\frac{1}{2}(I + U_T)y &= Tx.
\end{aligned}
$$

The third equation shows that

$$(I - U_T)y = 0 \ \Rightarrow \ x = 0 \ \Rightarrow Tx = 0 \ \Rightarrow (I + U_T)y = 0$$

by the fourth equation. So

$$y = \frac{1}{2}([I - U_T]y + [I + U_T]y) = 0.$$

Thus $(I - U_T)^{-1}$ exists, and $y = (I - U_T)^{-1}(2ix)$ from the third of the four equations above, and the last equation gives

$$Tx = \frac{1}{2}(I + U_T)y = \frac{1}{2}(I + U_T)(I - U_T)^{-1}2ix$$

or

$$T = i(I + U_T)(I - U_T)^{-1}$$

as required. Furthermore, every $x \in D(T)$ is in $\mathrm{im}(I - U_T)$. This completes the proof of the first half of the theorem.

Now suppose we start with an isometry U and suppose that $(I - U)y = 0$ for some $y \in D(U)$. Let $z \in \mathrm{im}(I - U)$ so $z = w - Uw$ for some w. We have

$$(y, z) = (y, w) - (y, Uw) = (Uy, Uw) - (y, Uw) = (Uy - y, Uw) = 0.$$

Since we are assuming that $\mathrm{im}(I - U)$ is dense in $\mathbf{H}$, the condition $(y, z) = 0 \; \forall \, z \in \mathrm{im}(I - U)$ implies that $y = 0$. Thus $(I - U)^{-1}$ exists, and we may define

$$T = i(I + U)(I - U)^{-1}$$

with

$$D(T) = D\left((I - U)^{-1}\right) = \mathrm{im}(I - U)$$

dense in $\mathbf{H}$. Suppose that $x = (I - U)u, \; y = (I - U)v \; \in D(T) = \mathrm{im}(I - U)$. Then

$$(Tx, y) = (i(I + U)u, (I - U)v) = i\left[(Uu, v) - (u, Uv)\right] + i\left[(u, v) - (Uu, Uv)\right].$$

The second expression in brackets vanishes since U is an isometry. So $(Tx, y) =$

$$i(Uu, v) - i(u, Uv) = (-Uu, iv) + (u, iUv) = ([I - U]u, i[I + U]v) = (x, Ty).$$

This shows that T is symmetric.

To see that $U_T = U$ we again write $x = (I - U)u$. We have

$$Tx = i(I + U)u \; \text{ so } \; (T + iI)x = 2iu \; \text{ and } \; (T - iI)x = 2iUu.$$

Thus $D(U_T) = \{2iu \; u \in D(U)\} = D(U)$ and

$$U_T(2iu) = 2iUu = U(2iu).$$

Thus $U = U_T$.

We must still show that T is a closed operator. T maps $x_n = (I - U)u_n$ to $(I + U)u_n$. If both $(I - U)u_n$ and $(I + U)u_n$ converge, then u_n and Uu_n converge. The fact that U is closed implies that if $u = \lim u_n$ then $u \in D(U)$ and $Uu = \lim Uu_n$. But this that $(I - U)u_n \to (I - U)u$ and $i(I + U)u_n \to i(I + U)u$ so T is closed. QED

The map $T \mapsto U_T$ from symmetric operators to isometries is called the **Cayley transform**.

Recall that an isometry is unitary if its domain and image are all of $\mathbf{H}$. If U is a closed isometry, then $x_n \in D(U)$ and $x_n \to x$ implies that Ux_n is convergent, hence $x \in D(U)$ and $Ux = \lim Ux_n$. Similarly, if $Ux_n \to y$ then the x_n are Cauchy, hence convergent to an x with $Ux = y$. So for any closed isometry U the spaces $D(U)^\perp$ and $\mathrm{im}(U)^\perp$ measure how far U is from being unitary: If they both reduce to the zero subspace then U is unitary.

For a closed symmetric operator T define

$$\mathbf{H}_T^+ = \{x \in \mathbf{H} | T^* x = ix\} \quad \text{and} \quad \mathbf{H}_T^- = \{x \in \mathbf{H} | T^* x = -ix\}. \qquad (11.2)$$

The main theorem of this section is

Theorem 11.1.2 *Let T be a closed symmetric operator and $U = U_T$ its Cayley transform. Then*

$$\mathbf{H}_T^+ = D(U)^\perp \quad and \quad \mathbf{H}_T^- = (\mathrm{im}(U))^\perp.$$

Every $x \in D(T^*)$ is uniquely expressible as

$$x = x_0 + x_+ + x_-$$

with $x_0 \in D(T)$, $x_+ \in \mathbf{H}_T^+$ and $x_- \in \mathbf{H}_T^-$, so

$$T^* x = T x_0 + i x_+ - i x_-.$$

In particular, T is self adjoint if and only if U is unitary.

Proof.   To say that $x \in D(U)^\perp = D\left((T + iI)^{-1}\right)^\perp$ says that

$$(x, (T + iI)y) = 0 \quad \forall\, y \in D(T).$$

This says that

$$(x, Ty) = -(x, iy) = (ix, y) \quad \forall\, y \in D(T).$$

This is precisely the assertion that $x \in D(T^*)$ and $T^* x = ix$. We can read these equations backwards to conclude that $\mathbf{H}_T^+ = D(U)^\perp$. Similarly, if $x \in \mathrm{im}(U)^\perp$ then $(x, (T - iI)z) = 0 \ \forall z \in D(T)$ implying $T^* x = -ix$ and conversely.

We know that $D(U)$ and $\mathrm{im}(U)$ are closed subspaces of $\mathbf{H}$ so any $w \in \mathbf{H}$ can be written as the sum of an element of $D(U)$ and an element of $D(U)^\perp$. Taking $w = (T^* + iI)x$ for some $x \in D(T^*)$ gives

$$(T^* + iI)x = y_0 + x_1, \quad y_0 \in D(U) = \mathrm{im}(T + iI), \ x_1 \in D(U)^\perp.$$

We can write $y_0 = (T + iI)x_0, \ x_0 \in D(T)$ so

$$(T^* + iI)x = (T + iI)x_0 + x_1.$$

Since $T^* = T$ on $D(T)$ and $T^* x_1 = ix_1$ as $x_1 \in D(U)^\perp$ we have

$$T^* x_1 + i x_1 = 2i x_1.$$

So if we set

$$x_+ = \frac{1}{2i}x_1$$

we have

$$x_1 = (T^* + iI)x_+, \quad x_+ \in D(U)^\perp.$$

so

$$(T^* + iI)x = (T^* + iI)(x_0 + x_+)$$

or

$$T^*(x - x_0 - x_+) = -i(x - x_0 - x_+).$$

This implies that $(x - x_0 - x_+) \in \mathbf{H}_T^- = \text{im}(U)^\perp$. So if we set

$$x_- := x - x_0 - x_+$$

we get the desired decomposition $x = x_0 + x_+ + x_-$.

To show that the decomposition is unique, suppose that

$$x_0 + x_+ + x_- = 0.$$

Applying $(T^* + iI)$ gives

$$0 = (T + iI)x_0 + 2ix_+.$$

But $(T + iI)x_0 \in D(U)$ and $x_+ \in D(U)^\perp$ so both terms above must be zero, so $x_+ = 0$. Also, from the preceding theorem we know that $(T + iI)x_0 = 0 \Rightarrow x_0 = 0$. Hence since $x_0 = 0$ and $x_+ = 0$ we must also have $x_- = 0$. QED

11.1.1 An elementary example.

Take $\mathbf{H} = L_2([0, 1])$ relative to the standard Lebesgue measure. Consider the operator $\frac{1}{i}\frac{d}{dt}$ which is defined on all elements of $\mathbf{H}$ whose derivative, in the sense of distributions, is again in $L_2([0, 1])$. For any two such elements we have the integration by parts formula

$$\left(\frac{1}{i}\frac{d}{dt}x, y\right) = x(1)\overline{y(1)} - x(0)\overline{y(0)} + \left(x, \frac{1}{i}\frac{d}{dt}y\right).$$

(Even though in general the value at a point of an element in L_2 makes no sense, if x is such that $x' \in L_2$ then $\frac{1}{h}\int_0^h x(t)dt$ makes sense, and integration by parts using a continuous representative for x shows that the limit of this expression is well defined and equal to $x(0)$ for our continuous representative.) Suppose we take $T = \frac{1}{i}\frac{d}{dt}$ but with $D(T)$ consisting of those elements whose derivatives belong to L_2 as above, but which in addition satisfy

$$x(0) = x(1) = 0.$$

This space is dense in $\mathbf{H} = L_2$ but if y is *any* function whose derivative is in $\mathbf{H}$, we see from the integration by parts formula that

$$(Tx, y) = \left(x, \frac{1}{i}\frac{d}{dt}y \right).$$

In other words, using the Riesz representation theorem, we see that

$$T^* = \frac{1}{i}\frac{d}{dt}$$

defined on *all* y with derivatives in L_2. Notice that

$$T^* e^{\pm t} = \mp i e^{\pm t}$$

so in fact the spaces $\mathbf{H}_T^{\pm}$ are both one dimensional.

For each complex number $e^{i\theta}$ of absolute value one we can find a "self adjoint extension" A_θ of T, that is an operator A_θ such that

$$D(T) \subset D(A_\theta) \subset D(T^*)$$

with $D(A_\theta) = D(A_\theta^*)$, $A_\theta = A_\theta^*$ and $A_\theta = T$ on $D(T)$. Indeed, let $D(A_\theta)$ consist of all x with derivatives in L_2 and which satisfy the "boundary condition"

$$x(1) = e^{i\theta}x(0).$$

Let us compute A_θ^* and its domain. Since $D(T) \subset D(A_\theta)$, if $(A_\theta x, y) = (x, A_\theta^* y)$ we must have $y \in D(T^*)$ and $A_\theta^* y = \frac{1}{i}\frac{d}{dt}y$. But then the integration by parts formula gives

$$(Ax, y) - (x, \frac{1}{i}\frac{d}{dt}y) = e^{i\theta}x(0)\overline{y(1)} - x(0)\overline{y(0)}.$$

This will vanish for all $x \in D(A_\theta)$ if and only if $y \in D(A_\theta)$. So we see that A_θ is self adjoint.

The moral is that to construct a self adjoint operator from a differential operator which is symmetric, we may have to supplement it with appropriate boundary conditions.

On the other hand, consider the same operator $\frac{1}{i}\frac{d}{dt}$ considered as an unbounded operator on $L_2(\mathbf{R})$. We take as its domain the set of all elements of $x \in L_2(\mathbf{R})$ whose distributional derivatives belong to $L_2(\mathbf{R})$ and such that $\lim_{t \to \pm\infty} x = 0$. The functions $e^{\pm t}$ do not belong to $L_2(\mathbf{R})$ and so our operator is in fact self-adjoint. So the issue of whether or not we must add boundary conditions depends on the nature of the domain where the differential operator is to be defined. A deep analysis of this phenomenon for second order ordinary differential equations was provided by Hermann Weyl in a paper published in 1911. It is safe to say that much of the progress in the theory of self-adjoint operators was in no small measure influenced by a desire to understand and generalize the results of this fundamental paper.

11.2 Equibounded semi-groups on a Frechet space.

A Frechet space $\mathbf{F}$ is a vector space with a topology defined by a sequence of semi-norms and which is complete. An important example is the Schwartz space $\mathcal{S}$. Let $\mathbf{F}$ be such a space. We want to consider a one parameter family of operators T_t on $\mathbf{F}$ defined for all $t \geq 0$ and which satisfy the following conditions:

- $T_0 = I$

- $T_t \circ T_s = T_{t+s}$

- $\lim_{t \to t_0} T_t x = T_{t_0} x \ \forall t_0 \geq 0$ and $x \in \mathbf{F}$.

- For any defining seminorm p there is a defining seminorm q and a constant K such that $p(T_t x) \leq K q(x)$ for all $t \geq 0$ and all $x \in \mathbf{F}$.

We call such a family an **equibounded continuous semigroup**. We will usually drop the adjective "continuous" and even "equibounded" since we will not be considering any other kind of semigroup.

11.2.1 The infinitesimal generator.

We are going to begin by showing that every such semigroup has an " infinitesimal generator", i.e. can be written in some sense as $T_t = e^{At}$. It is important to observe that we have made a serious change of convention in that we are dropping the i that we have used until now. With this new notation, for example, the infinitesimal generator of a group of unitary transformations will be a skew-adjoint operator rather than a self-adjoint operator. In quantum mechanics, where an "observable" is a self-adjoint operator, there is a good reason for emphasizing the self-adjoint operators, and hence including the i. There are many good reasons for deviating from the physicists' notation, not the least having to do with the theory of Lie algebras. I do not want to go into these reasons now. Some will emerge from the ensuing notation. But the presence or absence of the i is a cultural divide between physicists and mathematicians.

So we define the operator A as

$$Ax = \lim_{t \searrow 0} \frac{1}{t}(T_t - I)x.$$

That is A is the operator defined on the domain $D(A)$ consisting of those x for which the limit exists.

Our first task is to show that $D(A)$ is dense in $\mathbf{F}$. For this we begin as promised with the putative resolvent

$$R(z) := \int_0^\infty e^{-zt} T_t dt \tag{11.3}$$

which is defined (by the boundedness and continuity properties of T_t) for all z with Re $z > 0$. We begin by checking that every element of im $R(z)$ belongs to

$D(A)$: We have

$$\frac{1}{h}(T_h - I)R(z)x = \frac{1}{h}\int_0^\infty e^{-zt}T_{t+h}xdt - \frac{1}{h}\int_0^\infty e^{-zt}T_txdt =$$

$$\frac{1}{h}\int_h^\infty e^{-z(r-h)}T_rxdr - \frac{1}{h}\int_0^\infty e^{-zt}T_txdt \;\; = \frac{e^{zh}-1}{h}\int_h^\infty e^{-zt}T_txdt - \frac{1}{h}\int_0^h e^{-zt}T_txdt$$

$$= \frac{e^{zh}-1}{h}\left[R(z)x - \int_0^h e^{-zt}T_tdt\right] - \frac{1}{h}\int_0^h e^{-zt}T_txdt.$$

If we now let $h \to 0$, the integral inside the bracket tends to zero, and the expression on the right tends to x since $T_0 = I$. We thus see that

$$R(z)x \in D(A)$$

and

$$AR(z) = zR(z) - I,$$

or, rewriting this in a more familiar form,

$$(zI - A)R(z) = I. \tag{11.4}$$

This equation says that $R(z)$ is a right inverse for $zI - A$. It will require a lot more work to show that it is also a left inverse.

We will first prove that $D(A)$ is dense in $\mathbf{F}$ by showing that $\text{im}(R(z))$ is dense. In fact, taking s to be real, we will show that

$$\lim_{s\to\infty} sR(s)x = x \quad \forall\, x \in \mathbf{F}. \tag{11.5}$$

Indeed,

$$\int_0^\infty se^{-st}dt = 1$$

for any $s > 0$. So we can write

$$sR(s)x - x = s\int_0^\infty e^{-st}[T_tx - x]dt.$$

Applying any seminorm p we obtain

$$p(sR(s)x - x) \le s\int_0^\infty e^{-st}p(T_tx - x)dt.$$

For any $\epsilon > 0$ we can, by the continuity of T_t, find a $\delta > 0$ such that

$$p(T_tx - x) < \epsilon \;\; \forall \;\; 0 \le t \le \delta.$$

Now let us write

$$s\int_0^\infty e^{-st}p(T_tx - x)dt = s\int_0^\delta e^{-st}p(T_tx - x)dt + s\int_\delta^\infty e^{-st}p(T_tx - x)dt.$$

The first integral is bounded by

$$\epsilon s \int_0^\delta e^{-st}dt \le \epsilon s \int_0^\infty e^{-st}dt = \epsilon.$$

As to the second integral, let M be a bound for $p(T_t x) + p(x)$ which exists by the uniform boundedness of T_t. The triangle inequality says that $p(T_t x - x) \le p(T_t x) + p(x)$ so the second integral is bounded by

$$M \int_\delta^\infty se^{-st}dt = Me^{-s\delta}.$$

This tends to 0 as $s \to \infty$, completing the proof that $sR(s)x \to x$ and hence that $D(A)$ is dense in $\mathbf{F}$.

11.3   The differential equation

Theorem 11.3.1 *If $x \in D(A)$ then for any $t > 0$*

$$\lim_{h \to 0} \frac{1}{h}[T_{t+h} - T_t]x = AT_t x = T_t Ax.$$

In colloquial terms, we can formulate the theorem as saying that

$$\frac{d}{dt}T_t = AT_t = T_t A$$

in the sense that the appropriate limits exist when applied to $x \in D(A)$.

Proof. Since T_t is continuous in t, we have

$$T_t Ax = T_t \lim_{h \searrow 0} \frac{1}{h}[T_h - I]x = \lim_{h \searrow 0} \frac{1}{h}[T_t T_h - T_t]x =$$

$$\lim_{h \searrow 0} \frac{1}{h}[T_{t+h} - T_t]x = \lim_{h \searrow 0} \frac{1}{h}[T_h - I]T_t x$$

for $x \in D(A)$. This shows that $T_t x \in D(A)$ and

$$\lim_{h \searrow 0} \frac{1}{h}[T_{t+h} - T_t]x = AT_t x = T_t Ax.$$

To prove the theorem we must show that we can replace $h \searrow 0$ by $h \to 0$. Our strategy is to show that with the information that we already have about the existence of right handed derivatives, we can conclude that

$$T_t x - x = \int_0^t T_s Ax \, ds.$$

Since T_t is continuous, this is enough to give the desired result. In order to establish the above equality, it is enough, by the Hahn-Banach theorem to prove that for any $\ell \in \mathbf{F}^*$ we have

$$\ell(T_t x) - \ell(x) = \int_0^t \ell(T_s A x) ds.$$

In turn, it is enough to prove this equality for the real and imaginary parts of ℓ.

So it all boils down to a lemma in the theory of functions of a real variable:

Lemma 11.3.1 *Suppose that f is a continuous real valued function of t with the property that the right hand derivative*

$$\frac{d^+}{dt} f := \lim_{h \searrow 0} \frac{f(t+h) - f(t)}{h} = g(t)$$

exists for all t and $g(t)$ is continuous. Then f is differentiable with $f' = g$.

Proof.   We first prove that $\frac{d^+}{dt} f \geq 0$ on an interval $[a, b]$ implies that $f(b) \geq f(a)$. Suppose not. Then there exists an $\epsilon > 0$ such that

$$f(b) - f(a) < -\epsilon(b - a).$$

Set

$$F(t) := f(t) - f(a) + \epsilon(t - a).$$

Then $F(a) = 0$ and

$$\frac{d^+}{dt} F > 0.$$

At a this implies that there is some $c > a$ near a with $F(c) > 0$. On the other hand, since $F(b) < 0$, and F is continuous, there will be some point $s < b$ with $F(s) = 0$ and $F(t) < 0$ for $s < t \leq b$. This contradicts the fact that $[\frac{d^+}{dt} F](s) > 0$.

Thus if $\frac{d^+}{dt} f \geq m$ on an interval $[t_1, t_2]$ we may apply the above result to $f(t) - mt$ to conclude that

$$f(t_2) - f(t_1) \geq m(t_2 - t_1),$$

and if $\frac{d^+}{dt} f(t) \leq M$ we can apply the above result to $Mt - f(t)$ to conclude that $f(t_2) - f(t_1) \leq M(t_2 - t_1)$. So if $m = \min g(t) = \min \frac{d^+}{dt} f$ on the interval $[t_1, t_2]$ and M is the maximum, we have

$$m \leq \frac{f(t_2) - f(t_1)}{t_2 - t_1} \leq M.$$

Since we are assuming that g is continuous, this is enough to prove that f is indeed differentiable with derivative g. QED

11.3.1 The resolvent.

We have already verified that

$$R(z) = \int_0^\infty e^{-zt} T_t dt$$

maps $\mathbf{F}$ into $D(A)$ and satisfies

$$(zI - A)R(z) = I$$

for all z with Re $z > 0$, cf (11.4).

We shall now show that for this range of z

$$(zI - A)x = 0 \quad \Rightarrow \quad x = 0 \quad \forall \, x \in D(A)$$

so that $(zI - A)^{-1}$ exists and that it is given by $R(z)$. Suppose that

$$Ax = zx \quad x \in D(A)$$

and choose $\ell \in \mathbf{F}^*$ with $\ell(x) = 1$. Consider

$$\phi(t) := \ell(T_t x).$$

By the result of the preceding section we know that ϕ is a differentiable function of t and satisfies the differential equation

$$\phi'(t) = \ell(T_t Ax) = \ell(T_t zx) = z\ell(T_t x) = z\phi(t), \quad \phi(0) = 1.$$

So

$$\phi(t) = e^{zt}$$

which is impossible since $\phi(t)$ is a bounded function of t and the right hand side of the above equation is not bounded for $t \geq 0$ since the real part of z is positive.

We have from (11.4) that

$$(zI - A)R(z)(zI - A)x = (zI - A)x$$

and we know that $R(z)(zI - A)x \in D(A)$. From the injectivity of $zI - A$ we conclude that $R(z)(zI - A)x = x$.

From $(zI - A)R(z) = I$ we see that $zI - A$ maps im $R(z) \subset D(A)$ onto $\mathbf{F}$ so certainly $zI - A$ maps $D(A)$ onto $\mathbf{F}$ bijectively. Hence

$$\operatorname{im}(R(z)) = D(A), \quad \operatorname{im}(zI - A) = \mathbf{F}$$

and

$$R(z) = (zI - A)^{-1}.$$

We have already established the following:

The resolvent $R(z) = R(z, A) := \int_0^\infty e^{-zt}T_t dt$ is defined as a strong limit for Re $z > 0$ and, for this range of z:

$$
\begin{align}
D(A) &= \operatorname{im}(R(z, A)) \tag{11.6}\\
AR(z, A)x = R(z, A)Ax &= (zR(z, A) - I)x \quad x \in D(A) \tag{11.7}\\
AR(z, A)x &= (zR(z, A) - I)x \quad \forall\ x \in \mathbf{F} \tag{11.8}\\
\lim_{z \nearrow \infty} zR(z, A)x &= x \quad \text{for } z \text{ real } \forall x \in \mathbf{F}. \tag{11.9}
\end{align}
$$

We also have

Theorem 11.3.2 *The operator A is closed.*

Proof. Suppose that $x_n \in D(A)$, $x_n \to x$ and $y_n \to y$ where $y_n = Ax_n$. We must show that $x \in D(A)$ and $Ax = y$. Set

$$z_n := (I - A)x_n \quad \text{so} \quad z_n \to x - y.$$

Since $R(1, A) = (I - A)^{-1}$ is a bounded operator, we conclude that

$$x = \lim x_n = \lim(I - A)^{-1}z_n = (I - A)^{-1}(x - y).$$

From (11.6) we see that $x \in D(A)$ and from the preceding equation that $(I - A)x = x - y$ so $Ax = y$. QED

Application to Stone's theorem.

We now have enough information to complete the proof of Stone's theorem:

Suppose that $U(t)$ is a one-parameter group of unitary transformations on a Hilbert space. We have $(U(t)x, y) = (x, U(t)^{-1}y) = (x, U(-t)y)$ and so differentiating at the origin shows that the infinitesimal generator A, which we know to be closed, is skew-symmetric:

$$(Ax, y) = (x, Ay) \quad \forall\ x, y \in D(A).$$

Also the resolvents $(zI - A)^{-1}$ exist for all z which are not purely imaginary, and $(zI - A)$ maps $D(A)$ onto the whole Hilbert space $\mathbf{H}$.

Writing $A = iT$ we see that T is symmetric and that its Cayley transform U_T has zero kernel and is surjective, i.e. is unitary. Hence T is self-adjoint. This proves Stone's theorem that every one parameter group of unitary transformations is of the form e^{iTt} with T self-adjoint.

11.3.2   Examples.

For $r > 0$ let

$$J_r := (I - r^{-1}A)^{-1} = rR(r, A)$$

so by (11.8) we have

$$AJ_r = r(J_r - I). \tag{11.10}$$

Translations.

Consider the one parameter group of translations acting on $L_2(\mathbf{R})$:

$$[U(t)x](s) = x(s-t). \tag{11.11}$$

This is defined for all $x \in \mathcal{S}$ and is an isometric isomorphism there, so extends to a unitary one parameter group acting on $L_2(\mathbf{R})$. Equally well, we can take the above equation in the sense of distributions, where it makes sense for all elements of $\mathcal{S}'$, in particular for all elements of $L_2(\mathbf{R})$. We know that we can differentiate in the distributional sense to obtain

$$A = -\frac{d}{ds}$$

as the "infinitesimal generator" in the distributional sense. Let us see what the general theory gives. Let $y_r := J_r x$ so

$$y_r(s) = r \int_0^\infty e^{-rt} x(s-t) dt = r \int_{-\infty}^s e^{-r(s-u)} x(u) du.$$

The right hand expression is a differentiable function of s and

$$y_r'(s) = rx(s) - r^2 \int_s^\infty e^{-r(s-u)} x(u) du = rx(s) - ry_r(s).$$

On the other hand we know from (11.10) that

$$Ay_r = AJ_r x = r(y_r - x).$$

Putting the two equations together gives

$$A = -\frac{d}{ds}$$

as expected. This is a skew-adjoint operator in accordance with Stone's theorem.

We can now go back and give an intuitive explanation of what goes wrong when considering this same operator A but on $L_2[0,1]$ instead of on $L_2(\mathbf{R})$. If x is a smooth function of compact support lying in $(0,1)$, then x can not tell whether it is to be thought of as lying in $L_2([0,1])$ or $L_2(\mathbf{R})$, so the only choice for a unitary one parameter group acting on x (at least for small $t > 0$) if the shift to the right as given by (11.11). But once t is large enough that the support of $U(t)x$ hits the right end point, 1, this transformation can not continue as is. The only hope is to have what "goes out" the right hand side come in, in some form, on the left, and unitarity now requires that

$$\int_0^1 |x(s-t)|^2 dt = \int_0^1 |x(t)|^2 dt$$

where now the shift in (11.11) means mod 1. This still allows freedom in the choice of phase between the exiting value of the x and its incoming value. Thus we specify a unitary one parameter group when we fix a choice of phase as the effect of "passing go". This choice of phase is the origin of the θ that are needed to introduce in finding the self adjoint extensions of $\frac{1}{i}\frac{d}{dt}$ acting on functions vanishing at the boundary.

The heat equation.

Let **F** consist of the bounded uniformly continuous functions on **R**. For $t > 0$ define

$$[T_t x](s) = \frac{1}{\sqrt{2\pi t}} \int_{-\infty}^{\infty} e^{-(s-v)/2t} x(v) dv.$$

In other words, T_t is convolution with

$$n_t(u) = \frac{1}{\sqrt{2\pi t}} e^{-u^2/2t}.$$

We have already verified in our study of the Fourier transform that this is a continuous semi-group (when we set $T_0 = I$) when acting on $\mathcal{S}$. In fact, for $x \in \mathcal{S}$, we can take the Fourier transform and conclude that

$$[T_t x]\hat{\ }(\sigma) = e^{-i\sigma^2 t/2} \hat{x}(\sigma).$$

Differentiating this with respect to t and setting $t = 0$ (and taking the inverse Fourier transform) shows that

$$\left[\frac{d}{dt} T_t x \right]_{t=0} = \frac{1}{2} \frac{d^2}{ds^2} x$$

for $x \in \mathcal{S}$. We wish to arrive at the same result for T_t acting on **F**. It is easy enough to verify that the operators T_t are continuous in the uniform norm and hence extend to an equibounded semigroup on **F**. We will now verify that the infinitesimal generator A of this semigroup is

$$A = \frac{1}{2} \frac{d^2}{ds^2}$$

with domain consisting of all twice differentiable functions.

Let us set $y_r = J_r x$ so

$$
\begin{aligned}
y_r(s) &= \int_{-\infty}^{\infty} x(v) \left[\int_0^{\infty} r \frac{1}{\sqrt{2\pi t}} e^{-rt-(s-v)^2/2t} dt \right] dv \\
&= \int_{-\infty}^{\infty} x(v) \left[\int_0^{\infty} 2\sqrt{r} \frac{1}{\sqrt{2\pi}} e^{-\sigma^2 - r(s-v)^2/2\sigma^2} d\sigma \right] dv \quad \text{setting } t = \sigma^2/r \\
&= \int_{-\infty}^{\infty} x(v) (r/2)^{\frac{1}{2}} e^{-\sqrt{2r}|s-v|} dv
\end{aligned}
$$

since for any $c > 0$ we have

$$\int_0^{\infty} e^{-(\sigma^2 + c^2/\sigma^2)} d\sigma = \frac{\sqrt{\pi}}{2} e^{-2c}. \tag{11.12}$$

Let me postpone the calculation of this integral to the end of the subsection. Assuming the evaluation of this integral we can write

$$y_r(s) = \left(\frac{r}{2} \right)^{\frac{1}{2}} \left[\int_s^{\infty} x(v) e^{-\sqrt{2r}(v-s)} dv + \int_{-\infty}^{s} x(v) e^{-\sqrt{2r}(s-v)} dv \right].$$

This is a differentiable function of s and we can differentiate to obtain

$$y'_r(s) = r\left[\int_s^\infty x(v)e^{-\sqrt{2r}(v-s)}dv - \int_{-\infty}^s x(v)e^{-\sqrt{2r}(s-v)}dv\right].$$

This is also differentiable and compute its derivative to obtain

$$y''_r(s) = -2rx(s) + r^{3/2}\sqrt{2}\int_{-\infty}^\infty x(v))e^{-\sqrt{2r}|v-s|}dv,$$

or

$$y''_r = 2r(y_r - x).$$

Comparing this with (11.10) which says that $Ay_r = r(y_r - x)$ we see that indeed

$$A = \frac{1}{2}\frac{d^2}{ds^2}.$$

Let us now verify the evaluation of the integral in (11.12): Start with the known integral

$$\int_0^\infty e^{-x^2}dx = \frac{\sqrt{\pi}}{2}.$$

Set $x = \sigma - c/\sigma$ so that $dx = (1 + c/\sigma^2)d\sigma$ and $x = 0$ corresponds to $\sigma = \sqrt{c}$. Thus $\frac{\sqrt{\pi}}{2} =$

$$\int_{\sqrt{c}}^\infty e^{-(\sigma-c/\sigma)^2}(1 + c/\sigma)^2 d\sigma = e^{2c}\int_{\sqrt{c}}^\infty e^{-(\sigma^2+c^2/\sigma^2)}(1 + c/\sigma^2)d\sigma$$

$$= e^{2c}\left[\int_{\sqrt{c}}^\infty e^{-(\sigma^2+c^2/\sigma^2)}d\sigma + \int_{\sqrt{c}}^\infty e^{-(\sigma^2+c^2/\sigma^2)}\frac{c}{\sigma^2}d\sigma\right].$$

In the second integral inside the brackets set $t = -c/\sigma$ so $dt = \frac{c}{\sigma^2}d\sigma$ and this second integral becomes

$$\int_0^{\sqrt{c}} e^{-(t^2+c^2/t^2)}dt$$

and hence

$$\frac{\sqrt{\pi}}{2} = e^{2c}\int_0^\infty e^{-(\sigma^2+c^2/\sigma^2)}d\sigma$$

which is (11.12).

Bochner's theorem.

A complex valued continuous function F is called **positive definite** if for every continuous function ϕ of compact support we have

$$\int_{\mathbf{R}}\int_{\mathbf{R}} F(t-s)\phi(t)\overline{\phi(s)}dtds \geq 0. \tag{11.13}$$

We can write this as

$$(F \star \overline{\phi}, \overline{\phi}) \geq 0$$

where the convolution is taken in the sense of generalized functions. If we write $F = \hat{G}$ and $\overline{\phi} = \hat{\psi}$ then by Plancherel this equation becomes

$$(G\psi, \psi) \geq 0$$

or

$$\langle G, |\psi|^2 \rangle \geq 0$$

which will certainly be true if G is a finite non-negative measure. Bochner's theorem asserts the converse: that any positive definite function is the Fourier transform of a finite non-negative measure. We shall follow Yosida pp. 346-347 in showing that Stone's theorem implies Bochner's theorem.

Let $\mathcal{F}$ denote the space of functions on $\mathbf{R}$ which have finite support, i.e. vanish outside a finite set. This is a complex vector space, and has the semi-scalar product

$$(x, y) := \sum_{t,s} F(t - s)x(t)\overline{y(s)}.$$

(It is easy to see that the fact that F is a positive definite function implies that $(x, x) \geq 0$ for all $x \in \mathcal{F}$.) Passing to the quotient by the subspace of null vectors and completing we obtain a Hilbert space $\mathbf{H}$.

Let U_r be defined by $[U_r x](t) = x(t - r)$ as usual. Then

$$(U_r x, U_r y) = \sum_{t,s} F(t-s)x(t-r)\overline{y(s - r)} = \sum_{t,s} F(t+r-(s+r))x(t)\overline{y(s)} = (x, y).$$

So U_r descends to $\mathbf{H}$ to define a unitary operator which we shall continue to denote by U_r. We thus obtain a one parameter group of unitary transformations on $\mathbf{H}$. According to Stone's theorem there exists a resolution E_λ of the identity such that

$$U_t = \int_{-\infty}^{\infty} e^{it\lambda} dE_\lambda.$$

Now choose $\delta \in \mathcal{F}$ to be defined by

$$\delta(t) = \begin{cases} 1 & \text{if} \quad t = 0 \\ 0 & \text{if} \quad t \neq 0 \end{cases}.$$

Let x be the image of δ in $\mathbf{H}$. Then

$$(U_r x, x) = \sum F(t - s)\delta(t - r)\delta(s) = F(r).$$

But by Stone we have

$$F(r) = \int_{-\infty}^{\infty} e^{ir\lambda} d\mu_{x,x} = \int_{-\infty}^{\infty} e^{ir\lambda} d\|E_\lambda x\|^2$$

so we have represented F as the Fourier transform of a finite non-negative measure. QED

The logic of our argument has been - the Spectral Theorem implies Stone's theorem implies Bochner's theorem. In fact, assuming the Hille-Yosida theorem on the existence of semigroups to be proved below, one can go in the opposite direction. Given a one parameter group $U(t)$ of unitary transformations, it is easy to check that for any $x \in \mathbf{H}$ the function $t \mapsto (U(t)x, x)$ is positive definite, and then use Bochner's theorem to derive the spectral theorem on the cyclic subspace generated by x under $U(t)$. One can then get the full spectral theorem in multiplication operator form as we did in the handout on unbounded self-adjoint operators.

11.4 The power series expansion of the exponential.

In finite dimensions we have the formula

$$e^{tB} = \sum_0^\infty \frac{t^k}{k!} B^k$$

with convergence guaranteed as a result of the convergence of the usual exponential series in one variable. (There are serious problems with this definition from the point of view of numerical implementation which we will not discuss here.)

In infinite dimensional spaces some additional assumptions have to be placed on an operator B before we can conclude that the above series converges. Here is a very stringent condition which nevertheless suffices for our purposes.

Let $\mathbf{F}$ be a Frechet space and B a continuous map of $\mathbf{F} \to \mathbf{F}$. We will assume that the B^k are **equibounded** in the sense that for any defining semi-norm p there is a constant K and a defining semi-norm q such that

$$p(B^k x) \leq K q(x) \quad \forall \, k = 1, 2, \dots \quad \forall \, x \in \mathbf{F}.$$

Here the K and q are required to be independent of k and x.

Then

$$p\left(\sum_m^n \frac{t^k}{k!} B^k x\right) \leq \sum_m^n \frac{t^k}{k!} p(B^k x) \leq K q(x) \sum_n^n \frac{t^k}{k!}$$

and so

$$\sum_0^n \frac{t^k}{k!} B^k x$$

is a Cauchy sequence for each fixed t and x (and uniformly in any compact interval of t). It therefore converges to a limit. We will denote the map $x \mapsto \sum_0^\infty \frac{t^k}{k!} B^k x$ by

$$\exp(tB).$$

This map is linear, and the computation above shows that

$$p(\exp(tB)x) \le K \exp(t)q(x).$$

The usual proof (using the binomial formula) shows that $t \mapsto \exp(tB)$ is a one parameter equibounded semi-group. More generally, if B and C are two such operators then if $BC = CB$ then $\exp(t(B + C)) = (\exp tB)(\exp tC)$.

Also, from the power series it follows that the infinitesimal generator of $\exp tB$ is B.

11.5   The Hille Yosida theorem.

Let us now return to the general case of an equibounded semigroup T_t with infinitesimal generator A on a Frechet space $\mathbf{F}$ where we know that the resolvent $R(z, A)$ for Re $z > 0$ is given by

$$R(z, A)x = \int_0^\infty e^{-zt} T_t x dt.$$

This formula shows that $R(z, A)x$ is continuous in z. The resolvent equation

$$R(z, A) - R(w, A) = (w - z)R(z, A)R(w, A)$$

then shows that $R(z, A)x$ is complex differentiable in z with derivative $-R(z, A)^2 x$. It then follows that $R(z, A)x$ has complex derivative of all order given by

$$\frac{d^n R(z, A)x}{dz^n} = (-1)^n n! R(z, A)^{n+1} x.$$

On the other hand, differentiating the integral formula for the resolvent n- times gives
$$\frac{d^n R(z, A)x}{dz^n} = \int_0^\infty e^{-zt}(-t)^n T_t dt$$

where differentiation under the integral sign is justified by the fact that the T_t are equicontinuous in t. Putting the previous two equations together gives

$$(zR(z, A))^{n+1} x = \frac{z^{n+1}}{n!} \int_0^\infty e^{-zt} t^n T_t x dt.$$

This implies that for any semi-norm p we have

$$p((zR(z, A))^{n+1} x) \le \frac{z^{n+1}}{n!} \int_0^\infty e^{-zt} t^n \sup_{t \ge 0} p(T_t x) dt = \sup_{t \ge 0} p(T_t x)$$

since
$$\int_0^\infty e^{-zt} t^n dt = \frac{n!}{z^{n+1}}.$$

Since the T_t are equibounded by hypothesis, we conclude

Proposition 11.5.1 *The family of operators $\{(zR(z, A))^n\}$ is equibounded in Re $z > 0$ and $n = 0, 1, 2, \ldots$.*

We now come to the main result of this section:

Theorem 11.5.1 [Hille-Yosida.] *Let A be an operator with dense domain $D(A)$, and such that the resolvents*

$$R(n, A) = (nI - A)^{-1}$$

exist and are bounded operators for $n = 1, 2, \ldots$. Then A is the infinitesimal generator of a uniquely determined equibounded semigroup if and only if the operators

$$\{(I - n^{-1}A)^{-m}\}$$

are equibounded in $m = 0, 1, 2 \ldots$ and $n = 1, 2, \ldots$.

Proof. If A is the infinitesimal generator of an equibounded semi-group then we know that the $\{(I - n^{-1}A)^{-m}\}$ are equibounded by virtue of the preceding proposition. So we must prove the converse.

Set
$$J_n = (I - n^{-1}A)^{-1}$$

so $J_n = n(nI - A)^{-1}$ and so for $x \in D(A)$ we have

$$J_n(nI - A)x = nx$$

or

$$J_n Ax = n(J_n - I)x.$$

Similarly $(nI - A)J_n = nI$ so $AJ_n = n(J_n - I)$. Thus we have

$$AJ_n x = J_n Ax = n(J_n - I)x \quad \forall\, x \in D(A). \tag{11.14}$$

The idea of the proof is now this: By the results of the preceding section, we can construct the one parameter semigroup $s \mapsto \exp(sJ_n)$. Set $s = nt$. We can then form $e^{-nt} \exp(ntJ_n)$ which we can write as $\exp(tn(J_n - I)) = \exp(tAJ_n)$ by virtue of (11.14). We expect from (11.5) that

$$\lim_{n \to \infty} J_n x = x \quad \forall \quad x \in \mathbf{F}. \tag{11.15}$$

This then suggests that the limit of the $\exp(tAJ_n)$ be the desired semi-group.

So we begin by proving (11.15). We first prove it for $x \in D(A)$. For such x we have $(J_n - I)x = n^{-1}J_n Ax$ by (11.14) and this approaches zero since the J_n are equibounded. But since $D(A)$ is dense in $\mathbf{F}$ and the J_n are equibounded we conclude that (11.15) holds for all $x \in \mathbf{F}$.

Now define

$$T_t^{(n)} = \exp(tAJ_n) := \exp(nt(J_n - I)) = e^{-nt} \exp(ntJ_n).$$

We know from the preceding section that

$$p(\exp(ntJ_n)x) \le \sum \frac{(nt)^k}{k!} p(J_n^k x) \le e^{nt} K q(x)$$

which implies that

$$p(T_t^{(n)} x) \le K q(x). \tag{11.16}$$

Thus the family of operators $\{T_t^{(n)}\}$ is equibounded for all $t \ge 0$ and $n = 1, 2, \dots$. We next want to prove that the $\{T_t^{(n)}\}$ converge as $n \to \infty$ uniformly on each compact interval of t:

The J_n commute with one another by their definition, and hence J_n commutes with $T_t^{(m)}$. By the semi-group property we have

$$\frac{d}{dt} T_t^m x = A J_m T_t^{(m)} x = T_t^{(m)} A J_m x$$

so

$$T_t^{(n)} x - T_t^{(m)} x = \int_0^t \frac{d}{ds} (T_{t-s}^{(m)} T_s^{(n)}) x \, ds = \int_0^t T_{t-s}^{(m)} (A J_n - A J_m) T_s^{(n)} x \, ds.$$

Applying the semi-norm p and using the equiboundedness we see that

$$p(T_t^{(n)} x - T_t^{(m)} x) \le K t q((J_n - J_m) A x).$$

From (11.15) this implies that the $T_t^{(n)} x$ converge (uniformly in every compact interval of t) for $x \in D(A)$, and hence since $D(A)$ is dense and the $T_t^{(n)}$ are equicontinuous for all $x \in \mathbf{F}$. The limiting family of operators T_t are equicontinuous and form a semi-group because the $T_t^{(n)}$ have this property.

We must show that the infinitesimal generator of this semi-group is A. Let us temporarily denote the infinitesimal generator of this semi-group by B, so that we want to prove that $A = B$. Let $x \in D(A)$. We claim that

$$\lim_{n \to \infty} T_t^{(n)} A J_n x = T_t A x \tag{11.17}$$

uniformly in in any compact interval of t. Indeed, for any semi-norm p we have

$$
\begin{aligned}
p(T_t A x - T_t^{(n)} A J_n x) &\le & p(T_t A x - T_t^{(n)} A x) + p(T_t^{(n)} A x - T_t^{(n)} A J_n x) \\
&\le & p((T_t - T_t^{(n)}) A x) + K q(A x - J_n A x)
\end{aligned}
$$

where we have used (11.16) to get from the second line to the third. The second term on the right tends to zero as $n \to \infty$ and we have already proved that the first term converges to zero uniformly on every compact interval of t. This establishes (11.17).

We thus have, for $x \in D(A)$,

$$
\begin{aligned}
T_t x - x &= \lim_{n \to \infty} (T_t^{(n)} x - x) \\
&= \lim_{n \to \infty} \int_0^t T_s^{(n)} A J_n x \, ds \\
&= \int_0^t (\lim_{n \to \infty} T_s^{(n)} A J_n x) \, ds \\
&= \int_0^t T_s A x \, ds
\end{aligned}
$$

where the passage of the limit under the integral sign is justified by the uniform convergence in t on compact sets. It follows from $T_t x - x = \int_0^t T_s A x \, ds$ that x is in the domain of the infinitesimal operator B of T_t and that $Bx = Ax$. So B is an extension of A in the sense that $D(B) \supset D(A)$ and $Bx = Ax$ on $D(A)$.

But since B is the infinitesimal generator of an equibounded semi-group, we know that $(I - B)$ maps $D(B)$ onto $\mathbf{F}$ bijectively, and we are assuming that $(I - A)$ maps $D(A)$ onto $\mathbf{F}$ bijectively. Hence $D(A) = D(B)$. QED

In case $\mathbf{F}$ is a Banach space, so there is a single norm $p = \| \, \|$, the hypotheses of the theorem read: $D(A)$ is dense in $\mathbf{F}$, the resolvents $R(n, A)$ exist for all integers $n = 1, 2, \ldots$ and there is a constant K independent of n and m such that

$$
\|(I - n^{-1} A)^{-m}\| \leq K \quad \forall \, n = 1, 2, \ldots, \ m = 1, 2, \ldots. \tag{11.18}
$$

11.6 Contraction semigroups.

In particular, if A satisfies

$$
\|(I - n^{-1} A)^{-1}\| \leq 1 \tag{11.19}
$$

condition (11.18) is satisfied, and such an A then generates a semi-group. Under this stronger hypothesis we can draw a stronger conclusion: In (11.16) we now have $p = q = \| \cdot \|$ and $K = 1$. Since $\lim_{n \to \infty} T_t^n x = T_t x$ we see that under the hypothesis (11.19) we can conclude that

$$
\|T_t\| \leq 1 \quad \forall \, t \geq 0.
$$

A semi-group T_t satisfying this condition is called a **contraction semi-group**. We will study another useful condition for recognizing a contraction semigroup in the following subsection.

We have already given a direct proof that if S is a self-adjoint operator on a Hilbert space then the resolvent exists for all non-real z and satisfies

$$
\|R(z, S)\| \leq \frac{1}{|\operatorname{Im}(z)|}.
$$

This implies (11.19) for $A = iS$ and $-iS$ giving us an independent proof of the existence of $U(t) = \exp(iSt)$ for any self-adjoint operator S. As we mentioned previously, we could then use Bochner's theorem to give a third proof of the spectral theorem for unbounded self-adjoint operators. I might discuss Bochner's theorem in the context of generalized functions later probably next semester if at all. Once we give an independent proof of Bochner's theorem then indeed we will get a third proof of the spectral theorem.

11.6.1   Dissipation and contraction.

Let $\mathbf{F}$ be a Banach space. Recall that a semi-group T_t is called a **contraction semi-group** if

$$\|T_t\| \leq 1 \quad \forall\, t \geq 0,$$

and that (11.19) is a sufficient condition on operator with dense domain to generate a contraction semi-group.

The Lumer-Phillips theorem to be stated below gives a necessary and sufficient condition on the infinitesimal generator of a semi-group for the semi-group to be a contraction semi-group. It is generalization of the fact that the resolvent of a self-adjoint operator has $\pm i$ in its resolvent set.

The first step is to introduce a sort of fake scalar product in the Banach space $\mathbf{F}$. A **semi-scalar product** on $\mathbf{F}$ is a rule which assigns a number $\langle\langle x, z \rangle\rangle$ to every pair of elements $x, z \in \mathbf{F}$ in such a way that

$$
\begin{aligned}
\langle\langle x + y, z \rangle\rangle &= \langle\langle x, z \rangle\rangle + \langle\langle y, z \rangle\rangle \\
\langle\langle \lambda x, z \rangle\rangle &= \lambda \langle\langle x, z \rangle\rangle \\
\langle\langle x, x \rangle\rangle &= \|x\|^2 \\
|\langle\langle x, z \rangle\rangle| &\leq \|x\| \cdot \|z\|.
\end{aligned}
$$

We can always choose a semi-scalar product as follows: by the Hahn-Banach theorem, for each $z \in \mathbf{F}$ we can find an $\ell_z \in \mathbf{F}^*$ such that

$$\|\ell_z\| = \|z\| \quad \text{and} \quad \ell_z(z) = \|z\|^2.$$

Choose one such ℓ_z for each $z \in \mathbf{F}$ and set

$$\langle\langle x, z \rangle\rangle := \ell_z(x).$$

Clearly all the conditions are satisfied. Of course this definition is highly unnatural, unless there is some reasonable way of choosing the ℓ_z other than using the axiom of choice. In a Hilbert space, the scalar product is a semi-scalar product.

An operator A with domain $D(A)$ on $\mathbf{F}$ is called **dissipative** relative to a given semi-scalar product $\langle\langle \cdot, \cdot \rangle\rangle$ if

$$\text{Re}\, \langle\langle Ax, x \rangle\rangle \leq 0 \quad \forall\, x \in D(A).$$

For example, if A is a symmetric operator on a Hilbert space such that

$$(Ax, x) \leq 0 \quad \forall\, x \in D(A) \tag{11.20}$$

then A is dissipative relative to the scalar product.

Theorem 11.6.1 [Lumer-Phillips.] *Let A be an operator on a Banach space* **F** *with $D(A)$ dense in* **F**. *Then A generates a contraction semi-group if and only if A is dissipative with respect to any semi-scalar product and*

$$\operatorname{im}(I - A) = \mathbf{F}.$$

Proof. Suppose first that $D(A)$ is dense and that $\operatorname{im}(I - A) = \mathbf{F}$. We wish to show that (11.19) holds, which will guarantee that A generates a contraction semi-group. Let $s > 0$. Then if $x \in D(A)$ and $y = sx - Ax$ then

$$s\|x\|^2 = s\langle\langle x, x\rangle\rangle \leq s\langle\langle x, x\rangle\rangle - \operatorname{Re} \langle\langle Ax, x\rangle\rangle = \operatorname{Re} \langle\langle y, x\rangle\rangle$$

implying

$$s\|x\|^2 \leq \|y\|\|x\|. \tag{11.21}$$

We are assuming that $\operatorname{im}(I - A) = \mathbf{F}$. This together with (11.21) with $s = 1$ implies that $R(1, A)$ exists and

$$\|R(1, A)\| \leq 1.$$

In turn, this implies that for all z with $|z - 1| < 1$ the resolvent $R(z, A)$ exists and is given by the power series

$$R(z, A) = \sum_{n=0}^{\infty} (z - 1)^n R(1, A)^{n+1}$$

by our general power series formula for the resolvent. In particular, for s real and $|s - 1| < 1$ the resolvent exists, and then (11.21) implies that $\|R(s, A)\| \leq s^{-1}$. Repeating the process we keep enlarging the resolvent set $\rho(A)$ until it includes the whole positive real axis and conclude from (11.21) that $\|R(s, A)\| \leq s^{-1}$ which implies (11.19). As we are assuming that $D(A)$ is dense we conclude that A generates a contraction semigroup.

Conversely, suppose that T_t is a contraction semi-group with infinitesimal generator A. We know that A is dense. Let $\langle\langle \cdot, \cdot \rangle\rangle$ be any semi-scalar product. Then

$$\operatorname{Re} \langle\langle T_t x - x, x\rangle\rangle = \operatorname{Re} \langle\langle T_t x, x\rangle\rangle - \|x\|^2 \leq \|T_t x\|\|x\| - \|x\|^2 \leq 0.$$

Dividing by t and letting $t \searrow 0$ we conclude that $\operatorname{Re} \langle\langle Ax, x\rangle\rangle \leq 0$ for all $x \in D(A)$, i.e. A is dissipative for $\langle\langle \cdot, \cdot \rangle\rangle$. QED

Once again, this gives a direct proof of the existence of the unitary group generated generated by a skew adjoint operator.

A useful way of verifying the condition $\operatorname{im}(I - A) = \mathbf{F}$ is the following: Let $A^* : \mathbf{F}^* \to \mathbf{F}^*$ be the adjoint operator which is defined if we assume that $D(A)$ is dense.

Proposition 11.6.1 *Suppose that A is densely defined and closed, and suppose that both A and A^* are dissipative. Then $\operatorname{im}(I - A) = \mathbf{F}$ and hence A generates a contraction semigroup.*

Proof. The fact that A is closed implies that $(I - A)^{-1}$ is closed, and since we know that $(I - A)^{-1}$ is bounded from the fact that A is dissipative, we conclude that $\operatorname{im}(I - A)$ is a closed subspace of F. If it were not the whole space there would be an $\ell \in F^*$ which vanished on this subspace, i.e.

$$\langle \ell, x - Ax \rangle = 0 \quad \forall \, x \in D(A).$$

This implies that that $\ell \in D(A^*$ and $A^* \ell = \ell$ which can not happen if A^* is dissipative by (11.21) applied to A^* and $s = 1$. QED

11.6.2   A special case: $\exp(t(B - I))$ with $\|B\| \leq 1$.

Suppose that $B : F \to F$ is a bounded operator on a Banach space with $\|B\| \leq 1$. Then for any semi-scalar product we have

$$\operatorname{Re} \langle \langle (B - I)x, x \rangle \rangle = \operatorname{Re} \langle \langle Tx, x \rangle \rangle - \|x\|^2 \leq \|Bx\| \|x\| - \|x\|^2 \leq 0$$

so $B - I$ is dissipative and hence $\exp(t(B - I))$ exists as a contraction semi-group by the Lumer-Phillips theorem. We can prove this directly since we can write

$$\exp(t(B - I)) = e^{-t} \sum_{k=0}^{\infty} \frac{t^k B^k}{k!}.$$

The series converges in the uniform norm and we have

$$\| \exp(t(B - I)) \| \leq e^{-t} \sum_{k=0}^{\infty} \frac{t^k \|B\|^k}{k!} \leq 1.$$

For future use (Chernoff's theorem and the Trotter product formula) we record (and prove) the following inequality:

$$\|[\exp(n(B - I)) - B^n]x\| \leq \sqrt{n} \|(B - I)x\| \quad \forall \, x \in \mathbf{F}, \text{ and } \forall \, n = 1, 2, 3 \ldots .$$
$$\tag{11.22}$$

Proof.

$$
\begin{aligned}
\|[\exp(n(B - I)) - B^n]x\| &= \left\| e^{-n} \sum_{k=0}^{\infty} \frac{n^k}{k!}(B^k - B^n)x \right\| \\
&\leq e^{-n} \sum_{k=0}^{\infty} \frac{n^k}{k!} \|(B^k - B^n)x\| \\
&\leq e^{-n} \sum_{k=0}^{\infty} \frac{n^k}{k!} \|(B^{|k-n|} - I)x\| \\
&= e^{-n} \sum_{k=0}^{\infty} \frac{n^k}{k!} \|(B - I)(I + B + \cdots + B^{(|k-n|-1)})x\| \\
&\leq e^{-n} \sum_{k=0}^{\infty} \frac{n^k}{k!} |k - n| \|(B - I)x\|.
\end{aligned}
$$

So to prove (11.22) it is enough establish the inequality

$$
e^{-n} \sum_{k=0}^{\infty} \frac{n^k}{k!} |k - n| \leq \sqrt{n}. \tag{11.23}
$$

Consider the space of all sequences $\mathbf{a} = \{a_0, a_1, \dots\}$ with scalar product

$$
(\mathbf{a}, \mathbf{b}) := e^{-n} \sum_{k=0}^{\infty} \frac{n^k}{k!} a_k \overline{b_k}.
$$

The Cauchy-Schwarz inequality applied to $\mathbf{a}$ with $a_k = |k - n|$ and $\mathbf{b}$ with $b_k \equiv 1$ gives

$$
e^{-n} \sum_{k=0}^{\infty} \frac{n^k}{k!} |k - n| \leq \sqrt{e^{-n} \sum_{k=0}^{\infty} \frac{n^k}{k!} (k - n)^2} \cdot \sqrt{e^{-n} \sum_{k=0}^{\infty} \frac{n^k}{k!}}.
$$

The second square root is one, and we recognize the sum under the first square root as the variance of the Poisson distribution with parameter n, and we know that this variance is n. QED

11.7 Convergence of semigroups.

We are going to be interested in the following type of result. We would like to know that if A_n is a sequence of operators generating equibounded one parameter semi-groups $\exp t A_n$ and $A_n \to A$ where A generates an equibounded semi-group $\exp t A$ then the semi-groups converge, i.e. $\exp t A_n \to \exp t A$. We will prove such a result for the case of contractions. But before we can even formulate the result, we have to deal with the fact that each A_n comes equipped with its own domain of definition, $D(A_n)$. We do not want to make the overly

restrictive hypothesis that these all coincide, since in many important applications they won't.

For this purpose we make the following definition. Let us assume that $\mathbf{F}$ is a Banach space and that A is an operator on $\mathbf{F}$ defined on a domain $D(A)$. We say that a linear subspace $\mathbf{D} \subset D(A)$ is a **core** for A if the closure $\overline{A}$ of A and the closure of A restricted to $\mathbf{D}$ are the same: $\overline{A} = \overline{A|\mathbf{D}}$. This certainly implies that $D(A)$ is contained in the closure of $A|\mathbf{D}$. In the cases of interest to us $D(A)$ is dense in $\mathbf{F}$, so that every core of A is dense in $\mathbf{F}$.

We begin with an important preliminary result:

Proposition 11.7.1 *Suppose that A_n and A are dissipative operators, i.e. generators of contraction semi-groups. Let $\mathbf{D}$ be a core of A. Suppose that for each $x \in \mathbf{D}$ we have that $x \in D(A_n)$ for sufficiently large n (depending on x) and that*

$$A_n x \to Ax. \tag{11.24}$$

Then for any z with $\mathrm{Re}\, z > 0$ and for all $y \in \mathbf{F}$

$$R(z, A_n)y \to R(z, A)y. \tag{11.25}$$

Proof. We know that the $R(z, A_n)$ and $R(z, A)$ are all bounded in norm by $1/\mathrm{Re}\, z$. So it is enough for us to prove convergence on a dense set. Since $(zI - A)D(A) = \mathbf{F}$, it follows that $(zI - A)D$ is dense in $\mathbf{F}$. So in proving (11.25) we may assume that $y = (zI - A)x$ with $x \in D$. Then

$$
\begin{aligned}
\|R(z, A_n)y - R(z, A)y\| &= \|R(z, A_n)(zI - A)x - x\| \\
&= \|R(z, A_n)(zI - A_n)x + R(z, A_n)(A_n x - Ax) - x\| \\
&= \|R(z, A_n)(A_n - A)x\| \\
&\leq \frac{1}{\mathrm{Re}\, z}\|(A_n - A)x\| \to 0,
\end{aligned}
$$

where, in passing from the first line to the second we are assuming that n is chosen sufficiently large that $x \in D(A_n)$. QED

Theorem 11.7.1 *Under the hypotheses of the preceding proposition,*

$$(\exp(tA_n))x \to (\exp(tA))x$$

for each $x \in \mathbf{F}$ uniformly on every compact interval of t.

Proof. Let

$$\phi_n(t) := e^{-t}[((\exp(tA_n))x - (\exp(tA))x)] \ \text{ for } t \geq 0$$

and set $\phi(t) = 0$ for $t < 0$. It will be enough to prove that these $\mathbf{F}$ valued functions converge uniformly in t to 0, and since $\mathbf{D}$ is dense and since the operators entering into the definition of ϕ_n are uniformly bounded in n, it is enough to prove this convergence for $x \in \mathbf{D}$ which is dense. We claim that

for fixed $x \in \mathbf{D}$ the functions $\phi_n(t)$ are uniformly equi-continuous. To see this observe that

$$\frac{d}{dt}\phi_n(t) = e^{-t}[(\exp(tA_n))A_nx - (\exp(tA))Ax] - e^{-t}[(\exp(tA_n))x - (\exp(tA))x]$$

for $t \geq 0$ and the right hand side is uniformly bounded in $t \geq 0$ and n.

So to prove that $\phi_n(t)$ converges uniformly in t to 0, it is enough to prove this fact for the convolution $\phi_n \star \rho$ where ρ is any smooth function of compact support, since we can choose the ρ to have small support and integral $\sqrt{2\pi}$, and then $\phi_n(t)$ is close to $(\phi_n \star \rho)(t)$.

Now the Fourier transform of $\phi_n \star \rho$ is the product of their Fourier transforms: $\hat{\phi}_n\hat{\rho}$. We have $\hat{\phi}_n(s) =$

$$\frac{1}{\sqrt{2\pi}} \int_0^\infty e^{(-1-is)t}[(\exp tA_n)x - (\exp(tA))x]dt = \frac{1}{\sqrt{2\pi}}[R(1+is, A_n)x - R(1+is, A)x].$$

Thus by the proposition

$$\hat{\phi}_n(s) \to 0,$$

in fact uniformly in s. Hence using the Fourier inversion formula and, say, the dominated convergence theorem (for Banach space valued functions),

$$(\phi_n \star \phi)(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^\infty \hat{\phi}_n(s)\hat{\rho}(s)e^{ist}ds \to 0$$

uniformly in t. QED

The preceding theorem is the limit theorem that we will use in what follows. However, there is an important theorem valid in an arbitrary Frechet space, and which does not assume that the A_n converge, or the existence of the limit A, but only the convergence of the resolvent at a single point z_0 in the right hand plane!

In the following $\mathbf{F}$ is a Frechet space and $\{\exp(tA_n)\}$ is a family of of equibounded semi-groups which is also equibounded in n, so for every semi-norm p there is a semi-norm q and a constant K such that

$$p(\exp(tA_n)x) \leq Kq(x) \quad \forall\, x \in F$$

where K and q are independent of t and n. I will state the theorem here, and refer you to Yosida pp.269-271 for the proof.

Theorem 11.7.2 [Trotter-Kato.] *Suppose that $\{\exp(tA_n)\}$ is an equibounded family of semi-groups as above, and suppose that for some z_0 with positive real part there exist an operator $R(z_0)$ such that*

$$\lim_{n\to\infty} R(z_0, A_n) \to R(z_0)$$

and

$$\operatorname{im} R(z_0) \quad \text{is dense in } \mathbf{F}.$$

Then there exists an equibounded semi-group $\exp(tA)$ *such that*

$$R(z_0) = R(z_0, A)$$

and

$$\exp(tA_n) \to \exp(tA)$$

uniformly on every compact interval of $t \geq 0$.

11.8   The Trotter product formula.

In what follows, **F** is a Banach space. Eventually we will restrict attention to a Hilbert space.

11.8.1   Chernoff's theorem.

Theorem 11.8.1 [Chernoff.] *Let* $f : [0, \infty) \to$ *bounded operators on* **F** *with*

$$\|f(t)\| \leq 1 \quad \forall \, t$$

and

$$f(0) = I.$$

Let A be a dissipative operator and $\exp tA$ *the contraction semi-group it generates. Let* **D** *be a core of A. Suppose that*

$$\lim_{h \searrow 0} \frac{1}{h}[f(h) - I]x = Ax \quad \forall \, x \in \mathbf{D}.$$

Then for all $y \in \mathbf{F}$

$$\lim \left[f\left(\frac{t}{n}\right) \right]^n y = (\exp tA)y \tag{11.26}$$

uniformly in any compact interval of $t \geq 0$.

Proof.   For fixed $t > 0$ let

$$C_n := \frac{n}{t}\left[f\left(\frac{t}{n}\right) - I \right].$$

Then $\frac{t}{n}C_n$ generates a contraction semi-group by the special case of the Lumer-Phillips theorem discussed in Section 11.6.2, and therefore (by change of variable), so does C_n. So C_n is the generator of a semi-group

$$\exp tC_n$$

and the hypothesis of the theorem is that $C_n x \to Ax$ for $x \in \mathbf{D}$. Hence by the limit theorem in the preceding section

$$(\exp tC_n)y \to (\exp tA)y$$

for each $y \in \mathbf{F}$ uniformly on any compact interval of t. Now

$$\exp(tC_n) = \exp n \left[f\left(\frac{t}{n}\right) - I \right]$$

so we may apply (11.22) to conclude that

$$\left\| \left(\exp(tC_n) - f\left(\frac{t}{n}\right)^n \right) x \right\| \leq \sqrt{n} \left\| \left(f\left(\frac{t}{n}\right) - I \right) x \right\| = \frac{t}{\sqrt{n}} \left\| \frac{n}{t} \left(f\left(\frac{t}{n}\right) - I \right) x \right\|.$$

The expression inside the $\| \cdot \|$ on the right tends to Ax so the whole expression tends to zero. This proves (11.26) for all x in $\mathbf{D}$. But since $\mathbf{D}$ is dense in $\mathbf{F}$ and $f(t/n)$ and $\exp tA$ are bounded in norm by 1 if follows that (11.26) holds for all $y \in \mathbf{F}$. QED

11.8.2 The product formula.

Let A and B be the infinitesimal generators of the contraction semi-groups $P_t = \exp tA$ and $Q_t = \exp tB$ on the Banach space F. Then $A + B$ is only defined on $D(A) \cap D(B)$ and in general we know nothing about this intersection. However let us *assume* that $D(A) \cap D(B)$ is sufficiently large that the closure $\overline{A + B}$ is a densely defined operator and $\overline{A + B}$ is in fact the generator of a contraction semi-group R_t. So $\mathbf{D} := D(A) \cap D(B)$ is a core for $\overline{A + B}$.

Theorem 11.8.2 [Trotter.] *Under the above hypotheses*

$$R_t y = \lim \left(P_{\frac{t}{n}} Q_{\frac{t}{n}} \right)^n y \quad \forall \ y \in \mathbf{F} \tag{11.27}$$

uniformly on any compact interval of $t \geq 0$.

Proof. Define

$$f(t) = P_t Q_t.$$

For $x \in D$ we have

$$f(t)x = P_t(I + tB + o(t))x = (I + At + Bt + o(t))x$$

so the hypotheses of Chernoff's theorem are satisfied. The conclusion of Chernoff's theorem asserts (11.27). QED

A symmetric operator on a Hilbert space is called **essentially self adjoint** if its closure is self-adjoint. So a reformulation of the preceding theorem in the case of self-adjoint operators on a Hilbert space says

Theorem 11.8.3 *Suppose that S and T are self-adjoint operators on a Hilbert space H and suppose that $S + T$ (defined on $D(S) \cap D(T)$) is essentially self-adjoint. Then for every $y \in H$*

$$\exp(it((\overline{S + T}))y = \lim_{n \to \infty} \left(\exp(\frac{t}{n}iA)(\exp \frac{t}{n}iB) \right)^n y \tag{11.28}$$

where the convergence is uniform on any compact interval of t.

11.8.3   Commutators.

An operator A on a Hilbert space is called skew-symmetric if $A^* = -A$ on $D(A)$. This is the same as saying that iA is symmetric. So we call an operator skew adjoint if iA is self-adjoint. We call an operator A **essentially skew adjoint** if iA is essentially self-adjoint.

If A and B are bounded skew adjoint operators then their Lie bracket

$$[A, B] := AB - BA$$

is well defined and again skew adjoint.

In general, we can only define the Lie bracket on $D(AB) \cap D(BA)$ so we again must make some rather stringent hypotheses in stating the following theorem.

Theorem 11.8.4 *Let A and B be skew adjoint operators on a Hilbert space H and let*

$$\mathbf{D} := D(A^2) \cap D(B^2) \cap D(AB) \cap D(BA).$$

Suppose that the restriction of $[A, B]$ to $\mathbf{D}$ is essentially skew-adjoint. Then for every $y \in \mathbf{H}$

$$\exp t\overline{[A, B]}y = \lim_{n \to \infty} \left((\exp -\sqrt{\frac{t}{n}}A)(\exp -\sqrt{\frac{t}{n}}B)(\exp \sqrt{\frac{t}{n}}A)(\exp \sqrt{\frac{t}{n}}B) \right)^n y \tag{11.29}$$

uniformly in any compact interval of $t \geq 0$.

Proof. The restriction of $[A, B]$ to $\mathbf{D}$ is assumed to be essentially skew-adjoint, so $[A, B]$ itself (which has the same closure) is also essentially skew adjoint.

We have

$$\exp(tA)x = (I + tA + \frac{t^2}{2}A^2)x + o(t^2)$$

for $x \in D$ with similar formulas for $\exp(-tA)$ etc.

Let

$$f(t) := (\exp -tA)(\exp -tB)(\exp tA)(\exp tB).$$

Multiplying out $f(t)x$ for $x \in D$ gives a whole lot of cancellations and yields

$$f(s)x = (I + s^2[A, B])x + o(s^2)$$

so (11.29) is a consequence of Chernoff's theorem with $s = \sqrt{t}$. QED

We still need to develop some methods which allow us to check the hypotheses of the last three theorems.

11.8.4   The Kato-Rellich theorem.

This is the starting point of a class of theorems which asserts that that if A is self-adjoint and if B is a symmetric operator which is "small" in comparison to A then $A + B$ is self adjoint.

Theorem 11.8.5 [Kato-Rellich.] *Let A be a self-adjoint operator and B a symmetric operator with*

$$D(B) \supset D(A)$$

and

$$\|Bx\| \le a\|Ax\| + b\|x\| \quad 0 \le a < 1, \quad \forall\, x \in D(A).$$

Then $A + B$ is self-adjoint, and is essentially self-adjoint on any core of A.

Proof. [Following Reed and Simon II page 162.] To prove that $A + B$ is self adjoint, it is enough to prove that $\text{im}(A + B \pm i\mu_0) = H$. We do this for $A + B + i\mu_0$. The proof for $A + B - i\mu_0$ is identical.

Let $\mu > 0$. Since A is self-adjoint, we know that

$$\|(A + i\mu)x\|^2 = \|Ax\|^2 + \mu^2\|x\|^2$$

from which we concluded that $(A + i\mu)^{-1}$ maps $\mathbf{H}$ onto $D(A)$ and

$$\|(A + i\mu)^{-1}\| \le \frac{1}{\mu}, \quad \|A(A + i\mu)^{-1}\| \le 1.$$

Applying the hypothesis of the theorem to $x = (A + i\mu)^{-1}y$ we conclude that

$$\|B(A + i\mu)^{-1}y\| \le a\|A((A + i\mu)^{-1}y\| + b\|(A + i\mu)^{-1}y\| \le \left(a + \frac{b}{\mu}\right)\|y\|.$$

Thus for $\mu \gg 1$, the operator

$$C := B(A + i\mu)^{-1}$$

satisfies

$$\|C\| < 1$$

since $a < 1$. Thus $-1 \notin \text{Spec}(A)$ so $\text{im}(I + C) = \mathbf{H}$. We know that $\text{im}(A + \mu I) = \mathbf{H}$ hence

$$\mathbf{H} = \text{im}(I + C) \circ (A + i\mu I) = \text{im}(A + B + i\mu I)$$

proving that $A + B$ is self-adjoint.

If $\mathbf{D}$ is any core for A, it follows immediately from the inequality in the hypothesis of the theorem that the closure of $A + B$ restricted to $\mathbf{D}$ contains $D(A)$ in its domain. Thus $A + B$ is essentially self-adjoint on any core of A. QED

11.8.5   Feynman path integrals.

Consider the operator

$$H_0 : \quad L_2(\mathbf{R}^3) \to L_2(\mathbf{R}^3)$$

given by

$$H_0 := -\left(\frac{\partial^2}{\partial x_1^2} + \frac{\partial^2}{\partial x_2^2} + \frac{\partial^2}{\partial x_3^2}\right).$$

Here the domain of H_0 is taken to be those $\phi \in L_2(\mathbf{R}^3)$ for which the differential operator on the right, taken in the distributional sense, when applied to ϕ gives an element of $L_2(\mathbf{R}^3)$.

The operator H_0 is called the "free Hamiltonian of non-relativistic quantum mechanics". The Fourier transform $\mathcal{F}$ is a unitary isomorphism of $L_2(\mathbf{R}^3)$ into $L_2(\mathbf{R}_3)$ and carries H_0 into multiplication by ξ^2 whose domain consists of those $\hat{\phi} \in L_2(\mathbf{R}_3)$ such that $\xi^2 \hat{\phi}(\xi)$ belongs to $L_2(\mathbf{R}_3)$. The operator consisting of multiplication by $e^{-it\xi^2}$ is clearly unitary, and provides us with a unitary one parameter group. Transferring this one parameter group back to $L_2(\mathbf{R}^3)$ via the Fourier transform gives us a one parameter group of unitary transformations whose infinitesimal generator is $-iH_0$.

Now the Fourier transform carries multiplication into convolution, and the inverse Fourier transform (in the distributional sense) of $e^{-i\xi^2 t}$ is $(2it)^{-3/2} e^{ix^2/4t}$. Hence we can write, in a formal sense,

$$(\exp(-itH_0)f)(x) = (4\pi it)^{-3/2} \int_{\mathbf{R}^3} \exp\left(\frac{i(x-y)^2}{4t} \right) f(y) dy.$$

Here the right hand side is to be understood as a long winded way of writing the left hand side which is well defined as a mathematical object. The right hand side can also be regarded as an actual integral for certain classes of f, and as the L_2 limit of such such integrals. We shall discuss this interpretation in Section 11.10.

Let V be a function on $\mathbf{R}^3$. We denote the operator on $L_2(\mathbf{R}^3)$ consisting of multiplication by V also by V. Suppose that V is such that $H_0 + V$ is again self-adjoint. For example, if V were continuous and of compact support this would certainly be the case by the Kato-Rellich theorem. (Realistic "potentials" V will not be of compact support or be bounded, but nevertheless in many important cases the Kato-Rellich theorem does apply.)

Then the Trotter product formula says that

$$\exp -it(H_0 + V) = \lim_{n \to \infty} \left(\exp(-i\frac{t}{n}H_0)(\exp -i\frac{t}{n}V) \right)^n.$$

We have

$$\left((\exp -i\frac{t}{n}V)f \right)(x) = e^{-i\frac{t}{n}V(x)} f(x).$$

Hence we can write the expression under the limit sign in the Trotter product formula, when applied to f and evaluated at x_0 as the following formal expression:

$$\left(\frac{4\pi it}{n} \right)^{-3n/2} \int_{\mathbf{R}^3} \cdots \int_{\mathbf{R}^3} \exp(iS_n(x_0, \ldots, x_n)) f(x_n) dx_n \cdots dx_1$$

where

$$S_n(x_0, x_1, \ldots, x_n, t) := \sum_{i=1}^{n} \frac{t}{n} \left[\frac{1}{4} \left(\frac{(x_i - x_{i-1})}{t/n} \right)^2 - V(x_i) \right].$$

If $X : s \mapsto X(s)$, $0 \le s \le t$ is a piecewise differentiable curve, then the **action** of a particle of mass m moving along this curve is defined in classical mechanics as

$$S(X) := \int_0^t \left(\frac{m}{2}\dot{X}(s)^2 - V(X(s))\right) ds$$

where $\dot{X}$ is the velocity (defined at all but finitely many points).

Take $m = \frac{1}{2}$ and let X be the polygonal path which goes from x_0 to x_1, from x_1 to x_2 etc., each in time t/n so that the velocity is $|x_i - x_{i-1}|/(t/n)$ on the i-th segment. Also, the integral of $V(X(s))$ over this segment is approximately $\frac{t}{n}V(x_i)$. The formal expression written above for the Trotter product formula can be thought of as an integral over polygonal paths (with step length t/n) of $e^{iS_n(X)}f(X(t))d_n X$ where S_n approximates the classical action and where $d_n X$ is a measure on this space of polygonal paths.

This suggests that an intuitive way of thinking about the Trotter product formula in this context is to imagine that there is some kind of "measure" dX on the space Ω_{x_0} of *all* continuous paths emanating from x_0 and such that

$$\exp(-it(H_0 + V)f)(x) = \int_{\Omega_{x_0}} e^{iS(X)}f(X(t))dX.$$

This formula was suggested in 1942 by Feynman in his thesis (Trotter's paper was in 1959), and has been the basis of an enormous number of important calculations in physics, many of which have given rise to exciting mathematical theorems which were then proved by other means. I am unaware of any general mathematical justification of these "path integral" methods in the form that they are used.

11.9 The Feynman-Kac formula.

An important advance was introduced by Mark Kac in 1951 where the unitary group $\exp -it(H_0 + V)$ is replaced by the contraction semi-group $\exp -t(H_0 + V)$. Then the techniques of probability theory (in particular the existence of Wiener measure on the space of continuous paths) can be brought to bear to justify a formula for the contractive semi-group as an integral over path space. I will state and prove an elementary version of this formula which follows directly from what we have done. The assumptions about the potential are physically unrealistic, but I choose to regard the extension to a more realistic potential as a technical issue rather than a conceptual one.

Let V be a continuous real valued function of compact support. To each continuous path ω on $\mathbf{R}^n$ and for each fixed time $t \ge 0$ we can consider the integral

$$\int_0^t V(\omega(s))ds.$$

The map

$$\omega \mapsto \int_0^t V(\omega(s))ds \tag{11.30}$$

is a continuous function on the space of continuous paths, and we have

$$\frac{t}{m}\sum_{j=1}^{m} V\left(\omega\left(\frac{jt}{m}\right)\right) \to \int_0^t V(\omega(s))ds \qquad (11.31)$$

for each fixed ω.

Theorem 11.9.1 The Feynman-Kac formula. *Let V be a continuous real valued function of compact support on $\mathbf{R}^n$. Let*

$$H = \Delta + V$$

as an operator on $\mathbf{H} = L^2(\mathbf{R}^n)$. Then H is self-adjoint and for every $f \in \mathbf{H}$

$$\left(e^{-tH}f\right)(x) = \int_{\Omega_x} f(\omega(t)) \exp\left(\int_0^t V(\omega(s))ds\right) d_x\omega \qquad (11.32)$$

where Ω_x is the space of continuous paths emanating from x and $d_x\omega$ is the associated Wiener measure.

Proof. [From Reed-Simon II page 280.] Since multiplication by V is a bounded self-adjoint operator, we can apply the Kato-Rellich theorem (with $a = 0$!) to conclude that H is self-adjoint, and with the same domain as Δ. So we may apply the Trotter product formula to conclude that

$$(e^{-Ht})f = \lim_{m\to\infty}\left(e^{-\frac{t}{m}\Delta}e^{-\frac{t}{m}V}\right)^m f.$$

This convergence is in L^2, but by passing to a subsequence we may also assume that the convergence is almost everywhere. Now

$$\left[\left(e^{-\frac{t}{m}\Delta}e^{-\frac{t}{m}V}\right)^m f\right](x)$$

$$= \int_{\mathbf{R}^n}\cdots\int_{\mathbf{R}^n} p\left(x,x_m,\frac{t}{m}\right)\cdots p\left(x,x_m,\frac{t}{m}\right) f(x)_1 \exp\left(-\sum_{j=1}^{m}\frac{t}{m}V(x_j)\right) dx_1\cdots dx_m.$$

By the very definition of Wiener measure, this last expression is

$$\int_{\Omega_x} \exp\left(\frac{t}{m}\sum_{j=1}^{m} V\left(\omega\left(\frac{jt}{m}\right)\right)\right) f(\omega(t))d_x\omega.$$

The integrand (with respect to the Wiener measure $d_x\omega$) converges on all continuous paths, that is to say almost everywhere with respect to $d_x\mu$ to the integrand on right hand side of (11.32). So to justify (11.32) we must prove that the integral of the limit is the limit of the integral. We will do this by the dominated convergence theorem:

$$\int_{\Omega_x} \left|\exp\left(\frac{t}{m}\sum_{j=1}^{m} V\left(\omega\left(\frac{jt}{m}\right)\right)\right) f(\omega(t))\right| d_x\omega$$

$$\leq e^{t\,\max|V|} \int_{\Omega_x} |f(\omega(t))| d_x \omega = e^{t\,\max|V|} \left(e^{-t\Delta} |f| \right)(x) < \infty$$

for almost all x. Hence, by the dominated convergence theorem, (11.32) holds for almost all x. QED

11.10 The free Hamiltonian and the Yukawa potential.

In this section I want to discuss the following circle of ideas. Consider the operator

$$H_0 : \quad L_2(\mathbf{R}^3) \to L_2(\mathbf{R}^3)$$

given by

$$H_0 := -\left(\frac{\partial^2}{\partial x_1^2} + \frac{\partial^2}{\partial x_2^2} + \frac{\partial^2}{\partial x_3^2} \right).$$

Here the domain of H_0 is taken to be those $\phi \in L_2(\mathbf{R}^3)$ for which the differential operator on the right, taken in the distributional sense, when applied to ϕ gives an element of $L_2(\mathbf{R}^3)$.

The operator H_0 has a fancy name. It is called the "free Hamiltonian of non-relativistic quantum mechanics". Strictly speaking we should add "for particles of mass one in units where Planck's constant is one".

The Fourier transform is a unitary isomorphism of $L_2(\mathbf{R}^3)$ into $L_2(\mathbf{R}_3)$ and carries H_0 into multiplication by ξ^2 whose domain consists of those $\hat{\phi} \in L_2(\mathbf{R}_3)$ such that $\xi^2 \hat{\phi}(\xi)$ belongs to $L_2(\mathbf{R}_3)$. The operators

$$V(t) : L_2(\mathbf{R}_3) \to L_2(\mathbf{R}_3), \quad \hat{\phi}(\xi) \mapsto e^{-it\xi^2} \hat{\phi}$$

form a one parameter group of unitary transformations whose infinitesimal generator in the sense of Stone's theorem is operator consisting of multiplication by ξ^2 with domain as given above. [The minus sign before the i in the exponential is the convention used in quantum mechanics. So we write $\exp -itA$ for the one-parameter group associated to the self-adjoint operator A. I apologize for this (rather irrelevant) notational change, but I want to make the notation in this section consistent with what you will see in physics books.]

Thus the operator of multiplication by ξ^2, and hence the operator H_0 is a self-adjoint transformation. The operator of multiplication by ξ^2 is clearly non-negative and so every point on the negative real axis belongs to its resolvent set. Let us write a point on the negative real axis as $-\mu^2$ where $\mu > 0$. Then the resolvent is given by multiplication by $-f$ where

$$f(\xi) = f_\mu(\xi) := \frac{1}{\mu^2 + \xi^2}.$$

We can summarize what we "know" so far as follows:

1. The operator H_0 is self adjoint.

2. The one parameter group of unitary transformations it generates via Stone's theorem is

$$U(t) = \mathcal{F}^{-1}V(t)\mathcal{F}$$

   where $V(t)$ is multiplication by $e^{-it\xi^2}$.

3. Any point $-\mu^2$, $\mu > 0$ lies in the resolvent set of H_0 and

$$R(-\mu^2, H_0) = -\mathcal{F}^{-1}m_f\mathcal{F}$$

   where m_f denotes the operation of multiplication by f and f is as given above.

4. If $g \in \mathcal{S}$ and m_g denotes multiplication by g, then the the operator $\mathcal{F}^{-1}m_g\mathcal{F}$ consists of convolution by $\check{g}$. Neither the function $e^{-it\xi^2}$ nor the function f belongs to $\mathcal{S}$, so the operators $U(t)$ and $R(-\mu^2, H_0)$ can only be thought of as convolutions in the sense of generalized functions.

11.10.1   The Yukawa potential and the resolvent.

Nevertheless, we will be able to give some slightly more explicit (and very instructive) representations of these operators as convolutions. For example, we will use the Cauchy residue calculus to compute $\check{f}$ and we will find, up to factors of powers of 2π that $\check{f}$ is the function

$$Y_\mu(x) := \frac{e^{-\mu r}}{r}$$

where r denotes the distance from the origin, i.e. $r^2 = x^2$. This function has an integrable singularity at the origin, and vanishes rapidly at infinity. So convolution by Y_μ will be well defined and given by the usual formula on elements of $\mathcal{S}$ and extends to an operator on $L_2(\mathbf{R}^3)$.

   The function Y_μ is known as the **Yukawa potential**. Yukawa introduced this function in 1934 to explain the forces that hold the nucleus together. The exponential decay with distance contrasts with that of the ordinary electromagnetic or gravitational potential $1/r$ and, in Yukawa's theory, accounts for the fact that the nuclear forces are short range. In fact, Yukawa introduced a "heavy boson" to account for the nuclear forces. The role of mesons in nuclear physics was predicted by brilliant theoretical speculation well before any experimental discovery. Here are the details:

   Since $f \in L_2$ we can compute its inverse Fourier transform as

$$(2\pi)^{-3/2}\check{f} = \lim_{R \to \infty} (2\pi)^{-3} \int_{|\xi| \le R} \frac{e^{i\xi \cdot x}}{\mu^2 + \xi^2} d\xi. \tag{11.33}$$

Here lim means the L_2 limit and $|\xi|$ denotes the length of the vector ξ, i.e. $|\xi| = \sqrt{\xi^2}$ and we will use similar notation $|x| = r$ for the length of x. Assume $x \ne 0$. Let

$$u := \frac{\xi \cdot x}{|\xi||x|}$$

$$-R + i\sqrt{R} \qquad\qquad\qquad R + i\sqrt{R}$$

$$-R \qquad\qquad\qquad 0 \qquad\qquad\qquad R$$
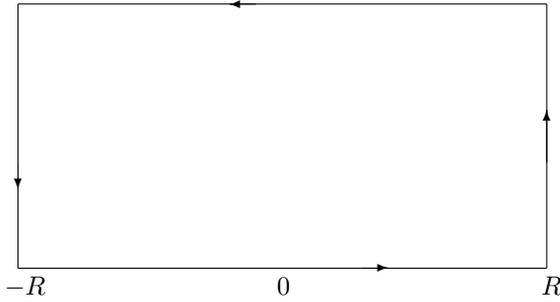
so u is the cosine of the angle between x and ξ. Fix x and introduce spherical coordinates in ξ space with x at the north pole and $s = |\xi|$ so that

$$(2\pi)^{-3} \int_{|\xi|\le R} \frac{e^{i\xi\cdot x}}{\mu^2 + \xi^2} d\xi = (2\pi)^{-2} \int_0^R \int_{-1}^1 \frac{e^{is|x|u}}{s^2 + \mu^2} s^2 du\, ds$$

$$=$$

$$\frac{1}{(2\pi)^2 i|x|} \int_{-R}^R \frac{s e^{is|x|}}{(s + i\mu)(s - i\mu)} ds.$$

This last integral is along the bottom of the path in the complex s-plane consisting of the boundary of the rectangle as drawn in the figure.

On the two vertical sides of the rectangle, the integrand is bounded by some constant time $1/R$, so the contribution of the vertical sides is $O(1/\sqrt{R})$. On the top the integrand is $O(e^{-\sqrt{R}})$. So the limits of these integrals are zero. There is only one pole in the upper half plane at $s = i\mu$, so the integral is given by $2\pi i\times$ this residue which equals

$$2\pi i \frac{i\mu e^{-\mu|x|}}{2i\mu} = \pi i e^{-\mu|x|}.$$

Inserting this back into (11.33) we see that the limit exists and is equal to

$$(2\pi)^{-3/2} \hat{f} = \frac{1}{4\pi} \frac{e^{-\mu|x|}}{|x|}.$$

We conclude that for $\phi \in \mathcal{S}$

$$[(H_0 + \mu^2)^{-1}\phi](x) = \frac{1}{4\pi} \int_{\mathbf{R}^3} \frac{e^{-\mu|x-y|}}{|x - y|} \phi(y) dy,$$

and since $(H_0 + \mu^2)^{-1}$ is a bounded operator on L_2 this formula extends in the L_2 sense to L_2.

11.10.2   The time evolution of the free Hamiltonian.

The "explicit" calculation of the operator $U(t)$ is slightly more tricky. The function $\xi \mapsto e^{-it\xi^2}$ is an "imaginary Gaussian", so we expect is inverse Fourier

transform to also be an imaginary Gaussian, and then we would have to make sense of convolution by a function which has absolute value one at all points. There are several ways to proceed. One involves integration by parts, and I hope to explain how this works later on in the course in conjunction with the method of stationary phase.

Here I will follow Reed-Simon vol II p.59 and add a little positive term to t and then pass to the limit. In other words, let α be a complex number with positive real part and consider the function

$$\xi \mapsto e^{-\xi^2 \alpha}$$

This function belongs to $\mathcal{S}$ and its inverse Fourier transform is given by the function

$$x \mapsto (2\alpha)^{-3/2} e^{-x^2/4\alpha}.$$

(In fact, we verified this when α is real, but the integral defining the inverse Fourier transform converges in the entire half plane $\mathrm{Re}\,\alpha > 0$ uniformly in any $\mathrm{Re}\,\alpha > \epsilon$ and so is holomorphic in the right half plane. So the formula for real positive α implies the formula for α in the half plane.)

We thus have

$$(e^{-H_0\alpha}\phi)(x) = \left(\frac{1}{4\pi\alpha}\right)^{3/2} \int_{\mathbf{R}^3} e^{-|x-y|^2/4\alpha}\phi(y)dy.$$

Here the square root in the coefficient in front of the integral is obtained by continuation from the positive square root on the positive axis. For example, if we take $\alpha = \epsilon + it$ so that $-\alpha = -i(t - i\epsilon)$ we get

$$(U(t)\phi)(x) = \lim_{\epsilon \searrow 0}(U(t-i\epsilon)\phi)(x) = \lim_{\epsilon \searrow 0}(4\pi i(t-i\epsilon))^{-\frac{3}{2}} \int e^{-|x-y|^2/4i(t-i\epsilon)}\phi(y)dy.$$

Here the limit is in the sense of L_2. We thus could write

$$(U(t))(\phi)(x) = (4\pi i)^{-3/2} \int e^{i|x-y|^2/4t}\phi(y)dy$$

if we understand the right hand side to mean the $\epsilon \searrow 0$ limit of the preceding expression.

Actually, as Reed and Simon point out, if $\phi \in L_1$ the above integral exists for any $t \neq 0$, so if $\phi \in L_1 \cap L_2$ we should expect that the above integral is indeed the expression for $U(t)\phi$. Here is their argument: We know that

$$\exp(-i(t - i\epsilon))\phi \to U(t)\phi$$

in the sense of L_2 convergence as $\epsilon \searrow 0$. Here we use a theorem from measure theory which says that if you have an L_2 convergent sequence you can choose a subsequence which also converges pointwise almost everywhere. So choose a subsequence of ϵ for which this happens. But then the dominated convergence

theorem kicks in to guarantee that the integral of the limit is the limit of the integrals.

To sum up: The function

$$P_0(x, y; t) := (4\pi it)^{-3/2} e^{i|x-y|/4t}$$

is called the **free propagator**. For $\phi \in L_1 \cap L_2$

$$[U(t)\phi](x) = \int_{\mathbf{R}^3} P_0(x, y; t)\phi(y)dy$$

and the integral converges. For general elements ψ of L_2 the operator $U(t)\psi$ is obtained by taking the L_2 limit of the above expression for any sequence of elements of $L_1 \cap L_2$ which approximate ψ in L_2. Alternatively, we could interpret the above integral as the $\epsilon \searrow$ limit of the corresponding expression with t replaced by $t - i\epsilon$.