Class Notes for MATH 355.

by
S. W. Drury

Contents

Starred chapters and sections should be omitted on a first reading. Double starred sections should be omitted on a second reading.

1

Measuring sets on the Line

In this chapter we look at the question of how to assign a "length" to a subset of $\mathbb{R}$. It's fairly clear that this is something that might be desirable to do. The motivation comes from the desire to define the Lebesgue Integral. The Riemann integral is defined by making a "vertical" decomposition of the space on which the function is defined. The advantage of doing this is that the sets of the decomposition are intervals and it is easy to decide what the length of an interval is. Lebesgue himself mentioned the situation of someone collecting money and wanting to discover what the day's takings are. There are two ways of doing this. The first way, is to keep a running total of the takings at each point in time. So, if at some point we receive a quarter and then a dime, we first add 25 cents and then 5 cents. The second approach is to collect all the nickels together, all the dimes together and all the quarters together and at the end of the day find out how many of each there are. This corresponds to a "horizontal" decomposition in defining an integral. The horizontal method turns out to have distinct advantages over the vertical, but there is one initial problem to be tackled. The horizontal decomposition of a function will lead to sets of the form

$$\{x; y_1 \leq f(x) < y_2\}$$

and we will need to decide what the length of this set is. So, we will need a theory of length for rather general subsets of the real line.

Another major impetus for developing so called measure theory is the theory of probability. Here we have a space Ω called the sample space. A point in Ω typically represents a particular outcome of an experiment. Usually we are not interested in individual outcomes, but rather in sets of outcomes that satisfy some criterion. Such a set is called an "event". Events are assigned a probability of occurring. This is like assigning a length to a subset of $[0,1]$. The situation is only slightly special in that the probability of the event of all possible outcomes Ω is necessarily equal to unity. We then talk of random variables which are functions defined on the sample space and their integrals with respect to the given probability measure defines the expectation of the random variable.

Measure theory is necessarily a complicated subject because, in many situations, it turns out to be impossible to assign a sensible length to every subset. A telling example is due to Banach and Tarski. It relates to paradoxical decompositions. Let a group G act on a space X. The action is ***paradoxical*** if for positive integers m and n, there are disjoint subsets $A_1, A_2, \ldots, A_m, B_1, B_2, \ldots, B_n$ of X and elements $g_1, \ldots, g_m, h_1, \ldots, h_n$ of G such that

$$X = \bigcup_{j=1}^{m} g_j \cdot A_j$$

$$X = \bigcup_{k=1}^{n} h_k \cdot B_k$$

The paradox is that if μ is some kind of G-invariant measure which applies to all subsets of X, then we will be forced to have (because of the disjointness) that

$$\mu(X) \geq \sum_{j=1}^{m} \mu(A_j) + \sum_{k=1}^{n} \mu(B_k)$$

and yet

$$\mu(X) \leq \sum_{j=1}^{m} \mu(g_j \cdot A_j) = \sum_{j=1}^{m} \mu(A_j)$$

$$\mu(X) \leq \sum_{k=1}^{n} \mu(h_k \cdot B_k) = \sum_{k=1}^{n} \mu(B_k)$$

leading to $2\mu(X) = \mu(X) + \mu(X) \leq \sum_{j=1}^{m} \mu(A_j) + \sum_{k=1}^{n} \mu(B_k) \leq \mu(X)$.

A weak form of the Banach–Tarski paradox states that the action of the rotation group on the sphere in Euclidean 3-space is paradoxical. One is forced to conclude that if one wants a viable theory, then it will only be possible to measure nice sets. Usually, there will be nasty sets that have pathological properties.

1.1 Fields and σ-fields

This means that we need to look at collections of subsets with certain properties. We will meet various types of such collections in this course. The most prevalent one will be the σ-field . Our main focus will be the real line. Before defining σ-fields we will establish the following result which will be the starting point for Lebesgue measure.

THEOREM 1 *Let K be a countable index set. $I = [a, b[$ and $I_k = [a_k, b_k[$ for $k \in K$ We have*

(i) *If $\bigcup_k I_k \subseteq I$ and the I_k are disjoint, then $\sum_k \mathrm{length(I_k)} \leq \mathrm{length(I)}$.*

(ii) *If $I \subseteq \bigcup_k I_k$, then $\mathrm{length(I)} \leq \sum_k \mathrm{length(I_k)}$.*

(iii) *If $\bigcup_k I_k = I$ and the I_k are disjoint, then $\sum_k \mathrm{length(I_k)} = \mathrm{length(I)}$.*

Proof.

(i) Let us assume that K is finite to start with. Then we can assume that $K = \{1, 2, \ldots, n\}$. We will proceed by induction on n. If $n = 1$ then $I_1 \subseteq I$ clearly implies that $a_1 \geq a$ and $b_1 \leq b$ so $\mathrm{length(I_1)} = b_1 - a_1 \leq b - a = \mathrm{length(I)}$. The induction starts.

Now let us assume that the result is true for $n - 1$ intervals. Let us reorder the intervals such that a_k are increasing with k. This affects neither the hypotheses nor the conclusion. Now for $1 \leq k < n$, $b_k \leq a_n$ for, if not then $a_n \in I_k$ since certainly $a_k \leq a_n$. Thus

$$\bigcup_{k=1}^{n-1} I_k \subseteq I \cap]-\infty, a_n[= [a, a_n[.$$

Applying the induction hypothesis, this gives $\sum_{k=1}^{n-1} \mathrm{length(I_k)} \leq a_n - a$. But, we also have that $b_n \leq b$, for if not, there is a point of I_n close to b_n which is not in I. So, $\mathrm{length(I_n)} = b_n - a_n \leq b - a_n$. It follows that $\sum_{k=1}^{n} \mathrm{length(I_k)} \leq (a_n - a) + (b - a_n) = b - a$, completing the induction step.

To establish the result when K is infinite, it suffices to assume without loss of generality that $K = \mathbb{N}$ and to let n tend to infinity in the finite case.

(ii) Let us assume that K is finite to start with. Then we can assume that $K = \{1, 2, \ldots, n\}$. We will proceed by induction on n. If $n = 1$ then the induction starts as before. Now let us assume that the result is true for $n-1$ intervals. Let us reorder the intervals such that a_k are increasing with k. Again, this affects neither the hypotheses nor the conclusion. First we consider the case where $a \geq b_1$. Then $I \cap I_1 = \emptyset$ and already $I \subseteq \bigcup_{k=2}^n I_k$ so that by the induction hypothesis $(b-a) \leq \sum_{k=2}^n (b_k - a_k) \leq \sum_{k=1}^n (b_k - a_k)$ as required. So, we may always assume that $a < b_1$. Now observe that

$$[b_1, b[\subseteq \bigcup_{k=2}^n I_k$$

since if $x \in [b_1, b[$, then $x \in [a, b[\subseteq \bigcup_{k=1}^n I_k$, but $x \notin I_1$. Therefore by the induction hypothesis $b - b_1 \leq \sum_{k=2}^n (b_k - a_k)$. But $a \geq a_1$ for otherwise $a \notin \bigcup_{k=1}^n I_k$. So we get

$$b - a = (b_1 - a) + (b - b_1) \leq (b_1 - a_1) + \sum_{k=2}^n (b_k - a_k) = \sum_{k=1}^n (b_k - a_k)$$

as required.

Now for the case K infinite. We can assume that $k = \mathbb{N}$. We want to use compactness to reduce to the case of finitely many intervals. But this won't work directly, so we want to make the contained interval closed and bounded and the containing intervals open. Let $\epsilon > 0$. Then we have

$$[a, b - \epsilon] \subseteq \bigcup_{k=1}^\infty]a_k - 2^{-k}\epsilon, b_k[$$

and by compactness, there exists an integer n such that

$$[a, b - \epsilon] \subseteq \bigcup_{k=1}^n]a_k - 2^{-k}\epsilon, b_k[$$

and

$$[a, b - \epsilon[\subseteq \bigcup_{k=1}^n [a_k - 2^{-k}\epsilon, b_k[$$

Therefore, from the finite result, we have

$$b - a - \epsilon \leq \sum_{k=1}^n (b_k - a_k + 2^{-k}\epsilon) \leq \epsilon + \sum_{k=1}^\infty (b_k - a_k)$$

Since ϵ is an arbitrary positive number, the result follows.

(iii) Follows immediately from (i) and (ii) above. ■

In Theorem 1 above, we had in mind that the endpoints of the intervals should be real numbers. We can also ask what happens if we allow either I of I_k to take either of the forms $]-\infty, b[$, $[a, \infty[$ or $]-\infty, \infty[$.

COROLLARY 2 *Let K be a countable index set. I and I_k for $k \in K$ be general intervals, closed on the left and open on the right. If $\bigcup_k I_k = I$ and the I_k are disjoint, then* $\sum_k \mathrm{length(I_k)} = \mathrm{length(I)}$.

Proof. The only problem is when one or more of the intervals has infinite length. Obviously, if one of the I_k has infinite length, so does I and $\sum_k \mathrm{length(I_k)} = \infty = \mathrm{length(I)}$. The only contentious case is when I has infinite length, but all the I_k have finite length. In this case, let $c > 0$. We have by Theorem 1, (ii) that $\mathrm{length(I \cap [-c, c[)} \leq \sum_{k \in K} \mathrm{length(I_k)}$. We find that $\mathrm{length(I \cap [-c, c[)} \longrightarrow \infty$ as $c \longrightarrow \infty$ and it follows that $\sum_k \mathrm{length(I_k)} = \infty$ as required. ■

We can now make progress on general measure theory.

DEFINITION *Let X be a set. Then a collection $\mathcal{F}$ of subsets of X is a* ***field*** *(sometimes called an algebra) if and only if*

(i) $X \in \mathcal{F}$.

(ii) $A \in \mathcal{F} \Longrightarrow X \setminus A \in \mathcal{F}$.

(iii) $A \in \mathcal{F}, B \in \mathcal{F} \Longrightarrow A \cup B \in \mathcal{F}$.

The immediate consequences of this definition are:

- $\emptyset \in \mathcal{F}$.
- $A_k \in \mathcal{F}$ for $k \in K$, K finite $\Longrightarrow \bigcup_{k \in K} A_k \in \mathcal{F}$.
- $A_k \in \mathcal{F}$ for $k \in K$, K finite $\Longrightarrow \bigcap_{k \in K} A_k \in \mathcal{F}$.

In the same vein we have the following definition.

DEFINITION *Let X be a set. Then a collection $\mathcal{F}$ of subsets of X is a **σ-field** (sometimes called a **σ-algebra**) if and only if*

(i) $X \in \mathcal{F}$.

(ii) $A \in \mathcal{F} \Longrightarrow X \setminus A \in \mathcal{F}$.

(iii) $A_k \in \mathcal{F}$ *for* $k \in K$, K *countable* $\Longrightarrow \bigcup_{k\in K} A_k \in \mathcal{F}$.

The immediate consequences of this definition are:

- $\emptyset \in \mathcal{F}$.
- $A_k \in \mathcal{F}$ for $k \in K$, K countable $\Longrightarrow \bigcap_{k\in K} A_k \in \mathcal{F}$.
- $\mathcal{F}$ a σ-field $\Longrightarrow \mathcal{F}$ a field.

DEFINITION *We can now define the concept of a measure (sometimes called a countably additive set function) on a field $\mathcal{F}$ of subsets of X as a function $\mu : \mathcal{F} \longrightarrow [0, \infty]$ such that*

(i) $\mu(\emptyset) = 0$.

(ii) $\mu\left(\bigcup_{k\in K} A_k\right) = \sum_{k\in K} \mu(A_k)$ *whenever K is a countable index set and A_k are pairwise disjoint subsets of X with $A_k \in \mathcal{F}$ and $\bigcup_{k\in K} A_k \in \mathcal{F}$.*

Sometimes if $\mathcal{F}$ is a field rather than a σ-field , μ is called a ***premeasure*** rather than a measure.

It's worth observing explicitly that we are allowing measures to take the value ∞. We interpret sums of nonnegative series with possibly infinite terms in the obvious way. So, if just one term in the series $\sum_{k\in K} \mu(A_k)$ is infinite, the whole sum is infinite. If not, then the series is treated as a series of nonnegative terms and it evaluates to a real number if the series converges and to ∞ if the series diverges.

Now we need to look at complementation, because this will come to plague us later. If $A, B \in \mathcal{F}$ and $\mu(A) = \mu(B)$ where μ is a measure on $\mathcal{F}$, then can we deduce that $\mu(A^c) = \mu(B^c)$? We have of course $\mu(X) = \mu(A) + \mu(A^c)$ and $\mu(X) = \mu(B) + \mu(B^c)$, so with the normal laws of arithmetic we have $\mu(A^c) = \mu(B^c)$. Indeed, if $\mu(X) < \infty$ this is clearly the case because the measures of all the sets involved are nonnegative real numbers. But if $\mu(A) = \mu(B) = \mu(X) = \infty$ then nothing whatever can be said about $\mu(A^c)$ and $\mu(B^c)$.

There is another important property of measures.

LEMMA 3 *Let $\mathcal{F}$ be a σ-field and suppose that μ is a measure on $\mathcal{F}$. Let $F_j \in \mathcal{F}$ be an increasing sequence of sets, then $\mu\left(\bigcup_j F_j\right) = \sup_j \mu(F_j)$.*

Proof. We define $A_1 = F_1$, $A_j = F_j \setminus F_{j-1}$ for $j = 2, 3, \ldots$. Then A_j are disjoint subsets in $\mathcal{F}$. Therefore

$$\mu\left(\bigcup_{j=1}^{\infty} A_j\right) = \sum_{j=1}^{\infty} \mu(A_j) = \mu(F_1) + \sum_{j=2}^{\infty} \Big(\mu(F_j) - \mu(F_{j-1})\Big) = \sup_j \mu(F_j).$$

∎

That's a lot of definitions, so we had better have some examples.

- Let X any set and let $\mathcal{F}$ be the collection of all subsets of X. That will be a σ-field. For a measure we can simply let $\mu(A)$ be the number of elements in A with the understanding that $\mu(A) = \infty$ if A is infinite. It's intuitively clear that μ is a measure on $\mathcal{F}$, but that would require some proof. This measure is called the ***counting measure*** because it simply counts the number of elements in the set.

- Let $X = \mathbb{N}$ and again let $\mathcal{F}$ be the collection of all subsets of X. Assign a weight $w_n \geq 0$ to each $n \in \mathbb{N}$. Now define

$$\mu(A) = \sum_{n \in A} w_n$$

  with the understanding that $\mu(A) = \infty$ if the series diverges. The terms of a series of positive terms can be rearranged without affecting the convergence or the value of the sum (or we can work with unconditional sums — see the notes for MATH 255).

- Let $X = \mathbb{N}$ and let $\mathcal{F}$ be the collection of subsets of X that are either finite or cofinite This is a field, but not a σ-field. You can assign a premeasure to $\mathcal{F}$ as in the last example.

- Let X be an uncountable set and let $\mathcal{F}$ be the collection of subsets of X that are either finite or cofinite This is a field, but not a σ-field. Now, let $\mu(A) = 0$ if A is finite and $\mu(A) = 1$ if A is cofinite.

- Let X and $\mathcal{F}$ be as in the last example. Now, let $\mu(A) = 0$ if A is finite and $\mu(A) = \infty$ if A is cofinite.

- Let $X = \{1,2,3,4,5,6\}$, $\mathcal{F}$ the set of all subsets of X and $\mu(A) = |A|/6$. Then $\mu(A)$ is the probability measure of a fair dice.

DEFINITION *A* ***probability measure*** *is a measure with the additional property that* $\mu(X) = 1$.

It follows easily from the definitions that if μ is a probability measure then μ takes its values in the interval $[0,1]$.

It will be noted that we are quite short on interesting examples. We work to remedy that situation. It will also be noted that we allow a measure to be defined on a field rather than a σ-field which might seem (correctly) to be its natural base of operations. Allowing this possibility gives us room for manœvre.

1.2 Extending the notion of length — first steps

So in this section, we will let $X = \mathbb{R}$ and $\mathcal{F}$ is the collection of all subsets of R that are *finite* unions of intervals closed on the left and open on the right. There is no restriction on the length of the intervals except perhaps that we can always consider it to be strictly positive for otherwise the interval would be empty. Let $F \in \mathcal{F}$ and consider C a component of F. This is a connected subset of F and therefore also of $\mathbb{R}$. So, C is an interval. Now every constituent interval of F is contained in some component C and it follows that C is just the (finite) union of those constituent intervals which it contains. Thus, C is closed on the left and open on the right and it follows that distinct components are not merely disjoint, but also cannot abut. So every set $F \in \mathcal{F}$ can be written in a unique way as

$$F = \bigcup_{k=1}^{n} I_k$$

where I_k are intervals closed on the left and open on the right that are disjoint and do not abut. It is now easy to see that the $\mathbb{R} \setminus F$ is also an element of $\mathcal{F}$. This means that $\mathcal{F}$ is a field. Let us now define $\mu(F) = \sum_{k=1}^{n} \mathrm{length(I_k)}$.

THEOREM 4 *The set function* μ *is a premeasure on* $\mathcal{F}$.

Proof. Let $F_j \in \mathcal{F}$ be disjoint for $j \in \mathbb{N}$, $F \in \mathcal{F}$ and $F = \bigcup_j F_j$. Then we have to show that

$$\mu(F) = \sum_j \mu(F_j).$$

We will write $F_j = \bigcup_{k=1}^{K_j} I_{j,k}$ and $F = \bigcup_{k=1}^{K} J_k$ where $I_{j,k}$ and J_k are intervals closed on the left and open on the right. All of these unions are of the disjoint variety. We get

$$\mu(F) = \sum_{k=1}^{K} \mathrm{length(J_k)} \tag{1.1}$$

$$= \sum_{k=1}^{K} \sum_{j=1}^{\infty} \sum_{\ell=1}^{K_j} \mathrm{length(J_k \cap I_{j,\ell})} \tag{1.2}$$

$$= \sum_{j=1}^{\infty} \sum_{\ell=1}^{K_j} \sum_{k=1}^{K} \mathrm{length(J_k \cap I_{j,\ell})} \tag{1.3}$$

$$= \sum_{j=1}^{\infty} \sum_{\ell=1}^{K_j} \mathrm{length(I_{j,\ell})} \tag{1.4}$$

$$= \sum_{j=1}^{\infty} \mu(F_j) \tag{1.5}$$

Here (1.1) is the definition of $\mu(F)$, (1.2) follows by Corollary 2 applied to $J_k = \bigcup_{j,\ell} J_k \cap I_{j,\ell}$, (1.3) follows from changing the order of summation in a series of positive terms, (1.4) follows since $I_{j,\ell} \subseteq F$ and by using Corollary 2 applied to $I_{j,\ell} = \bigcup_{k=1}^{K} J_k \cap I_{j,\ell}$. Finally (1.5) follows by the definition of $\mu(F_j)$. ■

1.3 Fields and σ-fields generated by a family of sets

Let X be a set and $\mathcal{A}$ a family of subsets of X. We define the field and the σ-field generated by $\mathcal{A}$ by considering the collection of all fields (respectively σ-fields) on X which contain the given collection $\mathcal{A}$ and take the intersection of the collection. Explicitly

$$\text{field generated by } \mathcal{A} = \bigcap_{\substack{\mathcal{F} \text{ is a field} \\ \mathcal{F} \supseteq \mathcal{A}}} \mathcal{F}$$

$$\sigma\text{-field generated by } \mathcal{A} = \bigcap_{\substack{\mathcal{F} \text{ is a } \sigma\text{-field} \\ \mathcal{F} \supseteq \mathcal{A}}} \mathcal{F}$$

There are two important considerations here. The first is that the power set of X, (i.e. the collection of all subsets of X) is a field (respectively σ-field) on X. The second is that an arbitrary intersection of fields (respectively σ-fields) is again a field (respectively σ-field). We can legitimately say that the field (respectively σ-field) generated by $\mathcal{A}$ is the smallest field (respectively σ-field) on X containing $\mathcal{A}$.

This *definition from the outside* is very unappealing. It's really difficult to get a handle on what it means. In the case of a field, it is possible to give a definition from the inside but for the σ-field, this is unfortunately not the case. For fields we have the following lemma.

LEMMA 5 *Let X be a set and $\mathcal{A}$ any collection of subsets of X. Let $\mathcal{F}$ be the field generated by $\mathcal{A}$. Then a subset F of X lies in $\mathcal{F}$ if and only if the following condition holds:*

> *There is an integer $N \geq 1$ and a chain of sets $(F_n)_{n=1}^N$ defined for $n = 1, 2, \ldots, N$ by one of the following options:*
>
> - *$F_n = X$.*
> - *$F_n \in \mathcal{A}$,*
> - *$F_n = F_{p_n} \setminus F_{q_n}$ with $1 \leq p_n, q_n < n$.*
> - *$F_n = F_{p_n} \cup F_{q_n}$ with $1 \leq p_n, q_n < n$.*
>
> *and with $F = F_N$.*

Proof. One proves easily by induction that $F_n \in \mathcal{F}$. In the opposite direction, let the collection of all subsets that are defined by a chain of this type by $\mathcal{G}$. Then clearly $\mathcal{A} \subseteq \mathcal{G}$. We only need to show that $\mathcal{G}$ is a field. We have $X \in \mathcal{G}$. Now let $G, H \in \mathcal{G}$. Then there are chains of sets $(G_n)_{n=1}^P$ and $(H_n)_{n=1}^Q$ with $G_P = G$ and $H_Q = H$. It is now evident that one may define and new chain $(F_n)_{n=1}^{P+Q+1}$ by

- $F_n = G_n$ if $n = 1, 2, \ldots, P$.
- $F_n = H_{n-P}$ if $n = P+1, P+2, \ldots, P+Q$.
- $F_{P+Q+1} = F_P \setminus F_{P+Q} (= G \setminus H)$ or $F_{P+Q+1} = F_P \cup F_{P+Q} (= G \cup H)$ depending on case.

This shows that $\mathcal{G}$ is closed under set-theoretic difference and union and is therefore a field. ■

It is the curse of measure theory that no corresponding result is true for σ-fields[1]. When discussing the σ-field generated by a family of sets we have to go through contortions.

EXAMPLE Consider for example the following question

> If $\mathcal{M}$ is a σ-field of subsets of X and S is a subset of X, show that the σ-field generated by $\mathcal{M} \cup \{S\}$ is
>
> $$\{(A \cap S) \cup (B \cap S^c); A, B \in \mathcal{M}\}.$$

We will prove in a moment that $\mathcal{H} = \{(A \cap S) \cup (B \cap S^c); A, B \in \mathcal{M}\}$ is a σ-field on X. On the other hand, if $\mathcal{G}$ is a σ-field with $\mathcal{M} \subseteq \mathcal{G}$ and $S \in \mathcal{G}$, then $(A \cap S) \cup (B \cap S^c) \in \mathcal{G}$ whenever $A, B \in \mathcal{M}$, so $\mathcal{H} \subseteq \mathcal{G}$. It follows that $\mathcal{H}$ is the smallest σ-field containing $\mathcal{M}$ and S.

To establish the claim, observe that $X = (X \cap S) \cup (X \cap S^c) \in \mathcal{H}$. Further

$$\begin{aligned}
((A \cap S) \cup (B \cap S^c))^c &= (A^c \cup S^c) \cap (B^c \cup S) \\
&= (A^c \cap B^c) \cup (S^c \cap B^c) \cup (A^c \cap S) \cup (S^c \cap S) \\
&= (S^c \cap B^c) \cup (A^c \cap S) \in \mathcal{H}
\end{aligned}$$

since $S^c \cap S = \emptyset$ and

$$A^c \cap B^c = (S^c \cap A^c \cap B^c) \cup (S \cap A^c \cap B^c) \subseteq (S^c \cap B^c) \cup (A^c \cap S).$$

On the other hand, we have

$$\bigcup_{k=1}^{\infty} \Big((A_k \cap S) \cup (B_k \cap S^c) \Big) = \left(\left(\bigcup_{k=1}^{\infty} A_k \right) \cap S \right) \cup \left(\left(\bigcup_{k=1}^{\infty} B_k \right) \cap S^c \right),$$

showing that $\mathcal{H}$ is a σ-field.

□

---

[1] This is not strictly true, but you will need to understand transfinite induction in order to state it.

1.4 Extending premeasures from fields to σ-fields

In this section our objective is the following result.

THEOREM 6 (CARATHÉODORY'S EXTENSION THEOREM) *Let μ be a premeasure on a field $\mathcal{F}$ of subsets of X. Let $\mathcal{G}$ be the σ-field generated by $\mathcal{F}$. Then there exists a measure ν on $\mathcal{G}$ which agrees with μ on $\mathcal{F}$.*

The first step in the proof of the Carathéodory Extension Theorem is the construction of an ***outer measure***. As opposed to measures, which are defined on fields, outer measures are defined on *all* subsets of the ambient space X.

DEFINITION *An outer measure θ on a set X is a map $\theta : \mathcal{P}_X \longrightarrow [0, \infty]$ with the following properties*

(i) $\theta(\emptyset) = 0$.

(ii) *If $A \subseteq B \subseteq X$, then $\theta(A) \le \theta(B)$. We refer to this as θ being **monotone**. The larger the subset, the larger the value of the set function.*

(iii) $\theta\left(\bigcup_{j=1}^{\infty} A_j\right) \le \sum_{j=1}^{\infty} \theta(A_j)$. *We express this condition as θ being **countably subadditive**.*

LEMMA 7 *Let μ be a premeasure on a field $\mathcal{F}$ of subsets of X. Let a set function $\mu^\star$ be defined on $\mathcal{P}_X$ by*

$$\mu^\star(A) = \inf \sum_{j=1}^{\infty} \mu(A_j) \tag{1.6}$$

where the infimum is taken over all possible sequences of sets $A_j \in \mathcal{F}$ such that $A \subseteq \bigcup_{j=1}^{\infty} A_j$. Then $\mu^\star$ is an outer measure on X.

Proof. First notice that in defining the infimum, we can always take $A_1 = X$ and $A_j = \emptyset$ for $j = 2, 3, \ldots$. Thus, there is always at least one covering over which the infimum is taken. Conditions (i) and (ii) in the definition of outer measure are trivially satisfied. We need only check the condition (iii). Towards this, let $A \subseteq X$ and $A_j \subseteq X$ with $A \subseteq \bigcup_{j=1}^{\infty} A_j$. We must show that

$$\mu^\star(A) \leq \sum_{j=1}^{\infty} \mu^\star(A_j).$$

If this fails to be true, then there exits $\epsilon > 0$ such that

$$\sum_{j=1}^{\infty} \mu^\star(A_j) < \mu^\star(A) - \epsilon$$

and indeed

$$\sum_{j=1}^{\infty} (\mu^\star(A_j) + \epsilon 2^{-j}) < \mu^\star(A).$$

But $\mu^\star(A_j)$ is defined as an infimum, so there do exist sets $F_{j,k} \in \mathcal{F}$ such that $A_j \subseteq \bigcup_{k=1}^{\infty} F_{j,k}$ and

$$\sum_{k=1}^{\infty} \mu(F_{j,k}) < \mu^\star(A_j) + \epsilon 2^{-j}.$$

But now we have $A \subseteq \bigcup_{j,k} F_{j,k}$ and the double indexed family $F_{j,k}$ is still countably indexed and could if necessary be written out as a sequence. We get

$$\mu^\star(A) \leq \sum_{j,k} \mu(F_{j,k}) \leq \sum_{j=1}^{\infty} \sum_{k=1}^{\infty} \mu(F_{j,k}) < \sum_{j=1}^{\infty} (\mu^\star(A_j) + \epsilon 2^{-j}) < \mu^\star(A).$$

This contradiction establishes the desired result. ∎

The next step in the saga is to define the concept of measurability with respect to an outer measure.

DEFINITION *Let θ be an outer measure on a set X. Then a subset M of X is said to be θ-measurable if and only if for every set E of X we have*

$$\theta(E) = \theta(E \cap M) + \theta(E \cap M^c).$$

We are using the notation $M^c = X \setminus M$ as a shorthand for complementation in X.

We can think of this in terms of the "cookie cutter" analogy. We imagine that M is the cookie cutter and E is the cookie dough. Then the cutter breaks up the dough into two disjoint pieces, $E \cap M$ and $E \cap M^c$. If M is a "good" cookie cutter, the amount of dough in the two pieces will always add up to the amount that was present originally no matter what the shape of the dough.

PROPOSITION 8 *Let θ be an outer measure on a set X. Let $\mathcal{M}$ be the collection of all subsets of X that are θ-measurable. Then $\mathcal{M}$ is a σ-field and the restriction of θ to $\mathcal{M}$ is a measure.*

We prove Proposition 8 in several steps.

Proof that $\mathcal{M}$ is a field. It is immediately obvious that $X \in \mathcal{M}$ and also that $M \in \mathcal{M}$ implies that $M^c \in \mathcal{M}$. So it remains only to establish that (iii) of the definition of a field holds. Towards this, it will suffice to show

$$A, B \in \mathcal{M} \qquad \Longrightarrow \qquad A \cap B \in \mathcal{M}$$

since we already know that $\mathcal{M}$ is closed under complementation. We have

$$\theta(E) = \theta(E \cap A) + \theta(E \cap A^c).$$

Now apply B as a "cookie cutter" to both $E \cap A$ and $E \cap A^c$

$$\theta(E) = \theta(E \cap A \cap B) + \theta(E \cap A \cap B^c) + \theta(E \cap A^c \cap B) + \theta(E \cap A^c \cap B^c).$$

Next we use the subadditivity of θ to get

$$\begin{aligned}\theta(E) &\geq \theta(E \cap A \cap B) + \theta\Big((E \cap A \cap B^c) \cup (E \cap A^c \cap B) \cup (E \cap A^c \cap B^c)\Big),\\ &= \theta(E \cap A \cap B) + \theta(E \cap ((A \cap B^c) \cup (A^c \cap B) \cup (A^c \cap B^c))),\\ &= \theta(E \cap A \cap B) + \theta(E \cap (A \cap B)^c).\end{aligned}$$

Using the subadditivity again, we get

$$\theta(E) \leq \theta(E \cap A \cap B) + \theta(E \cap (A \cap B)^c).$$

Combining the two inequalities gives

$$\theta(E) = \theta(E \cap A \cap B) + \theta(E \cap (A \cap B)^c)$$

and this completes the proof that $\mathcal{M}$ is a field. ■

LEMMA 9 *Let $A_j \in \mathcal{M}$ for $j \in J$ where J is a countable index set and suppose that the A_j are pairwise disjoint. Then for every $E \subseteq X$ we have*

$$\theta\left(E \cap \left(\bigcup_{j \in J} A_j\right)\right) = \sum_{j \in J} \theta(E \cap A_j). \tag{1.7}$$

Proof. Let us first consider the case where J is finite. Let J have n elements. If $n = 1$, then (1.7) is a tautology. If $n = 2$, then since $A_1 \in \mathcal{M}$ and by the disjointness of A_1 and A_2,

$$\begin{aligned}\theta(E \cap (A_1 \cup A_2)) &= \theta(E \cap (A_1 \cup A_2) \cap A_1) + \theta(E \cap (A_1 \cup A_2) \cap A_1^c) \\ &= \theta(E \cap A_1) + \theta(E \cap A_2)\end{aligned}$$

For $n \geq 3$ we use strong induction on n. We have

$$\begin{aligned}\theta\left(E \cap \left(\bigcup_{k=1}^{n} A_k\right)\right) &= \theta(E \cap A_1) + \theta\left(E \cap \left(\bigcup_{k=2}^{n} A_k\right)\right) \\ &= \theta(E \cap A_1) + \sum_{k=2}^{n} \theta(E \cap A_k) \\ &= \sum_{k=1}^{n} \theta(E \cap A_k)\end{aligned}$$

where the induction hypothesis has been used with 2 sets in the first line and with $n-1$ sets in the second line. This completes the finite case. For the infinite case, we have

$$\theta\left(E \cap \left(\bigcup_{k=1}^{\infty} A_k\right)\right) \geq \theta\left(E \cap \left(\bigcup_{k=1}^{n} A_k\right)\right) = \sum_{k=1}^{n} \theta(E \cap A_k).$$

Letting n tend to infinity now gives

$$\theta\left(E \cap \left(\bigcup_{k=1}^{\infty} A_k\right)\right) \geq \sum_{k=1}^{\infty} \theta(E \cap A_k).$$

On the other hand using the countable subadditivity of θ we have the reverse inequality

$$\theta\left(E \cap \left(\bigcup_{k=1}^{\infty} A_k\right)\right) \leq \sum_{k=1}^{\infty} \theta(E \cap A_k)$$

and combining the two inequalities completes the proof of the lemma. ∎

Completion of the proof of Proposition 8.

We first show that $\mathcal{M}$ is a σ-field . For this it is enough to let A_j be a sequence of disjoint subsets of $\mathcal{M}$ and to show that $\bigcup_{j=1}^{\infty} A_j \in \mathcal{M}$. Let $B_n = \bigcup_{j=1}^{n} A_j$ and $B = \bigcup_{j=1}^{\infty} A_j$. Then, we have

$$\theta(E) = \theta(E \cap B_n) + \theta(E \cap B_n^c) \geq \left\{ \sum_{j=1}^{n} \theta(E \cap A_j) \right\} + \theta(E \cap B^c) \quad (1.8)$$

by Lemma 9 and since $B_n^c \supseteq B^c$. Letting n tend to infinity in (1.8) we get

$$\theta(E) \geq \left\{ \sum_{j=1}^{\infty} \theta(E \cap A_j) \right\} + \theta(E \cap B^c) = \theta(E \cap B) + \theta(E \cap B^c)$$

again by Lemma 9. On the other hand, since θ is subadditive

$$\theta(E) \leq \theta(E \cap B) + \theta(E \cap B^c),$$

and it follows that $\theta(E) = \theta(E \cap B) + \theta(E \cap B^c)$ for all subsets E of X and hence that $B \in \mathcal{M}$.

Finally, setting $E = X$ in (1.7) shows that θ is countably additive on $\mathcal{M}$. ■

We can finally tackle our long term objective.

Proof of the Carathéodory Extension Theorem.

Starting with μ and the field $\mathcal{F}$ we first define an outer measure $\mu^\star$ on X by (1.6). Now let $\mathcal{M}$ be the σ-field of sets that are measurable with respect to $\mu^\star$. We will show that $\mathcal{F} \subseteq \mathcal{M}$.

Let $A \in \mathcal{F}$ and $E \subseteq X$. Let $\epsilon > 0$. Then, from the definition of $\mu^\star(E)$, there exist sets $F_j \in \mathcal{F}$ such that $E \subseteq \bigcup_{j=1}^{\infty} F_j$ and $\sum_{j=1}^{\infty} \mu(F_j) < \mu^\star(E) + \epsilon$. Then we have

$$\begin{aligned}
\mu^\star(E \cap A) + \mu^\star(E \cap A^c) &\leq \sum_{j=1}^{\infty} \mu(F_j \cap A) + \sum_{j=1}^{\infty} \mu(F_j \cap A^c) \\
&= \sum_{j=1}^{\infty} \Big(\mu(F_j \cap A) + \mu(F_j \cap A^c) \Big) \\
&= \sum_{j=1}^{\infty} \mu(F_j) \\
&< \mu^\star(E) + \epsilon.
\end{aligned}$$

Passing to the limit as $\epsilon \to 0$ we get $\mu^\star(E \cap A) + \mu^\star(E \cap A^c) \leq \mu^\star(E)$. By subadditivity we get $\mu^\star(E \cap A) + \mu^\star(E \cap A^c) \geq \mu^\star(E)$ and it follows that $A \in \mathcal{M}$.

Next, we need to show that μ and $\mu^\star$ agree on $\mathcal{F}$. Given $F \in \mathcal{F}$, we can set $A_1 = F$, $A_j = \emptyset$ for $j = 1, 2, \ldots$ to get $\mu^\star(F) \leq \mu(F)$. To get the inequality in the opposite direction, we must show that whenever $A_j \in \mathcal{F}$ and $F \subseteq \bigcup_{j=1}^\infty A_j$ we necessarily have

$$\theta(F) \leq \sum_{j=1}^{\infty} \mu(A_j). \tag{1.9}$$

We do this in two steps by manipulating the A_j. Firstly we ensure that the A_j are disjoint by replacing A_1 with itself, A_2 with $A_2 \setminus A_1$, A_3 with $A_3 \setminus (A_1 \cup A_2)$, etc. This process makes the A_j smaller, so showing that (1.9) holds with the "new" A_j implies that it holds with the "original" A_j. Secondly, we replace each A_j by $A_j \cap F$. Again, since the process makes the A_j smaller, it is enough to show (1.9) for the "new" sets. Note that both of these processes generate subsets of $\mathcal{F}$. So, it is enough to show (1.9) in case that A_j are disjoint subsets of F. But in that case we also have $F = \bigcup_{j=1}^\infty A_j$ and $\mu^\star(F) = \sum_{j=1}^\infty \mu(A_j)$ holds since μ is a measure on $\mathcal{F}$.

Now we are done because $\mathcal{G} \subseteq \mathcal{M}$ and if we define ν to be the restriction of $\mu^\star$ to $\mathcal{G}$, then ν is clearly a measure on $\mathcal{G}$. ■

1.5 Borel sets and Lebesgue sets

We can now apply the rather general results of the previous section to the case of the length premeasure on the field generated by the intervals closed on the left and open on the right. In this case, we obtain two σ-fields $\mathcal{G}$ and $\mathcal{M}$ and these are called the Borel σ-field of $\mathbb{R}$ and the Lebesgue σ-field of $\mathbb{R}$ respectively. To be explicit, the Borel σ-field of $\mathbb{R}$ is the smallest σ-field containing the intervals closed on the left and open on the right. The Lebesgue σ-field on the other hand is the σ-field of all sets that are measurable with respect to the "length" outer measure.

In fact it is possible to define the Borel σ-field for any metric space X. It is clear that every open interval in $\mathbb{R}$ is a countable union of intervals closed on the left and open on the right. So every open interval is in $\mathcal{G}$. On the other hand, every interval closed on the left and open on the right is a countable intersection of open intervals. So, in fact, the open intervals and the intervals closed on the left and open on the right generate the same σ-field . Also, every open subset of $\mathbb{R}$ is a countable union of open intervals, so that $\mathcal{G}$ is also the σ-field generated by the open sets. This prompts the following definition.

DEFINITION *Let X be a metric space. Then the* ***Borel σ-field*** $\mathcal{B}_X$ *of X is the smallest σ-field containing the open subsets of X. A subset B of X is said to be a* ***Borel subset*** *of X (or just a Borel set if the context is clear) if $B \in \mathcal{B}_X$.*

It goes without saying that Borel sets are very difficult to understand and to get a grip on. The only way in practice of showing that a set is Borel is to build it explicitly out of countable unions and countable intersections starting from open sets. In fact there is a special terminology for this. A subset of a metric space X is said to be a G_δ if it is a countable intersection of open subsets. It is an F_σ if it is a countable union of closed sets. It is a $G_{\delta\sigma}$ if it is a countable union of G_δ subsets and so on. The greek letters δ and σ stand for *durchschnitt* and *summe* in this context. One might hope that after some fixed finite number of such operations one would have captured all Borel subsets, but this unfortunately is not the case. To show that a subset is not a Borel subset, in practice, we have to find a σ-field containing all the open sets, but which does not contain the given subset.

We can now state the following Corollary of the Carathéodory Extension Theorem

COROLLARY 10 *There is a measure ν defined on the Borel σ-field of $\mathbb{R}$ which assigns to every interval its length.*

Proof. We apply the Carathéodory Extension Theorem to the field $\mathcal{F}$ generated by intervals closed on the left and open on the right. Let $\mathcal{G}$ be the σ-field generated by $\mathcal{F}$. Then $\mathcal{G}$ is just the Borel field of $\mathbb{R}$ and hence the length premeasure on $\mathcal{F}$ extends to the σ-field of all Borel subsets of $\mathbb{R}$.

At the moment, it is not clear what difference there might be between the Borel subsets and the Lebesgue subsets of $\mathbb{R}$. It is clear that every Borel subset is Lebesgue, but could the converse also be true? Well, it turns out that this is not the case.

PROPOSITION 11 *Both the Borel σ-field and the Lebesgue σ-field are translation invariant. Also, Lebesgue measure is translation invariant.*

Proof. Since the open subsets of $\mathbb{R}$ are translation invariant, it follows that the Borel σ-field of $\mathbb{R}$ is also translation invariant. The field $\mathcal{F}$ generated by intervals closed on the left and open on the right is translation invariant and also so is the length premeasure μ. It follows that Lebesgue outer measure is translation invariant

$$\mu^\star(E + x) = \mu^\star(E) \qquad \forall E \subseteq \mathbb{R}, x \in \mathbb{R},$$

and therefore the Lebesgue σ-field is translation invariant and also the Lebesgue measure ν which is just the restriction of $\mu^\star$. ∎

EXAMPLE The following example is rather counterintuitive. Let us take $X = \mathbb{Z}$, the set of all integers and let $\mathcal{F}$ be the collection of periodic subsets of $\mathbb{Z}$. To be explicit, a subset A of $\mathbb{Z}$ is periodic, if there exists an integer $n \geq 1$ and a subset B of $\{0, 1, 2, \ldots, n-1\}$ such that $A = B + n\mathbb{Z}$. The smallest n for which this can be done is called the ***period*** of A. Now it is clear that $\mathcal{F}$ is a field. If for example A_j is a periodic subset with period n_j for $j = 1, 2$, it is fairly straightforward to show that $\mathbb{Z} \setminus A_1$ is periodic with period n_1 and that $A_1 \cup A_2$ can be represented in the form $B + n\mathbb{Z}$ for $B \subseteq \{0, 1, 2, \ldots, n-1\}$ where n is the LCM of n_1 and n_2. For an element A of $\mathcal{F}$ we can define a density $\mu(A)$ by $\mu(A) = \dfrac{|B|}{n}$. Intuitively, this is the proportion of integers that are in the subset A. It is also possible to show that μ is finitely-additive on $\mathcal{F}$. This is true, because whenever one is dealing with only finitely many sets of $\mathcal{F}$, we can view everything on the period of the lowest common multiple of the periods of the given subsets. It seems reasonable that μ would also be a premeasure on $\mathcal{F}$. However, let us consider the consequences of this statement. Carathéodory's Extension Theorem would then guarantee an extension ν of μ to the σ-field $\mathcal{G}$ generated by $\mathcal{F}$. Now clearly $\{0\} = \bigcap_{n=1}^\infty n\mathbb{Z}$, so $\{0\} \in \mathcal{G}$. In fact, we can write $\mathbb{Z} \setminus \{0\}$ as a union $\bigcup_{k=1}^\infty A_k$ of disjoint periodic sets A_k. We will have $\nu(\{0\}) = 1 - \sum_{k=1}^\infty \mu(A_k)$. Equally well, we can realize $\mathbb{Z} \setminus \{n\}$ as a union $\bigcup_{k=1}^\infty (n + A_k)$ and then

$$\nu(\{n\}) = 1 - \sum_{k=1}^\infty \mu(n + A_k) = 1 - \sum_{k=1}^\infty \mu(A_k) = \nu(\{0\}),$$

so that $\nu(\{n\})$ will be independent of n. Now $\nu(\{0\}) > 0$ is not possible because we can find a positive integer N such that $N\nu(\{0\}) > 1$ and then $\nu(\{0, 1, 2, \ldots, N\}) > 1$ which is impossible. So it must be the case that $\nu(\{0\}) = 0$. But then $1 = \nu(\mathbb{Z}) = \nu(\bigcup_{n\in\mathbb{Z}}\{n\}) = \sum_{n\in\mathbb{Z}} \nu(\{n\}) = 0$. This contradiction shows that the original μ cannot be countably additive. In fact, one can deduce the existence of disjoint periodic subsets A_k with $\bigcup_{k=1}^\infty A_k = \mathbb{Z}$ but such that $\sum_{k=1}^\infty \mu(A_k) < 1$.

We can understand this by means of an explicit example. Let

$$\begin{aligned} B_1 &= \{-1, 0, 1\} + 8\mathbb{Z} \\ B_2 &= \{-2, 2\} + 16\mathbb{Z} \\ B_3 &= \{-3, 3\} + 32\mathbb{Z} \\ B_k &= \{-k, k\} + 2^{k+2}\mathbb{Z} \qquad (k \geq 2) \end{aligned}$$

Then it is clear that $\bigcup_{k=1}^\infty B_k = \mathbb{Z}$ and $\sum_{k=1}^\infty \mu(B_k) = 9/16 < 1$. The B_k are not disjoint, but they can be made so following standard procedures. □

1.6 Uniqueness of the Extension

In the previous section we proved the existence of an extension. What about the uniqueness? Could there be more than one possible extension. Well under reasonable hypotheses, the answer is no. The extra condition that is needed to make this conclusion possible is one that occurs a great deal in measure theory. It is called σ-finiteness.

DEFINITION *Let $\mathcal{F}$ be field or σ-field on a set X. Let μ be a measure on $\mathcal{F}$. Then μ is said to be finite if and only if $\mu(X) < \infty$.*

DEFINITION *Let $\mathcal{F}$ be field or σ-field on a set X. Let μ be a measure on $\mathcal{F}$. Then μ is said to be σ-finite if and only if there exists a sequence of subsets (X_j) of X with $X_j \in \mathcal{F}$, $\mu(X_j) < \infty$ for all j and $X = \bigcup_{j=1}^\infty X_j$.*

There are two approaches to the uniqueness question. We will develop them both. The first approach is the one favoured by probabilists.

DEFINITION *Let X be a set. Then a collection of subsets $\mathcal{P}$ of X is said to be a π-system if*

(π_1) $A, B \in \mathcal{P} \Longrightarrow A \cap B \in \mathcal{P}$.

On the other hand we also make the following definition.

DEFINITION *Let X be a set. Then a collection of subsets $\mathcal{L}$ of X is said to be a λ-system if*

(λ_1) $X \in \mathcal{L}$.

(λ_2) $A \in \mathcal{L} \Longrightarrow X \setminus A \in \mathcal{L}$.

(λ_3) *Whenever $A_1, A_2, \dots$ are disjoint subsets of X in $\mathcal{L}$ then the union $\bigcup_{j=1}^{\infty} A_j$ is also in $\mathcal{L}$.*

In fact, if (λ_1) and (λ_3) are true, then (λ_2) implies the condition that $A, B \in \mathcal{P}, A \supseteq B \Longrightarrow A \setminus B \in \mathcal{P}$. This is because $(A \setminus B)^c = (X \setminus A) \cup B$ and $(X \setminus A)$ and B are disjoint.

The following Lemma is now fairly clear

LEMMA 12 *A collection of subsets $\mathcal{F}$ of X which is both a π-system and a λ-system is also a σ-field .*

Proof. First of all $\emptyset = X \setminus X \in \mathcal{F}$. Since $\mathcal{F}$ is closed under finite intersections and complementation, it is clear that it is also closed under finite unions. Therefore $\mathcal{F}$ is a field. But now, given a sequence $A_1, A_2, \dots$ of subsets of X in $\mathcal{F}$, we can adjust them to make them disjoint, using the standard trick. We define

$$B_1 = A_1, \quad B_2 = A_2 \setminus A_1, \quad B_3 = A_3 \setminus (A_1 \cup A_2), \dots$$

We also have $\bigcup_{j=1}^{\infty} A_j = \bigcup_{j=1}^{\infty} B_j \in \mathcal{F}$. Thus $\mathcal{F}$ is a σ-field as required. ■

The key result in this section is the following.

THEOREM 13 (DYNKIN'S $\pi - \lambda$ THEOREM) *If $\mathcal{P}$ is a π-system and $\mathcal{L}$ is a λ-system and $\mathcal{P} \subseteq \mathcal{L}$, then $\sigma(\mathcal{P}) \subseteq \mathcal{L}$ where $\sigma(\mathcal{P})$ is the σ-field generated by $\mathcal{P}$.*

We will need the following lemma.

LEMMA 14 *Let $\mathcal{F}$ be a λ-system on X and suppose that $A \in \mathcal{F}$. We define $\mathcal{F}_A$ the system of all subsets B of X such that $A \cap B \in \mathcal{F}$. Then $\mathcal{F}_A$ is also a λ-system on X.*

Proof. There are three axioms to check. Since $X \cap A = A \in \mathcal{F}$, $X \in \mathcal{F}_A$. If $B \in \mathcal{F}_A$, then $A \cap B \in \mathcal{F}$ and so $A \setminus B = A \setminus (A \cap B) \in \mathcal{F}$ since $\mathcal{F}$ preserves contained differences. But $(X \setminus B) \cap A = A \setminus B$, so $X \setminus B \in \mathcal{F}_A$. This shows that $\mathcal{F}_A$ is closed under complementation. Finally, let B_j be a disjoint sequence in $\mathcal{F}_A$. Then the intersections $A \cap B_j$ are also disjoint and lie in $\mathcal{F}$. It follows that $A \cap \bigcup_{j=1}^{\infty} B_j = \bigcup_{j=1}^{\infty}(A \cap B_j) \in \mathcal{F}$. So, $\bigcup_{j=1}^{\infty} B_j \in \mathcal{F}_A$. This shows that $\mathcal{F}_A$ is closed under disjoint countable unions. ■

Proof of Dynkin's $\pi - \lambda$ Theorem. Let $\mathcal{F}$ be the λ-system generated by $\mathcal{P}$. In other words, $\mathcal{F}$ is the intersection of all λ-systems containing $\mathcal{P}$. It is clear that $\mathcal{F} \subseteq \mathcal{L}$. If we can show that $\mathcal{F}$ is also a π-system, then it will be a σ-field and the desired conclusion will follow.

Now, let $A \in \mathcal{P}$ and $B \in \mathcal{P}$. Then $A \cap B \in \mathcal{P} \subseteq \mathcal{F}$ so that $B \in \mathcal{F}_A$. By Lemma 14, $\mathcal{F}_A$ is a λ-system containing $\mathcal{P}$. It now follows that $\mathcal{F} \subseteq \mathcal{F}_A$ because $\mathcal{F}$ is the intersection of all λ-systems containing $\mathcal{P}$. But, this means that $A \in \mathcal{P}$ and $B \in \mathcal{F}$ implies that $A \cap B \in \mathcal{F}$. So, if $B \in \mathcal{F}$ then $\mathcal{P} \subseteq \mathcal{F}_B$. But then, $\mathcal{F} \subseteq \mathcal{F}_B$, because again by Lemma 14 $\mathcal{F}_B$ is a λ-system containing $\mathcal{P}$ and $\mathcal{F}$ is the smallest such animal. So, finally we have shown that

$$A, B \in \mathcal{F} \Longrightarrow A \cap B \in \mathcal{F}.$$

In other words, $\mathcal{F}$ is a π-system and the proof is complete. ■

We can now use this result to obtain information about the uniqueness of extensions.

PROPOSITION 15 *Let $\mathcal{P}$ be a π-system and μ_1 and μ_2 be finite measures on $\sigma(\mathcal{P})$ which agree on $\mathcal{P}$ and on X (i.e. $\mu_1(X) = \mu_2(X)$). Then μ_1 and μ_2 agree on $\sigma(\mathcal{P})$.*

Proof. Let $\mathcal{L}$ be the collection of subsets in $\sigma(\mathcal{P})$ on which μ_1 and μ_2 agree. We will show that $\mathcal{L}$ is a λ-system. By hypothesis, $X \in \mathcal{L}$. Now suppose that $A \in \mathcal{L}$. Then $\mu_1(A) = \mu_2(A)$. We find that $\mu_j(X \setminus A) + \mu_j(A) = \mu_j(X)$ for $j = 1, 2$. It follows that $\mu_1(X \setminus A) = \mu_2(X \setminus A)$. It's very important here that $\mu_j(X)$ is finite. If both the $\mu_j(X)$ and the $\mu_j(A)$ are infinite, we cannot deduce the value of $\mu_j(X \setminus A)$. This shows that $\mathcal{L}$ satisfies both (λ_1) and (λ_2). The fact that it satisfies (λ_3) uses the fact that μ_j are measures. If A_j are disjoint subsets in $\mathcal{L}$, then we have

$$\mu_1\left(\bigcup_{j=1}^{\infty} A_j\right) = \sum_{j=1}^{\infty} \mu_1(A_j) = \sum_{j=1}^{\infty} \mu_2(A_j) = \mu_2\left(\bigcup_{j=1}^{\infty} A_j\right)$$

and it follows that $\bigcup_{j=1}^{\infty} A_j \in \mathcal{L}$. This completes the verification that $\mathcal{L}$ is a λ-system. Finally, Dynkin's π-λ Theorem shows that $\sigma(\mathcal{P}) \subseteq \mathcal{L}$ and the proof is complete. ∎

The usual application is to the uniqueness of the extension in Carathéodory's Extension Theorem.

THEOREM 16 *Let μ be a σ-finite premeasure on a field $\mathcal{F}$ of subsets of X. Let $\mathcal{G}$ be the σ-field generated by $\mathcal{F}$. Then there exists a unique measure ν on $\mathcal{G}$ which agrees with μ on $\mathcal{F}$.*

Proof. The existence of ν is already contained in Carathéodory's Extension Theorem. It is the uniqueness with which we are really concerned here. So, let ν_1 and ν_2 be two possible extensions. Now, since μ is σ-finite, we can find a sequence of subsets (X_j) of X in $\mathcal{F}$, such that $\mu(X_j)$ is finite and $X = \bigcup_{j=1}^{\infty} X_j$. So let $\mathcal{F}_j$ be the field of those sets of $\mathcal{F}$ that are contained inside X_j. We view this as a field on X_j. Note that this field is exactly the same as the field of traces $\{F \cap X_j; F \in \mathcal{F}\}$. Let $\mathcal{G}_j = \sigma(\mathcal{F}_j)$. Since ν_1 and ν_2 agree on $\mathcal{F}_j$, by Proposition 15, they also agree on $\mathcal{G}_j$. We now construct a new σ-field $\mathcal{H}_j = \{H; H \subseteq X,\ H \cap X_j \in \mathcal{G}_j\}$ which clearly contains $\mathcal{F}$. Therefore, by definition of $\mathcal{G}$ we have $\mathcal{G} \subseteq \mathcal{H}_j$. So $G \in \mathcal{G}$ implies that for each j, $G \cap X_j \in \mathcal{G}_j$. Now, without loss of generality, we can arrange that the X_j are disjoint. So $G = \bigcup_{j=1}^{\infty} G \cap X_j$ is a disjoint union and

$$\nu_1(G) = \sum_{j=1}^{\infty} \nu_1(G \cap X_j) = \sum_{j=1}^{\infty} \nu_2(G \cap X_j) = \nu_2(G).$$

Thus ν_1 and ν_2 agree on $\mathcal{G}$. ∎

1.7 Monotone Classes*

The second approach to the uniqueness problem is the one that is usually adopted by mathematicians (as distinct from probabilists). It uses a new construct, that of a ***monotone class*** .

DEFINITION *Let X be a set. Then a collection of subsets $\mathcal{M}$ of X is said to be a monotone class if*

(i) Whenever $A_1, A_2, \ldots$ are increasing subsets of X in $\mathcal{M}$ then the union $\bigcup_{j=1}^{\infty} A_j$ is also in $\mathcal{M}$.

(ii) *Whenever $A_1, A_2, \ldots$ are decreasing subsets of X in $\mathcal{M}$ then the intersection $\bigcap_{j=1}^{\infty} A_j$ is also in $\mathcal{M}$.*

THEOREM 17 (MONOTONE CLASS THEOREM) *Let $\mathcal{F}$ be a field. If $\mathcal{M}$ is a monotone class and $\mathcal{F} \subseteq \mathcal{M}$, then $\sigma(\mathcal{F}) \subseteq \mathcal{M}$.*

Proof. Let $\mathcal{M}$ be the smallest monotone class containing $\mathcal{F}$. It will suffice to show that $\sigma(\mathcal{F}) = \mathcal{M}$. Note that every σ-field is a monotone class, so we have $\sigma(\mathcal{F}) \supseteq \mathcal{M}$. We will show that $\mathcal{M}$ is in fact a σ-field . We define

$$\mathcal{M}_P = \{Q; P \setminus Q \in \mathcal{M},\ Q \setminus P \in \mathcal{M},\ P \cup Q \in \mathcal{M}\}.$$

It is easy to check that

- $\mathcal{M}_P$ is a monotone class.
- $P \in \mathcal{M}_Q \iff Q \in \mathcal{M}_P$.

Now, let $P \in \mathcal{F}$ be temporarily frozen. Then, since $\mathcal{F}$ is a field, $Q \in \mathcal{F}$ implies that $Q \in \mathcal{M}_P$, i.e. $\mathcal{F} \subseteq \mathcal{M}_P$. Therefore $\mathcal{M} \subseteq \mathcal{M}_P$, because $\mathcal{M}$ is the smallest monotone class containing $\mathcal{F}$. So, in fact, $Q \in \mathcal{M}$ implies that $Q \in \mathcal{M}_P$ and this in turn implies that $P \in \mathcal{M}_Q$. So, unfreezing P and freezing Q temporarily, we have $\mathcal{F} \subseteq \mathcal{M}_Q$ whenever $Q \in \mathcal{M}$. Again because $\mathcal{M}$ is the smallest monotone class containing $\mathcal{F}$ this gives $\mathcal{M} \subseteq \mathcal{M}_Q$ for all $Q \in \mathcal{M}$. But this just says that $\mathcal{M}$ is closed under finite unions and set theoretic differences. So, in fact $\mathcal{M}$ is a σ-field and the reverse inclusion $\sigma(\mathcal{F}) \subseteq \mathcal{M}$ follows. ∎

To illustrate the application of monotone classes, let us give a second proof of Theorem 16 at least in the case where μ is a finite premeasure. The extension to the σ-finite case follows the same ideas that are found in the first proof of Theorem 16.

It is only the uniqueness of the extension that is at issue. So, let ν_1 and ν_2 be two possible extensions to $\sigma(\mathcal{F})$. We consider

$$\mathcal{M} = \{M; M \in \sigma(\mathcal{F}), \nu_1(M) = \nu_2(M)\}.$$

It is easy see that $\mathcal{M}$ is a monotone class. For example in case $M_j \in \mathcal{M}$ and $M_j \uparrow M \in \sigma(\mathcal{F})$ we have

$$\nu_1(M) = \sup_j \nu_1(M_j) = \sup_j \nu_2(M_j) = \nu_2(M).$$

For decreasing sequences of sets we use complementation together with the condition $\mu(X) < \infty$. ∎

1.8 Completions of Measure Spaces

In this section we look at a way of extending abstract measure spaces. So, let X be a set, $\mathcal{G}$ a σ-field on X and μ a measure on $\mathcal{G}$. We say that a subset N of $\mathcal{G}$ is a ***null set*** if $\mu(N) = 0$. If there is a possibility of confusion between several different measures we use the terminology μ-null. Now it seems intuitively clear that if the measure of a subset N is zero, then the measure of any subset Z of N should also be zero. This is very clear for instance in the case where μ is supposed to give the "length" of a subset of $\mathbb{R}$. But there is a snag. The subset Z may not lie in $\mathcal{G}$ even if $N \in \mathcal{G}$ so that $\mu(Z)$ may not even be defined. For this reason, we introduce the notion of completion of a measure space.

DEFINITION *We define a family of sets $\overline{\mathcal{G}}$ as follows. We say that $H \in \overline{\mathcal{G}}$ iff there exist $G, N \in \mathcal{G}$ and $Z \subseteq N$ with N μ-null such that $H = G \triangle Z$. Here we have used $\triangle$ to denote symmetric difference $G \triangle Z = (G \setminus Z) \cup (Z \setminus G)$. We also define a quantity $\overline{\mu}(H) = \mu(G)$. Then $(X,\ \overline{\mathcal{G}},\ \overline{\mu})$ will eventually define the completion of $(X,\ \mathcal{G},\ \mu)$.*

Note that symmetric differences satisfy the identity $A \triangle (A \triangle B) = B$, so we can equally well define $\overline{\mathcal{G}}$ by the relation $H \triangle G \subseteq N$. Yet another way of describing $\overline{\mathcal{G}}$ comes from writing $H = (G \setminus N) \cup \Big((N \setminus Z) \cup (Z \setminus G)\Big)$ where $G \setminus N \in \mathcal{G}$ and $(N \setminus Z) \cup (Z \setminus G) \subseteq N$. This means that the completion $\overline{\mathcal{G}}$ can be defined with the union operation rather than the symmetric difference. Conceptually, this may be a little easier. We have the following Theorem.

THEOREM 18 *Let $\mathcal{G}$ be a σ-field on X and μ a measure on $\mathcal{G}$. Then*

(i) *$\overline{\mathcal{G}}$ is a σ-field .*

(ii) *For $H \in \overline{\mathcal{G}}$, $\overline{\mu}(H)$ is well-defined.*

(iii) *$\overline{\mu}$ is a measure on $\overline{\mathcal{G}}$ extending μ.*

Proof. We have $X = X \triangle \emptyset$ and $\emptyset$ is a null set, so $X \in \overline{\mathcal{G}}$. Now let $H \in \overline{\mathcal{G}}$. Then $H = G \triangle Z$ where $G, N \in \mathcal{G}$, N is null and $Z \subseteq N$. But $H^c = G^c \triangle Z$ and $G^c \in \mathcal{G}$, so $H^c \in \overline{\mathcal{G}}$. Now let $H_j \in \overline{\mathcal{G}}$ for $j \in \mathbb{N}$. Then we can write $H_j = G_j \cup Z_j$ where $G_j, N_j \in \mathcal{G}$, N_j is null and $Z_j \subseteq N_j$. We now have

$$\bigcup_{j=1}^{\infty} H_j = \left(\bigcup_{j=1}^{\infty} G_j\right) \cup \left(\bigcup_{j=1}^{\infty} Z_j\right) \tag{1.10}$$

where $\bigcup_{j=1}^{\infty} G_j \in \mathcal{G}$ and $\bigcup_{j=1}^{\infty} Z_j \subseteq \bigcup_{j=1}^{\infty} N_j$ and $\bigcup_{j=1}^{\infty} N_j$ is a null set in $\mathcal{G}$. Hence $\bigcup_{j=1}^{\infty} H_j \in \overline{\mathcal{G}}$. The proof of (i) is complete.

Next we show that $\overline{\mu}$ is well-defined. Let $H \triangle G_j \subseteq N_j$ for $j = 1, 2$. Then we have $G_1 \triangle G_2 = (H \triangle G_1) \triangle (H \triangle G_2) \subseteq N_1 \cup N_2$. Thus $\mu(G_1 \triangle G_2) = 0$. Then $G_1 \subseteq G_2 \cup (G_1 \triangle G_2)$ and it follows that $\mu(G_1) \leq \mu(G_2)$ and the reverse inequality holds by symmetry. So $\mu(G_1) = \mu(G_2)$.

The uniqueness just proved shows that $\overline{\mu}$ extends μ. Hence also $\overline{\mu}(\emptyset) = \mu(\emptyset) = 0$. If the H_j in (1.10) are disjoint, then so are the G_j and it follows immediately that

$$\overline{\mu}\left(\bigcup_{j=1}^{\infty} H_j\right) = \mu\left(\bigcup_{j=1}^{\infty} G_j\right) = \sum_{j=1}^{\infty} \mu(G_j) = \sum_{j=1}^{\infty} \overline{\mu}(H_j)$$

■

Finally we have

DEFINITION *A measure space $(X, \mathcal{G}, \mu)$ is complete if it is its own completion.*

and also the following proposition.

PROPOSITION 19 *Let θ be an outer measure on a set X. Let $\mathcal{M}$ be the σ-field of θ-measurable subsets and let μ be the restriction of θ to $\mathcal{M}$. Then we know from Proposition 8 that μ is a measure on $\mathcal{M}$. We have*

(i) *If $Z \subseteq X$ and $\theta(Z) = 0$, then $Z \in \mathcal{M}$.*

(ii) *If $A, Z \subseteq X$ and $\theta(Z) = 0$, then $\theta(A) = \theta(A \cup Z)$.*

(iii) *If $A, B \subseteq X$ and $\theta(A \triangle B) = 0$ then $\theta(A) = \theta(B)$.*

(iv) *The measure space $(X, \mathcal{M}, \mu)$ is complete.*

Proof. For (i), we have $\theta(E) \leq \theta(E \cap Z) + \theta(E \cap Z^c) = \theta(E \cap Z^c) \leq \theta(E)$. The first inequality comes from the subadditivity of θ and the second from the monotonicity. Hence by definition of $\mathcal{M}$, $Z \in \mathcal{M}$. For (ii) we simply have $\theta(A) \leq \theta(A \cup Z) \leq \theta(A) + \theta(Z) = \theta(A)$, where the left-hand inequality is from the monotonicity of θ and the right-hand one is from the subadditivity. For (iii) we simply apply (ii) twice to get $\theta(A) = \theta(A \cup B) = \theta(B)$. We have used the relation $(A \cup B) \setminus A \subseteq A \triangle B$ and the monotonicity of θ, to deduce that $\theta((A \cup B) \setminus A) = 0$.

Finally, for (iv), let $G \in \mathcal{M}$ and suppose that $\theta(G \triangle H) = 0$. Then for any subset E of X we have $(E \cap G) \triangle (E \cap H) \subseteq G \triangle H$ and it follows that $\theta((E \cap G) \triangle (E \cap H)) = 0$. Therefore $\theta(E \cap G) = \theta(E \cap H)$ by (iii). Similarly we have $\theta(E \cap G^c) = \theta(E \cap H^c)$. Since $\theta(E) = \theta(E \cap G) + \theta(E \cap G^c)$, it now follows that $\theta(E) = \theta(E \cap H) + \theta(E \cap H^c)$. Since E is an arbitrary subset of X we see that $H \in \mathcal{M}$. That completes the proof of (iv). ∎

COROLLARY 20 *The Lebesgue σ-field $\mathcal{L}$ on $\mathbb{R}$ is complete.*

Proposition 19 allows us to reconcile the approach that we have taken to measure theory in these notes and an approach that used to be used many years ago but has fallen out of favour as being unnecessarily complicated. To illustrate, let μ be a measure on a σ-field $\mathcal{F}$ of subsets of X and suppose that $\mu(X) < \infty$. We now define as before the outer measure $\mu^\star$

$$\mu^\star(A) = \inf \sum_{j=1}^{\infty} \mu(A_j)$$

where the infimum is taken over all possible sequences of sets $A_j \in \mathcal{F}$ such that $A \subseteq \bigcup_{j=1}^{\infty} A_j$. We can also define an ***inner measure*** $\mu_\star$ by

$$\mu_\star(A) = \mu(X) - \mu^\star(X \setminus A).$$

We don't say here precisely what an inner measure is, we leave that to your imagination. We will denote by ν the restriction of $\mu^\star$ to $\mathcal{M}$. We know that ν is a measure on $\mathcal{M}$.

Let A be an arbitrary subset of X. It is easy to see, from the definition of the infimum that given $\epsilon > 0$ there exists a set $B \in \mathcal{M}$ such that $A \subseteq B$ and $\nu(B) \leq \mu^\star(A) + \epsilon$. Here in fact the set B is just a set $\bigcup_{j=1}^{\infty} A_j$ taken from the situation defining the infimum. Now taking a sequence of ϵ_k decreasing to zero, we get $A \subseteq B_k$ and $\nu(B_k) \leq \mu^\star(A) + \epsilon_k$, and taking $B = \bigcap_{k=1}^{\infty} B_k$, we find $B \in \mathcal{M}$, $A \subseteq B$ and $\mu^\star(A) = \nu(B)$. In some sense, this justifies the term outer measure, because we have found a measurable subset B of X containing A with its measure equal to the outer measure of A.

Now repeat all this argument on $X \setminus A$. We will come up with a subset C in $\mathcal{M}$ with $C \subseteq A$ and $\nu(C) = \mu_\star(A)$. Now if $A \in \mathcal{M}$ then we can apply the "cookie cutter" to the whole space

$$\mu^\star(A) + \mu^\star(X \setminus A) = \mu^\star(X \cap A) + \mu^\star(X \cap A^c) = \mu^\star(X) = \mu(X) < \infty$$

and evidently $\mu_\star(A) = \mu^\star(A)$. It is the converse statement in which we are interested. If $\mu_\star(A) = \mu^\star(A)$ then in fact

$$\nu(C) = \mu_\star(A) = \mu^\star(A) = \nu(B)$$

and since we are working in a finite measure space, $\nu(B \setminus C) = 0$. It now follows from Proposition 19 and the completeness of $\mathcal{M}$ that $A \setminus C$ is in $\mathcal{M}$ and hence A is in $\mathcal{M}$. So, the $\mu^\star$-measurable sets are precisely those for which the inner and outer measures coincide.

1.9 Approximating sets in Lebesgue measure

In this section, we come back to the specific topic of Lebesgue measure. We have a concept of Lebesgue measurable set which is really very hard to grasp. So we need a way of knowing that a Lebesgue measurable set is not too bad. The way that we do this is to show that it can be approximated by nice sets. The approximation is carried out in terms of sets of small measure. Later on in these notes, we will look at this same topic in a more general light.

THEOREM 21 *Let $E \in \mathcal{L}$ and let $\epsilon > 0$.*

(i) *Then there exists U open in $\mathbb{R}$ such that $U \supseteq E$ and $\nu(U \setminus E) < \epsilon$.*

(ii) *Then there exists C closed in $\mathbb{R}$ such that $C \subseteq E$ and $\nu(E \setminus C) < \epsilon$.*

Proof. Let us assume to start with that E is bounded. Now by definition

$$\nu(E) = \mu^\star(E) = \inf\{\sum_{j=1}^{\infty} \mu(F_j); E \subseteq \bigcup_{j=1}^{\infty} F_j, F_j \in \mathcal{F}\}.$$

Now, each F_j is a finite union of intervals closed on the left and open on the right. We might as well assume that in fact each F_j is a single such interval. So, given $\epsilon > 0$, we have such intervals F_j with

$$\sum_{j=1}^{\infty} \mu(F_j) < \nu(E) + \epsilon/2$$

If $F_j = [a_j, b_j[$, we define $U_j =]a_j - 2^{-j-1}\epsilon, b_j[$ a slightly larger open interval. Then clearly $E \subseteq \bigcup_{j=1}^{\infty} U_j$ and $\sum_{j=1}^{\infty} \mu(U_j) < \nu(E) + \epsilon$. So now $U = \bigcup_{j=1}^{\infty} U_j$ is open and it is clear that $\nu(U) \le \sum_{j=1}^{\infty} \mu(U_j) < \nu(E) + \epsilon$. Finally, and this is a key point, since E is bounded, $\nu(E) < \infty$ and we can deduce that $\nu(U \setminus E) < \epsilon$.

In the case that E is unbounded, for each $k \in \mathbb{N}$, we find an open set $V_k \supseteq E \cap [-k, k]$ such that $\nu(V_k \setminus (E \cap [-k, k])) < 2^{-k-1}\epsilon$. Let $V = \bigcup_{k=1}^{\infty} V_k$. Then V is again an open subset of $\mathbb{R}$ and

$$\nu(V \setminus E) = \nu\left(\bigcup_k (V_k \setminus E)\right) \le \sum_k \nu(V_k \setminus E) \le \sum_k \nu\Big(V_k \setminus (E \cap [-k, k])\Big) < \epsilon$$

Furthermore, if $x \in E$, then find $k \in \mathbb{N}$ such that $|x| \le k$, so $x \in E \cap [-k, k] \subseteq V_k \subseteq V$. That is $E \subseteq V$. This completes the proof of (i).

To prove (ii) we simply apply (i) to $\mathbb{R} \setminus E$. ■

As a corollary, we can establish the so called regularity of Lebesgue measure. For reasons that may eventually become clear, we approximate from the inside with compact subsets of $\mathbb{R}$ rather than closed subsets.

COROLLARY 22 *Let $E \in \mathcal{L}$.*

(i) $\nu(E) = \inf\{\nu(U); U \textit{ open } \supseteq E\}$.

(ii) $\nu(E) = \sup\{\nu(K); K \textit{ compact } \subseteq E\}$.

Proof. (i) is an immediate consequence of Theorem 21. Also it is clear that $\nu(E) = \sup\{\nu(K); K \text{ closed } \subseteq E\}$. So, all that really needs to be shown here is that if C is closed, then we have $\nu(C) = \sup\{\nu(K); K \text{ compact } \subseteq C\}$. To see this, we set $K_n = C \cap [-n, n]$. Then K_n is closed (intersection of two closed sets) and bounded and hence, by the Heine–Borel Theorem, it is compact. But $\nu(C) = \sup_n \nu(K_n)$, since K_n are increasing and have union C. ■

Another corollary can now be established. The proof is an exercise.

COROLLARY 23 *The completion of the Borel σ-field $\mathcal{B}_{\mathbb{R}}$ on the real line with respect to the (restriction of) Lebesgue measure is the Lebesgue σ-field .*

1.10 A non-measurable set

Here is the simplest way of constructing a non-measurable subset of $\mathbb{R}$. Consider the cosets $a + \mathbb{Q}$ of the rational numbers in $\mathbb{R}$. Each such coset meets the interval $[0, 1]$. This is simply because for every real number a, the interval $[-a, 1 - a]$ meets $\mathbb{Q}$. So, from each coset $a + \mathbb{Q}$ pick an element $x(a) \in [0, 1]$ and let E be the totality of all such elements. The key observation is that if q_1 and q_2 are distinct rational numbers, then $q_1 + E$ and $q_2 + E$ are disjoint sets. To see why, suppose that $q_1 + x(a_1) = q_2 + x(a_2)$. Then $\mathbb{Q} + x(a_1) = \mathbb{Q} + x(a_2)$ and so $x(a_1)$ and $x(a_2)$ lie in the same coset of $\mathbb{Q}$ in $\mathbb{R}$. But we chose just one element from each coset, so it must be that $x(a_1) = x(a_2)$ and therefore $q_1 = q_2$, contradicting the fact that $q_1 \neq q_2$. Now if $\nu(E) > 0$ we can get a contradiction as follows. We clearly have

$$\bigcup_{q \in \mathbb{Q} \cap [0,1]} (q + E) \subseteq [0, 2]$$

and the sets in the union are disjoint. Each has the same measure as E because ν is translation invariant. Since there are infinitely many points in $\mathbb{Q} \cap [0, 1]$, we have a contradiction. We are forced to conclude that $\nu(E) = 0$. But this is no good either, because

$$\mathbb{R} = \bigcup_{q \in \mathbb{Q}} (q + E) \tag{1.11}$$

effectively forcing $\mathbb{R}$ to have zero measure. This contradiction forces us to conclude that $E \notin \mathcal{L}$. With a little bit more effort we can show the following.

PROPOSITION 24 *Let $A \subseteq \mathbb{R}$ be a measurable subset of $\mathbb{R}$ with $\nu(A) > 0$. Then there exists $B \subseteq A$ with B non-measurable.*

Proof. Suppose not. We will demonstrate a contradiction. So, A is a set of positive measure, every subset of which is measurable. Now, we have $A = \bigcup_{n \in \mathbb{Z}} ([n, n+1] \cap A)$. Choose one of the subsets $[n, n+1] \cap A$ which has positive measure — they cannot all have zero measure. Then it will suffice to work with $[n, n + 1] \cap A$ replacing A. So, after translating we can assume without loss of generality that $A \subseteq [0, 1]$.

Now, with E as above, let $A_q = A \cap (q + E)$ for $q \in \mathbb{Q}$ a fixed rational number. By hypothesis, A_q is measurable. We now have

$$\bigcup_{r \in \mathbb{Q} \cap [0,1]} (r + A_q) \subseteq [0, 2]$$

and once again, the $r + A_q$ are disjoint as r varies over $\mathbb{Q} \cap [0, 1]$. This is because $r+A_q \subseteq (r+q)+E$. Since the sets $r+A_q$ all have the same measure, we conclude as above that $\nu(A_q) = 0$. But now, unfreezing $q \in \mathbb{Q}$, we have $A = \bigcup_{q\in\mathbb{Q}} A_q$, by (1.11). It follows that $\nu(A) = 0$ contradicting the fact that A has positive measure. ■

COROLLARY 25 *There is a subset of* $\mathbb{R}$ *which is Lebesgue measurable but not Borel.*

Proof. We will prove this using the Cantor set. We will in fact use two Cantor sets K_1 and K_2. The first of these is the regular Cantor Ternary set based on $[-1, 1]$. We can write

$$K_1 = \{\frac{2}{3}\sum_{k=0}^{\infty} \omega_k 3^{-k}; \omega_k \in \{-1, 1\},\ k = 0, 1, 2, \ldots\}.$$

The second Cantor set will be constructed with a variable ratio of dissection. At the first step, the interval $[-1, 1]$ is split into two closed subintervals, $[-1, 1-2r_0]$ and $[2r_0-1, 1]$. So the right-hand interval is centred at r_0 and has length $2(1-r_0)$. We then repeat the decomposition with a different scale based on r_1. so, the extreme right-hand interval at the second step is $[r_0 + (1 - r_0)(2r_1 - 1), 1]$ etc. The points of K_2 have the form

$$\omega_0 r_0 + \omega_1(1 - r_0)r_1 + \omega_2(1 - r_0)(1 - r_1)r_2 + \cdots$$

Now, after removing the first $k + 1$ groups of intervals, the total length of the set remaining is $2\prod_{j=0}^{k}\Big(2(1 - r_j)\Big)$. If we choose for example $r_j = 1/2 + 2^{-2-j}$, then $2\prod_{j=0}^{\infty}\Big(2(1 - r_j)\Big) > 0$. So we can arrange that K_2 has positive measure. On the other hand, K_1 which is built with $r_j = 2/3$ for all $j = 0, 1, 2, \ldots$, has zero measure. Nevertheless, the two sets can be put into one-to-one correspondence using the omegas and the mappings involved are continuous in both directions. Let us denote this correspondence $\alpha : K_2 \longrightarrow K_1$.

If you believe all this, then we are practically done. Since K_2 has positive Lebesgue measure, it must contain a subset E which is not Lebesgue measurable.

If E is not Lebesgue measurable, then it is certainly not Borel. But now consider $\alpha(E)$ which is a subset of K_1. Since α is continuous in both directions, it preserves open subsets and hence also Borel subsets. Being a Borel subset is a topological property! We conclude that $\alpha(E)$ is not Borel. But K_1 is a closed set of zero Lebesgue measure and therefore every subset of K_1 for example $\alpha(E)$ is necessarily Lebesgue measurable. ■

2

Integration over Measure Spaces

2.1 Measurable functions

So, in this section we have a σ-field $\mathcal{F}$ of subsets of X and a measure μ on $\mathcal{F}$. The first task is to integrate nonnegative simple functions. We need the concept of ***measurability*** .

DEFINITION

(i) *Let $f : X \longrightarrow Y$ where Y is a metric space. Then f is measurable (as a mapping from $(X, \mathcal{F})$ to (Y, d_Y)) iff $f^{-1}(U) \in \mathcal{F}$ for every open subset U of Y.*

(ii) *Let $f : X \longrightarrow Y$ where $(Y, \mathcal{G})$ is a measurable space (this just means that $\mathcal{G}$ is a σ-field on Y). Then f is measurable (as a mapping from $(X, \mathcal{F})$ to $(Y, \mathcal{G})$) iff $f^{-1}(G) \in \mathcal{F}$ for every subset $G \in \mathcal{G}$.*

It is easy to prove that if f is measurable as a mapping from $(X, \mathcal{F})$ to (Y, d_Y), then f is measurable as a mapping from $(X, \mathcal{F})$ to $(Y, \mathcal{B}_Y)$, where $\mathcal{B}_Y$ denotes the Borel σ-field of Y. So, we can if we wish interpret (i) above as a special case of (ii). Another special case of the above situation is when both X and Y are metric spaces. Then we understand the statement that $f : X \longrightarrow Y$ is a ***Borel function*** as the statement that f is measurable with respect to the measurable spaces $(X, \mathcal{B}_X)$ and $(Y, \mathcal{B}_Y)$.

We need some results that will allow us to build combinations of measurable functions.

PROPOSITION 26 *Let $(X, \mathcal{F})$ be a measurable space and let Y_1, Y_2 be separable metric spaces. Let $f_j : X \longrightarrow Y_j$ be $\mathcal{F}$-measurable functions for $j = 1, 2$. Define a new function $f : X \longrightarrow Y_1 \times Y_2$ by*

$$f(x) = (f_1(x), f_2(x)).$$

Then f is also $\mathcal{F}$-measurable.

Proof. We take the product metric on $Y_1 \times Y_2$. Let U be an open subset of $Y_1 \times Y_2$. Then, according to Theorem 29 in the notes for MATH 354, U can be written as a countable union of open balls. But, in the product metric, open balls are just products of open balls, so we may write

$$U = \bigcup_{j=1}^{\infty} \Big(U_{Y_1}(y_{1,j}, t_j) \times U_{Y_2}(y_{2,j}, t_j) \Big)$$

and it follows that

$$f^{-1}(U) = \bigcup_{j=1}^{\infty} \Big(f_1^{-1}(U_{Y_1}(y_{1,j}, t_j)) \cap f_2^{-1}(U_{Y_2}(y_{2,j}, t_j)) \Big)$$

which is clearly in $\mathcal{F}$ because both f_1 and f_2 are measurable. ■

COROLLARY 27 *Let $(X, \mathcal{F})$ be a measurable space and let Y_1, Y_2 be separable metric spaces. Let $f_j : X \longrightarrow Y_j$ be $\mathcal{F}$-measurable functions for $j = 1, 2$. Let Y be a metric space and suppose that $\alpha : Y_1 \times Y_2 \longrightarrow Y$ is a borel mapping. Define a new function $g : X \longrightarrow Y$ by*

$$g(x) = \alpha(f_1(x), f_2(x)).$$

Then g is also measurable. In particular, sums and products of measurable real-valued functions are measurable.

LEMMA 28 *The σ-field generated by the sets $]a, \infty[$ as a runs through $\mathbb{R}$ is the Borel field $\mathcal{B}_{\mathbb{R}}$ of $\mathbb{R}$.*

Proof. Let $\mathcal{C}$ denote the σ-field generated by the sets $]a, \infty[$. We clearly have $]a, \infty[^c =] - \infty, a]$, so for every $b \in \mathbb{R}$ we have $] - \infty, b] \in \mathcal{C}$. Now we can also write

$$] - \infty, b[= \bigcup_{n=1}^{\infty}] - \infty, b - 2^{-n}]$$

so, $] - \infty, b[\in \mathcal{C}$ for every $b \in \mathbb{R}$. It now follows that every bounded open interval $]a, b[=]a, \infty[\cap] - \infty, b[\in \mathcal{C}$. Thus, in fact all open intervals, bounded or not are in $\mathcal{C}$ and hence, all open sets. Therefore $\mathcal{C}$ contains all borel sets. Obviously $\mathcal{C} \subseteq \mathcal{B}_{\mathbb{R}}$, so the two σ-fields are equal. ∎

COROLLARY 29 *Suppose that for each $n \in \mathbb{N}$, f_n is a measurable real-valued mapping. Then so is $f = \sup_{n=1}^{\infty} f_n$.*

Proof. We first observe that

$$\left\{x; \left(\sup_{n=1}^{\infty} f_n(x)\right) > a\right\} = \bigcup_{n=1}^{\infty} \{x; f_n(x) > a\}.$$

This means that the set on the left is measurable for every $a \in \mathbb{R}$. But, it is clear that $\sigma\left(f^{-1}(\mathcal{A})\right) = \{f^{-1}(B); B \in \sigma(\mathcal{A})\}$. We apply this with $f = \sup_{n=1}^{\infty} f_n$ and $\mathcal{A}$ the collection of intervals of the form $]a, \infty[$. It tells that $f^{-1}(B)$ is measurable for every Borel subset B of $\mathbb{R}$. ∎

Obviously, the same argument also works for infima. So it will also follow that $\limsup f_n = \inf_m \sup_{n=m}^{\infty} f_n$ is a measurable function if all of the f_n are. The same result also holds for liminfs. This means in particular, that a pointwise limit of measurable functions is measurable.

2.2 More on Measurable Functions*

We can extend this result to a more general context with the following lemma.

LEMMA 30 *Let Z be a separable metric space, and U a nonempty open subset of Z. Then there is a sequence $(z_k)_{k=1}^{\infty}$ of points of Z and $\delta_k > 0$ such that*

$$\bigcup_{k=1}^{\infty} B(z_k, \delta_k) \subseteq U \subseteq \bigcup_{k=1}^{\infty} U(z_k, \delta_k) \tag{2.1}$$

where we have denoted $B(z, \delta) = \{w \in Z; d(z, w) \leq \delta\}$ and $U(z, \delta) = \{w \in Z; d(z, w) < \delta\}$. In fact, the inclusions in (2.1) are equalities.

THEOREM 31 *Let Z be a separable metric space and $(X, \mathcal{M})$ a measurable space. Let $f_n : X \longrightarrow Z$ be $\mathcal{M}$-measurable functions and let $f_n \underset{n\to\infty}{\longrightarrow} f$ pointwise where $f : X \longrightarrow Z$. Then f is $\mathcal{M}$-measurable.*

Proof. We claim that for every $z \in Z$ and every $\delta > 0$, there exists a set $M \in \mathcal{M}$ such that

$$f^{-1}(U(z, \delta)) \subseteq M \subseteq f^{-1}(B(z, \delta)). \tag{2.2}$$

To do this, we simply set $M = \displaystyle\bigcap_{m=1}^{\infty} \bigcup_{n=m}^{\infty} \{x; d(f_n(x), z) < \delta\}$. This is the set of all x such that $d(f_n(x), z) < \delta$ holds for infinitely many n. With this formulation, and using the pointwise convergence of f_n to f we have that (2.2) holds. (We could equally well have taken $M = \displaystyle\bigcup_{m=1}^{\infty} \bigcap_{n=m}^{\infty} \{x; d(f_n(x), z) < \delta\}$ which is the set of all x such that $d(f_n(x), z) < \delta$ holds for all sufficiently large n i.e. eventually).

With the claim proved, we now apply Lemma 30. For each k we can find $M_k \in \mathcal{M}$ such that

$$f^{-1}(U(z_k, \delta_k)) \subseteq M_k \subseteq f^{-1}(B(z_k, \delta_k)).$$

and now we have for any nonempty open U in Z

$$f^{-1}(U) \subseteq \bigcup_{k=1}^{\infty} f^{-1}(U(z_k, \delta_k)) \subseteq \bigcup_{k=1}^{\infty} M_k \subseteq \bigcup_{k=1}^{\infty} f^{-1}(B(z_k, \delta_k)) \subseteq f^{-1}(U)$$

so that $f^{-1}(U) = \bigcup_{k=1}^{\infty} M_k \in \mathcal{M}$. ■

2.3 The Lebesgue Integral — first steps

A measurable simple function $f : X \longrightarrow [0, \infty]$ is a function between the above spaces which is measurable and takes only finitely many values $s_1, \ldots, s_n$. We set $A_j = f^{-1}(\{s_j\})$. These are disjoint sets and their union is X. We can then succinctly write

$$f = \sum_{j=1}^{n} s_j \mathbb{1}_{A_j}$$

This will be called the ***regimented form*** — the s_j are all distinct, the A_j are disjoint, measurable and nonempty and their union is the whole of X. The key point here is that the regimented decomposition of f is uniquely determined by f. We now define

$$\int f d\mu = \sum_{j=1}^{n} s_j \mu(A_j).$$

We use here standard conventions with regard to arithmetic involving ∞. Our convention is that $0 \cdot \infty = \infty \cdot 0 = 0$. So, the integral of a function that is identically zero on a set of infinite measure is zero. The integral of ∞ times the indicator function of a null set is also zero.

Now we need to know that the integral has the right properties.

PROPOSITION 32

(i) *If f, g are nonnegative measurable simple functions and if $f \leq g$ pointwise, then $\int f d\mu \leq \int g d\mu$.*

(ii) *If f_n, f are nonnegative measurable simple functions and if $f_n \uparrow f$ pointwise, then $\int f_n d\mu \uparrow \int f d\mu$.*

(iii) *If f, g are nonnegative measurable simple functions and $a, b \in [0, \infty]$, then*

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

We will need the following technical lemma.

LEMMA 33 *Let $f = \sum_{k=1}^{m} t_k \mathbb{1}_{B_k}$ where B_k are disjoint measurable sets, then $\int f d\mu = \sum_{k=1}^{m} t_k \mu(B_k)$.*

Proof. Note that we have not required that the union of the B_k is X. Also some of the B_k may be empty and the numbers t_k may not be distinct. Whenever a B_k is empty, we omit that value of k and renumber the remaining B_k. If the union of the B_k is not X, we include a new $B_{m+1} = X \setminus \bigcup_{k=1}^{m} B_k$, set $t_{m+1} = 0$ and replace m by $m + 1$. Neither of these operations affects the truth of the Lemma. Recall that $0 \cdot \infty = 0$, so that even if a newly created B has infinite measure, the lemma remains unaffected.

Thus we may assume that B_k are nonempty and disjoint for $k = 1, \ldots, m$ and that their union is X. On B_k, the function f takes the value t_k which must therefore be one of the s_j. So, we have a map $\alpha : \{1, 2, \ldots, m\} \longrightarrow \{1, 2, \ldots, n\}$ such that $t_k = s_{\alpha(k)}$. The set where f takes the value s_j is then A_j and also $\bigcup_{\alpha(k)=j} B_k$, so these sets must be equal.

$$\int f d\mu = \sum_{j=1}^{n} s_j \mu(A_j) = \sum_{j=1}^{n} s_j \sum_{\alpha(k)=j} \mu(B_k) = \sum_{j=1}^{n} \sum_{\alpha(k)=j} t_k \mu(B_k) = \sum_{k=1}^{m} t_k \mu(B_k)$$

This completes the proof. ■

Proof of Proposition 32. Let us write $f = \sum_{j=1}^{n} s_j \mathbb{1}_{A_j}$, $g = \sum_{k=1}^{m} t_k \mathbb{1}_{B_k}$ both in regimented form. We can write now $\mathbb{1}_{A_j} = \sum_{k=1}^{m} \mathbb{1}_{A_j \cap B_k}$ and therefore, we can write

$$f = \sum_{j=1}^{n} s_j \mathbb{1}_{A_j} = \sum_{j=1}^{n} s_j \sum_{k=1}^{m} \mathbb{1}_{A_j \cap B_k} = \sum_{j=1}^{n} \sum_{k=1}^{m} s_j \mathbb{1}_{A_j \cap B_k}$$

The $A_j \cap B_k$ are disjoint subsets, but possibly empty. We can write in the same way

$$g = \sum_{k=1}^{m} t_k \mathbb{1}_{B_k} = \sum_{k=1}^{m} t_k \sum_{j=1}^{n} \mathbb{1}_{A_j \cap B_k} = \sum_{j=1}^{n} \sum_{k=1}^{m} t_k 1_{A_j \cap B_k}$$

So, according to Lemma 33

$$\int f d\mu = \sum_{j=1}^{n} \sum_{k=1}^{m} s_j \mu(A_j \cap B_k), \qquad \int g d\mu = \sum_{j=1}^{n} \sum_{k=1}^{m} t_k \mu(A_j \cap B_k) \tag{2.3}$$

Now, for each pair (j, k) there are two cases. Either $A_j \cap B_k = \emptyset$ in which case $\mu(A_j \cap B_k) = 0$ or, there exists $x \in A_j \cap B_k$. In the second case $s_j = f(x) \leq g(x) = t_k$. So, it follows from (2.3) that $\int f d\mu \leq \int g d\mu$ by comparing the expansions term by term. This completes the proof of (i).

Next we turn to (iii), which we prove in the same manner. We find

$$af + bg = \sum_{j=1}^{n} \sum_{k=1}^{m} (a s_j + b t_k) 1_{A_j \cap B_k}$$

so that

$$\int (af + bg) d\mu = \sum_{j=1}^{n} \sum_{k=1}^{m} (a s_j + b t_k) \mu(A_j \cap B_k) = a \int f d\mu + b \int g d\mu$$

from (2.3) and manipulative algebra. That completes the proof of (iii).

Now (ii) is much harder and we first tackle the special case $f = t\mathbb{1}_B$, where $t \in [0,\infty]$ and B is a measurable set. We are supposing that f_n are nonnegative measurable simple functions and that $f_n \uparrow t\mathbb{1}_B$ pointwise. If $t = 0$, then $f_n \equiv 0$ for all n and the result is obvious. We can assume therefore that $t > 0$. Now, let $0 < r < t$ and think of r as being close to t. Let $B_n = \{x; f_n(x) > r\}$. Then B_n increases with n because the f_n are increasing. Also $B_n \subseteq B$ for otherwise find $x \in B_n \setminus B$ and we have $0 = t\mathbb{1}_B(x) \geq f_n(x) > r > 0$. Now for each $x \in B$, eventually we will have $f_n(x) > r$ for otherwise $\sup_n f_n(x) \leq r < t = \sup_n f_n(x)$. In other words $\bigcup_n B_n = B$, and it follows that $\sup_n \mu(B_n) = \mu(B)$.

Since $f_n \leq t\mathbb{1}_B$, it follows from (i) that $\int f_n d\mu \leq t\mu(B)$ and therefore that $\sup_n \int f_n d\mu \leq t\mu(B)$. This is the "easy" inequality. To get the "hard" one, we observe that $\int f_n d\mu \geq r\mu(B_n)$ since $f > r\mathbb{1}_{B_n}$ and on taking sups we find $\sup_n \int f_n d\mu \geq r \sup \mu(B_n) = r\mu(B)$. Since r can be taken as close to t as we like (or arbitrarily large if $t = \infty$), we get $\sup_n \int f_n d\mu \geq t\mu(B)$. This settles the special case $f = t\mathbb{1}_B$.

For the general case, we write $f = \sum_{j=1}^m s_j \mathbb{1}_{A_j}$ in regimented form. Now $f_n \uparrow f$, so it follows (for each j fixed) that $f_n\mathbb{1}_{A_j} \uparrow f\mathbb{1}_{A_j} = s_j\mathbb{1}_{A_j}$. Thus, by the special case we have that $\int f_n \mathbb{1}_{A_j} d\mu \uparrow s_j\mu(A_j)$. Finally, by applying an induction to extend (iii) we can deduce

$$\int f_n d\mu = \int \sum_{j=1}^m f_n \mathbb{1}_{A_j} d\mu = \sum_{j=1}^m \int f_n \mathbb{1}_{A_j} d\mu \uparrow \sum_{j=1}^m s_j \mu(A_j) = \int f d\mu$$

and the proof is complete. ■

We now extend the definition of the integral to nonnegative measurable functions.

DEFINITION *Let $f : X \longrightarrow [0,\infty]$ be measurable with respect to a σ-field $\mathcal{F}$ of subsets of X. Let μ be a measure on $(X,\mathcal{F})$. Then we define*

$$\int f d\mu = \sup_{\substack{0 \leq s \leq f \\ s \text{ simple measurable}}} \int s d\mu$$

If f is itself a nonnegative measurable simple function, then it follows immediately from Proposition 32 (ii) that the "old" and "new" definitions agree.

LEMMA 34 *For all $n = 1, 2, \ldots$ and $x \in [0, \infty]$ define*

$$\beta_n(x) = \begin{cases} 2^{-n}\lfloor 2^n x \rfloor & \text{if } 0 \leq x < n, \\ n & \text{if } x \geq n. \end{cases}$$

Then

- $\beta_n : [0, \infty] \longrightarrow [0, \infty[$ *is a Borel mapping.*
- β_n *takes only finitely many values.*
- $\beta_n(x) \uparrow x$ *for all* $x \in [0, \infty]$ *as* n *increases to* ∞.

Proof. The first two assertions are fairly obvious. For the third, there are two cases. If $x < \infty$, then eventually, $x < n$ so that the first definition applies. In that case, $0 \leq x - \beta_n(x) = 2^{-n}(2^n x - \lfloor 2^n x \rfloor) < 2^{-n}$. On the other hand, if $x = \infty$, then $\beta_n(x) = n$ and we are also done! ■

The next step is to extend Proposition 32 to the case of nonnegative measurable functions.

THEOREM 35

(i) *If f, g are nonnegative measurable functions and if $f \leq g$ pointwise, then $\int f d\mu \leq \int g d\mu$.*

(ii) *If f_n, f are nonnegative measurable functions and if $f_n \uparrow f$ pointwise, then $\int f_n d\mu \uparrow \int f d\mu$.*

(iii) *If f, g are nonnegative measurable functions and $a, b \in [0, \infty]$, then*

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

Item (ii) in Theorem 35 is called the Monotone Convergence Theorem.

Proof. (i) follows immediately from the definition and Proposition 32 (i). Now for (ii). By (i) we have $\int f_n d\mu \leq \int f d\mu$ so that $\sup_n \int f_n d\mu \leq \int f d\mu$ The danger is that $\sup_n \int f_n d\mu < \int f d\mu$. In that case, there will exist a simple measurable function s with $0 \le s \le f$ such that already

$$\sup_n \int f_n d\mu < \int s d\mu. \tag{2.4}$$

Unfortunately, s may take ∞ as a value, so we need to be quite careful. Let $s_n(x) = \dfrac{n-1}{n} \min(s(x), n)$. Then s_n is a bounded simple function increasing to s and $s_n(x) < s(x)$ unless $s(x) = 0$. So $\int s_n d\mu \uparrow \int s d\mu$ using Proposition 32 (ii). Replacing the s in (2.4) by a suitable s_n, we can assume without loss of generality that (2.4) holds for a simple measurable function s satisfying $0 \le s(x) \le f(x)$ and $s(x) < f(x)$ whenever $f(x) > 0$. We define $E_n = \{x; f_n(x) \ge s(x)\}$, increasing measurable sets with union X. (To see that $E_n \uparrow X$, observe first that if $f(x) = 0$ then $f_n(x) = s(x) = 0$ and $x \in E_n$ for all n. On the other hand, if $f(x) > 0$ then $s(x) < f(x)$ and $f_n(x) \uparrow f(x)$. We now have

$$\int f_n d\mu \ge \int \mathbb{1}_{E_n} f_n d\mu \ge \int \mathbb{1}_{E_n} s d\mu \uparrow \int s d\mu$$

using Proposition 32 (ii). So $\sup_n \int f_n d\mu \ge \int s d\mu$. The proof of (ii) is complete.

Finally, for (iii) it will suffice to show that if f is a nonnegative measurable function, we can find f_n nonnegative measurable simple functions with $f_n \uparrow f$. We will then be able to approximate g in the same way and we will have $af_n + bg_n \uparrow af + bg$. So, applying Proposition 32 (iii) to get

$$\int (af_n + bg_n) d\mu = a \int f_n d\mu + b \int g_n d\mu$$

and passing to the limit as $n \longrightarrow \infty$ we have the desired result. To prove the claim, we take $f_n = \beta_n \circ f$, where β_n is defined as in Lemma 34. ■

The following lemma describe the Tchebychev Inequality.

LEMMA 36 *Let f be a nonnegative measurable function and $0 < t < \infty$ a scalar. Then*

$$\mu\Big(\{x; f(x) \ge t\}\Big) \le t^{-1} \int f d\mu.$$

Also, if $\int f d\mu = 0$, then $f = 0$ μ-a.e.

Proof.

Let $E_t = \{x; f(x) \geq t\}$. Clearly, $t\mathbb{1}_{E_t} \leq f$. So we have $\int t\mathbb{1}_{E_t} d\mu \leq \int f d\mu$. Effectively that gives us $\mu(E_t) = t^{-1} \int f d\mu$. For the last assertion, we have for every $t > 0$, that $\mu(E_t) = 0$. But $\{x; f(x) > 0\} = \bigcup_{k=1}^{\infty} E_{2^{-k}}$ and so $\{x; f(x) > 0\}$ is a μ-null set. ∎

2.4 The Lebesgue Integral for real and complex valued functions

So far, we have defined the Lebesgue integral of a nonnegative function. We now look at the same issue for functions that are real-valued or complex-valued. The key point is that we only define the integral if

$$\int |f| d\mu < \infty$$

where of course, since $|f|$ is a nonnegative function, the integral $\int |f| d\mu$ is perfectly well defined. Functions that have this property are call ***integrable*** , ***absolutely integrable*** , ***summable*** or ***absolutely summable*** .

Let us start with a real-valued function f. We write

$$f_+(x) = \begin{cases} f(x) & \text{if } f(x) \geq 0, \\ 0 & \text{otherwise.} \end{cases} \qquad \text{and} \qquad f_-(x) = \begin{cases} -f(x) & \text{if } f(x) \leq 0, \\ 0 & \text{otherwise.} \end{cases}$$

In this way, we see that $f_+ \geq 0$, $f_- \geq 0$ and $f = f_+ - f_-$. Now, it is clear that $f_\pm \leq |f|$, so by Theorem 35, (i), we have

$$\int f_\pm d\mu \leq \int |f| d\mu < \infty$$

Therefore, the difference $\int f_+ d\mu - \int f_- d\mu$ is meaningful (it is not of the form $\infty - \infty$) and we define $\int f d\mu$ to be this quantity. We need to see that this integral behaves as it should.

THEOREM 37 *Let f, g be real-valued measurable functions such that $\int |f| d\mu < \infty$ and $\int |g| d\mu < \infty$. Let $a, b \in \mathbb{R}$. Then we have*

(i) $\int |af + bg| d\mu \leq |a| \int |f| d\mu + |b| \int |g| d\mu < \infty$.

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

(iii) $\left| \int f d\mu \right| \leq \int |f| d\mu$.

(iv) *If also $f \leq g$ pointwise, then $\int f d\mu \leq \int g d\mu$.*

Proof. The first statement is easy because $0 \leq |af + bg| \leq |a||f| + |b||g|$. We simply apply part (iii) of Theorem 35. This shows that the integral on the left-hand side of (ii) exists. For (ii), we split the statement up into two separate problems. Let us show first that

$$\int a f d\mu = a \int f d\mu. \tag{2.5}$$

If $a = 0$ this is obvious. If $a = -1$, then it boils down to

$$\int (-f) d\mu = \int (-f)_+ d\mu - \int (-f)_- d\mu = \int (f_-) d\mu - \int (f_+) d\mu = - \int f d\mu$$

so this leaves the case $a > 0$. But that case is easy, for then $(af)_\pm = a(f_\pm)$ and (2.5) follows straightforwardly from Theorem 35 part (iii). This leaves us to show that

$$\int (f + g) d\mu = \int f d\mu + \int g d\mu \tag{2.6}$$

Let us denote $h = f + g$. Then we have $h_+ - h_- = f_+ - f_- + g_+ - g_-$ and it follows that $h_+ + f_- + g_- = h_- + f_+ + g_+$ and it follows from an extended version of Theorem 35 part (iii) that

$$\int (h_+) d\mu + \int (f_-) d\mu + \int (g_-) d\mu = \int (h_-) d\mu + \int (f_+) d\mu + \int (g_+) d\mu \tag{2.7}$$

which is, after rearranging the terms precisely (2.6). Note that all the terms in (2.7) are finite so that there is no problem in subtracting off infinity from infinity. Now for (iii), we see that $\left|\int f d\mu\right|$ is one or other of the quantities $\int f_+ d\mu - \int f_- d\mu$ or $\int f_d \mu - \int f_+ d\mu$. But, both of these are bounded above by $\int |f| d\mu$.

Now for (iv) let $h = g - f \geq 0$ pointwise. So $\int h d\mu \geq 0$. Now apply (ii) with $a = -1$ and $b = 1$ to get the desired conclusion. ■

We now extend the definition of the integral to complex-valued measurable functions in the obvious way. We insist that $\int |f| d\mu < \infty$ and then since $|\Re f| \leq |f|$ and $|\Im f| \leq |f|$, we have that $\int |\Re f| d\mu < \infty$ and $\int |\Im f| d\mu < \infty$ allowing us to define

$$\int f d\mu = \int \Re f d\mu + i \int \Im f d\mu$$

We then have the expected theorem.

THEOREM 38 *Let f, g be complex-valued measurable functions satisfying the conditions $\int |f| d\mu < \infty$ and $\int |g| d\mu < \infty$. Let $a, b \in \mathbb{C}$. Then we have*

(i) $\int |af + bg| d\mu \leq |a| \int |f| d\mu + |b| \int |g| d\mu < \infty$.

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

(iii) $\left| \int f d\mu \right| \leq \int |f| d\mu$.

We leave the proof of (i) and (ii) to the reader.

Proof of (iii). If $\int f d\mu = 0$, then we are done. Otherwise, we can write $\int f d\mu = r\omega$ with $r > 0$ and $|\omega| = 1$. Now we have $\Re \overline{\omega} f \leq |\overline{\omega} f| = |f|$ pointwise so that

$$\int \Re \overline{\omega} f d\mu \leq \int |f| d\mu.$$

But $\int \Re \overline{\omega} f d\mu = \Re \int \overline{\omega} f d\mu = \Re \overline{\omega} \int f d\mu = \overline{\omega} r \omega = r = \left| \int f d\mu \right|$ and the proof is complete. ■

2.5 Interchanging limits and integrals

Here we discuss integrals of limits. With the Riemann Integral, there is very little that one can say. The definition of the Lebesgue integral allows some much more powerful theorems to be proved. Let us start by recalling Item (ii) in Theorem 35.

THEOREM 39 (MONOTONE CONVERGENCE THEOREM) *If f_n, f are nonnegative measurable functions and if $f_n \uparrow f$ pointwise, then $\int f_n d\mu \uparrow \int f d\mu$.*

It should be remarked that we can relax the pointwise convergence assumption in this and all other convergence theorems in the following way. A property is said to hold ***almost everywhere*** if it holds on the set $X \setminus N$ where X is the whole ambient space and N is a null set. If there is possible confusion over the measure μ that is being used to check the nullness of N, we will use the term μ-***almost everywhere***. Actually, these are normally abbreviated to *a.e.* or μ-*a.e.*. The probabilists use *a.s.* meaning ***almost surely*** in the context of a probability measure. So, the statement $f_n \longrightarrow f$ a.e. means that there exists a null set N such that $f_n \longrightarrow f$ on $X \setminus N$. We could then redefine f_n and f to be zero on N and this change would not affect the values of $\int f_n d\mu$ or $\int f d\mu$. For the redefined sequence and limit function, we do have pointwise convergence everywhere and so the Monotone Convergence Theorem applies.

The next step is the following intermediate result.

LEMMA 40 (FATOU'S LEMMA) *Let f_n be nonnegative measurable functions. Then*

$$\int \liminf_{n\to\infty} f_n d\mu \leq \liminf_{n\to\infty} \int f_n d\mu$$

The $\liminf$ on the left is taken over the sequence $(f_n(x))$ for every $x \in X$, while the one on the right is just the $\liminf$ of a sequence of real numbers.

Proof. The definition of the $\liminf$ is $\liminf_{n\to\infty} f_n(x) = \sup_n \inf_{m\geq n} f_m(x)$. So, let us set $g_n(x) = \inf_{m\geq n} f_m(x)$. Note that g_n is measurable by the same methods used to show Corollary 29. Now, $g_n \uparrow \liminf_{n\to\infty} f_n$, so by the Monotone Convergence Theorem we have

$$\int g_n d\mu \uparrow \int \liminf_{n\to\infty} f_n d\mu.$$

But, we clearly have $g_n(x) \leq f_n(x)$ pointwise and therefore

$$\int \liminf_{n\to\infty} f_n d\mu = \liminf_{n\to\infty} \int g_n d\mu \leq \liminf_{n\to\infty} \int f_n d\mu$$

This completes the proof. ∎

Of course there can be strict inequality in Fatou's Lemma. It suffices to take

$$f_n(x) = \begin{cases} \mathbb{1}_{[-1,0[}(x) & \text{if } n \text{ is odd,} \\ \mathbb{1}_{[0,1[}(x) & \text{if } n \text{ is even.} \end{cases}$$

Then $\liminf_{n\to\infty} f_n(x) = 0$ for all x, but $\int f_n d\mu = 1$ for all n.

These results apply only to nonnegative functions, so we need something that will work for signed functions or for complex valued functions.

THEOREM 41 (DOMINATED CONVERGENCE THEOREM) *Let f_n be a sequence of measurable functions and suppose that $f_n \longrightarrow f$ pointwise. Further suppose that there is a (nonnegative) function g such that $|f_n| \leq g$ pointwise for every $n \in \mathbb{N}$. If $\int g d\mu < \infty$, then necessarily*

$$\int f_n d\mu \underset{n\to\infty}{\longrightarrow} \int f d\mu. \tag{2.8}$$

Proof. Obviously $|f| \le g$, so we have $|f - f_n| \le |f| + |f_n| \le 2g$ pointwise. Let $h_n = 2g - |f - f_n| \ge 0$ and apply Fatou's Lemma. This gives

$$\begin{aligned}
\int 2g d\mu &= \int \liminf_{n\to\infty} \Big(2g - |f - f_n|\Big) d\mu \\
&\le \liminf_{n\to\infty} \int 2g - |f - f_n| d\mu \\
&= \int 2g d\mu - \limsup_{n\to\infty} \int |f - f_n| d\mu,
\end{aligned}$$

the last step using part (ii) of Theorem 37. Because $\int 2g d\mu < \infty$, we can deduce that $\int |f - f_n| d\mu \longrightarrow 0$ as $n \longrightarrow \infty$. So, (2.8) also holds. ■

EXAMPLE We look at some examples of the Dominated Convergence Theorem in action. We'll assume here that we can evaluate Lebesgue integrals as Riemann integrals. Consider

$$f_n(x) = \left(1 + \frac{x}{n}\right)^{-n} \sin\left(\frac{x}{n}\right)$$

on $[0, \infty[$. The pointwise limit of $f_n(x)$ is zero, because $\left(1 + \frac{x}{n}\right)^{-n} \longrightarrow e^{-x}$ and $\sin\left(\frac{x}{n}\right) \longrightarrow 0$. But we need to bound f_n somehow to satisfy the hypotheses of the Dominated Convergence Theorem. We use two fairly basic facts

$$\left|\sin\left(\frac{x}{n}\right)\right| \le 1$$

and

$$\left(1 + \frac{x}{2}\right)^2 \le \left(1 + \frac{x}{n}\right)^n \qquad n \ge 2, x \ge 0$$

The second of these can be proved by observing from the Binomial Theorem that the right-hand member is increasing with n. Together, these inequalities tell us that $|f_n(x)| \le g(x)$ for $n \ge 2$ where

$$g(x) = \left(1 + \frac{x}{2}\right)^{-2}$$

But $\int_0^\infty g(x) dx = 2$ and we can apply the Dominated Convergence Theorem. The conclusion is that

$$\int_0^\infty \left(1 + \frac{x}{n}\right)^{-n} \sin\left(\frac{x}{n}\right) dx \underset{n\to\infty}{\longrightarrow} 0.$$

Note that in this example, $\int_0^\infty |f_1(x)|dx = \infty$. It is therefore essential to throw away the first term of the sequence. In the Lebesgue theory, $\int_0^\infty f_1(x)dx$ is not even defined. In the Riemann theory, it has to be treated as an indefinite integral. $\square$

EXAMPLE Let $f_n(x) = \dfrac{1+nx^2}{(1+x^2)^n}$ on $[0,1]$. Then it is evident that $f_n(x) \longrightarrow \mathbb{1}_{\{0\}}(x)$ as $n \longrightarrow \infty$. We also have $1+nx^2 \le (1+x^2)^n$ so that for all n and x, $|f_n(x) \le 1$. So, the Dominated Convergence Theorem applies and

$$\int_0^1 \frac{1+nx^2}{(1+x^2)^n}dx \longrightarrow \int_0^1 \mathbb{1}_{\{0\}}(x)dx = 0.$$

$\square$

EXAMPLE Let $f_n(x) = n\sin\left(\dfrac{x}{n}\right)\left(x(1+x^2)\right)^{-1} = \varphi\left(\dfrac{x}{n}\right)(1+x^2)^{-1}$, on the set $[0,\infty[$. We have denoted $\varphi(x) = \dfrac{\sin(x)}{x}$. Note that $|\varphi(x)| \le 1$ and that $\int_0^\infty (1+x^2)^{-1}dx < \infty$. So, by the Dominated Convergence Theorem, we deduce

$$\int_0^\infty n\sin\left(\frac{x}{n}\right)\frac{dx}{x(1+x^2)} \longrightarrow \int_0^\infty \frac{dx}{1+x^2} = \frac{\pi}{2}.$$

$\square$

EXAMPLE The final example is $\lim\limits_{n\to\infty}\displaystyle\int_a^\infty \frac{n}{1+n^2x^2}dx$ for $a \ge 0$. Using the Riemann theory and a change of variables, it is easy to see that the value of the limit is 0 if $a > 0$ and $\dfrac{\pi}{2}$ if $a = 0$. What happens if we try to use convergence theorems without using a change of variables? Well

$$f_n(x) = \frac{n}{1+n^2x^2} \longrightarrow \infty\mathbb{1}_{\{0\}}(x).$$

Is the convergence dominated? If $x \ge 1$ then $f_n(x)$ decreases with n, so $\dfrac{1}{1+x^2}$ is an upper bound in this range. If $0 \le x < 1$, then it can be shown that

$$\frac{1}{4x} \le \sup_{n=1}^{\infty} f_n(x) \le \frac{1}{2x}. \tag{2.9}$$

Thus, if $a > 0$, the Dominated Convergence Theorem can be successfully applied and if $a = 0$, then it cannot be.

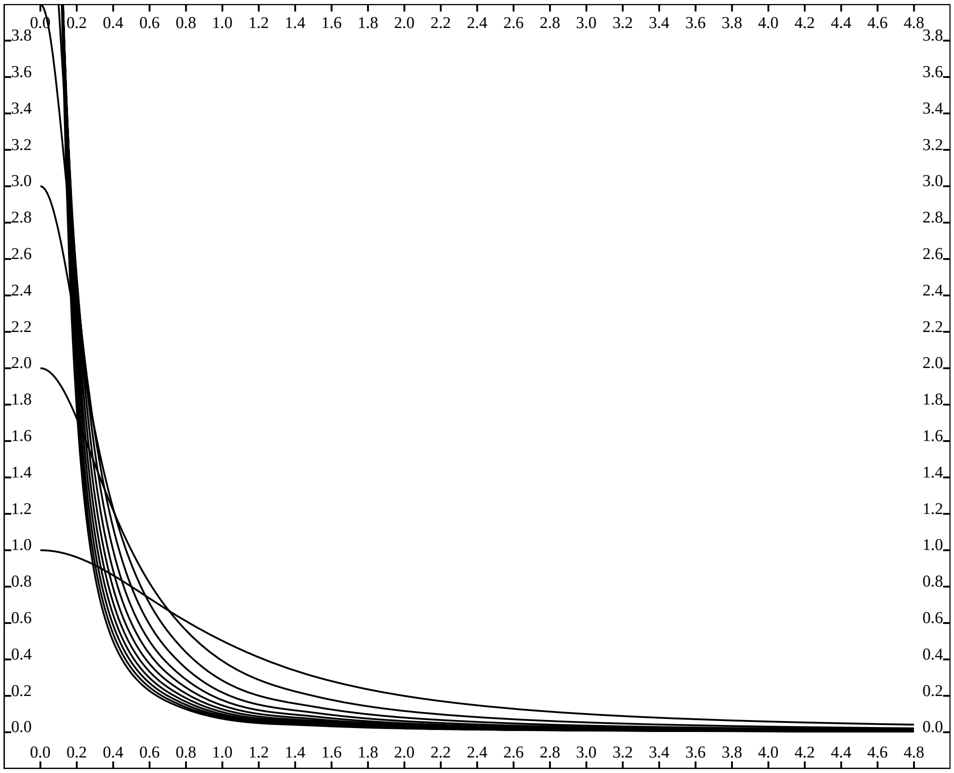

Figure 2.1: The functions f_n. Note that $\sup_n f_n(x)$ behaves like $f_1(x)$ for $x \geq 1$, but like Cx^{-1} for $0 < x \leq 1$.

To prove (2.9) the right-hand inequality follows from $(nx-1)^2 \geq 0$. For the left-hand inequality, we will take $n = \lfloor x^{-1} \rfloor$ in the sup. Then we have $n \geq 1$ and we can write $x = (n+t)^{-1}$ where $0 \leq t < 1$. It is then necessary to show that

$$\frac{n+t}{4} \leq \frac{n}{1 + \frac{n^2}{(n+t)^2}}$$

which boils down to $2n^2 + 2nt + t^2 \leq 4n^2 + 4nt$. This last inequality is satisfied because $2n^2 + t^2 \leq 2n^2 + 1 \leq 4n^2$ and since $2nt \leq 4nt$. $\square$

2.6 Riemann and Lebesgue Integrals

In the last section we used the Riemann theory of integration to compute our integrals. We start this section by justifying that.

THEOREM 42 *Let f be Riemann Integrable on $[a, b]$. Then, f is Lebesgue measurable on $[a, b]$ and the Riemann and Lebesgue integrals agree.*

Note that it is false in general that a Riemann integrable function is Borel. Let K be the Cantor Ternary set in $[0, 1]$ and let $S \subseteq K$ be a subset that is not Borel. Then $\mathbb{1}_S$ is Riemann integrable, but it is not a Borel function. We should also observe that there is a theorem of Lebesgue which characterizes Riemann integrable functions completely. We will not prove this theorem here.

THEOREM 43 *A bounded function f on $[a, b]$ is Riemann integrable on $[a, b]$ if and only if the set of points where f fails to be continuous has zero Lebesgue measure.*

Proof of Theorem 42. Without loss of generality, we can assume that $a = 0$ and that $b = 1$. Now we apply Theorem 71 from the notes from MATH 255. This implies that we can work with dyadic subintervals of $[0, 1]$. So, let us define

$$g_n(x) = \inf\{f(x); k2^{-n} \leq x \leq (k+1)2^{-n}\} \text{ for } k2^{-n} \leq x < (k+1)2^{-n}$$

and

$$h_n(x) = \sup\{f(x); k2^{-n} \leq x \leq (k+1)2^{-n}\} \text{ for } k2^{-n} \leq x < (k+1)2^{-n}$$

for each integer k with $0 \leq k < 2^n$. The quantities $g_n(1)$ and $h_n(1)$ not defined by the above equations can be defined to make g_n and h_n to be continuous on the left at 1. Now, we clearly have

$$g_n(x) \leq f(x) \leq h_n(x)$$

for all $x \in [0, 1]$. Furthermore the Riemann and Lebesgue integrals of g_n and h_n are easily seen to agree. The content of Theorem 71 is that

$$\int_0^1 g_n(x)dx \uparrow \int_0^1 f(x)dx \text{ and } \int_0^1 h_n(x)dx \downarrow \int_0^1 f(x)dx$$

as $n \longrightarrow \infty$. (The notation $\int_0^1 f(x)dx$ here is for the Riemann integral.) In fact, g_n is an increasing sequence of Lebesgue measurable functions and h_n is a decreasing such sequence. Of course Riemann integrable functions have to be bounded, so the functions g_n and h_n are uniformly bounded. Now, let $g = \sup g_n$ and $h = \inf h_n$, then, from the Dominated Convergence Theorem, we find that $\int_0^1 (h-g)(x)dx = 0$ and $h \geq g$. So g and h agree except on a null set. We also have $h \leq f \leq g$ and this is enough to show that f agrees with both g and h except on a null set. Hence f is Lebesgue measurable because g (or h) is. Again by Dominated Convergence, the Riemann and Lebesgue integrals coincide. ∎

3

L^p spaces

The L^p spaces are spaces of measurable functions on a measure space $(X, \mathcal{M}, \mu)$. Well, that's not entirely correct. For instance the space $L^1(X, \mathcal{M}, \mu)$ is the space of "functions" such that $\int |f| d\mu < \infty$ and we take $\|f\|_1 = \int |f| d\mu < \infty$ as the norm. It is easy to check all the standard inequalities and equalities for $\|\cdot\|_1$ to be a norm with one exception. If $\|f\|_1 = 0$, then we can only deduce that $f = 0$ μ-a.e. and that's different from f being identically zero. So we introduce an equivalence relation. Two functions f and g are viewed as being equivalent if and only if $f = g$ μ-a.e. or more precisely if and only if $\mu(\{x; f(x) \neq g(x)\}) = 0$. It's easy to see that this defines an equivalence relation and the equivalence classes are effectively functions defined up to a μ-null set. The elements of $L^1(X, \mathcal{M}, \mu)$ are then strictly speaking equivalence classes of functions rather than functions, and then $\|\cdot\|_1$ actually defines a norm. If $\tilde{f}$ is such an equivalence class, and f is a function in that equivalence class, f is called a ***version*** of $\tilde{f}$. We define $L^p(X, \mathcal{M}, \mu)$ for all values of p with $1 \leq p \leq \infty$. It consists of all equivalence classes of functions such that the p-norm is finite, the p-norm being given by

$$\|f\|_p = \left\{ \int |f|^p d\mu \right\}^{\frac{1}{p}}. \tag{3.1}$$

for $1 \leq p < \infty$. We'll define the infinity norm later. It's obvious that (3.1) actually defines a norm except for the subadditivity, which we'll verify eventually.

To define the infinity norm, we need a new concept called the essential supremum. We can define this by

$$\operatorname*{ess\,sup}_{x \in X} f(x) = \inf\{M; f(x) \leq M,\ \mu - \text{a.e.}\}.$$

This notion depends of course upon the measure. If there is possible confusion, we would use the notation μ-ess sup. The infimum defining the ess sup is actually attained. If ess sup $f = \infty$, then there is nothing to show, if on the other hand, it is finite, then there is a sequence M_k decreasing to ess sup f such that $f \le M_k$ μ-a.e. But then

$$\{x; f(x) > \text{ess sup} f\} \subseteq \bigcup_k \{x; f(x) \ge M_k\}.$$

The right hand side is a countable union of μ-null sets and hence μ-null. So actually $f(x) \le \operatorname{ess\,sup} f$ μ-a.e. Of course, one can also define the essential infimum in the same way.

Now we define the infinity norm by

$$\|f\|_\infty = \operatorname*{ess\,sup}_{x \in X} |f(x)|.$$

It's clear that functions that are equal almost everywhere have the same infinity norm. It's also clear that if $\|f\|_\infty < \infty$, then f has a version g that is actually bounded and indeed such that $\sup |g(x)| = \text{ess sup} |f|$. It's easy to show that the infinity norm is in fact a norm.

One can also define $\|f\|_p$ if $0 < p < 1$, but it's not a norm. The would be unit ball $\{f; \|f\|_p \le 1\}$ is no longer convex (except under really trivial circumstances).

In the $p \ge 1$ case, there's a very important concept called the conjugate index. We define $p' = \dfrac{p}{p-1}$. We have $1' = \infty$, $2' = 2$ and $\infty' = 1$. As p increases, p' decreases.

THEOREM 44 (HÖLDER'S INEQUALITY) *Let* $1 \le p \le \infty$. *Then*

$$\left| \int fg d\mu \right| \le \|f\|_p \|g\|_{p'}$$

provided that the right hand side is finite.

Proof. If $p = 1$ or $p = \infty$, then the result is straightforward. We will assume that $1 < p < \infty$, so that also $1 < p' < \infty$. It will be enough to show that $\int |fg| d\mu \le \|f\|_p \|g\|_{p'}$. So, without loss of generality, f and g are nonnegative functions and we need to show

$$\int fg d\mu \le \left\{ \int f^p d\mu \right\}^{\frac{1}{p}} \left\{ \int g^{p'} d\mu \right\}^{\frac{1}{p'}}$$

We can assume that the right hand side is finite. Also, after renormalizing, we can assume that

$$\int f^p d\mu = 1 \text{ and } \int g^{p'} d\mu = 1.$$

But now, we use the inequality $xy \leq \dfrac{x^p}{p} + \dfrac{y^{p'}}{p'}$ for $x, y \geq 0$ to obtain

$$\int fg d\mu \leq \frac{1}{p}\int f^p d\mu + \frac{1}{p'}\int g^{p'} d\mu = \frac{1}{p} + \frac{1}{p'} = 1$$

and we are done. ∎

COROLLARY 45 (MINKOWSKI'S INEQUALITY) *Let $1 \leq p \leq \infty$. Then*

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

Proof. Again, it's enough to show the result for nonnegative functions. We now proceed as follows

$$\begin{aligned}
\int (f+g)^p d\mu &= \int (f+g)(f+g)^{p-1} d\mu, \\
&= \int f(f+g)^{p-1} d\mu + \int g(f+g)^{p-1} d\mu, \\
&\leq \left\{\int f^p d\mu\right\}^{\frac{1}{p}} \left\{\int (f+g)^{(p-1)p'} d\mu\right\}^{\frac{1}{p'}}, \\
&\quad + \left\{\int g^p d\mu\right\}^{\frac{1}{p}} \left\{\int (f+g)^{(p-1)p'} d\mu\right\}^{\frac{1}{p'}},
\end{aligned}$$

by applying Hölder's inequality to each term,

$$\leq \Big(\|f\|_p + \|g\|_p\Big) \left\{\int (f+g)^p d\mu\right\}^{\frac{1}{p'}},$$

since $(p-1)p' = p$. Now, if $\int (f+g)^p d\mu = 0$, then f and g vanish almost everywhere and the result is easy. Otherwise $\int (f+g)^p d\mu > 0$ and it is legal to divide off giving

$$\left\{\int (f+g)^p d\mu\right\}^{\frac{1}{p}} \leq \Big(\|f\|_p + \|g\|_p\Big)$$

as required. ∎

This Corollary fills the gaps so that we know that $\|\cdot\|$ is a norm.

3.1 Completeness of the L^p spaces

Next comes the question of completeness. It is easiest to understand this first in the context of L^1. Usually in metric space theory, we use completeness to derive the Weierstrass M-test.

PROPOSITION 46 *Let V be a complete normed space and let v_j be elements of V for $j \in \mathbb{N}$. Suppose that*

$$\sum_{j=1}^{\infty} \|v_j\|_V < \infty.$$

Then the sequence of partial sums (s_n) given by

$$s_n = \sum_{j=1}^{n} v_j$$

converges to an element $s \in V$. Furthermore we have the norm estimate

$$\|s\|_V \le \sum_{j=1}^{\infty} \|v_j\|_V. \tag{3.2}$$

What we are going to do here is the reverse. We will use the M-test to get at the completeness of L^1.

PROPOSITION 47 *Let $(X, \mathcal{M}, \mu)$ be a measure space and let $f_n \in L^1(X, \mathcal{M}, \mu)$ be such that $\sum_{n=1}^{\infty} \|f_n\|_1 < \infty$. Then the series $\sum_{n=1}^{\infty} f_n$ converges in L^1.*

Proof. We have

$$\begin{aligned} \int \sum_{n=1}^{\infty} |f_n| d\mu &= \int \sup_N \sum_{n=1}^{N} |f_n| d\mu \\ &= \sup_N \int \sum_{n=1}^{N} |f_n| d\mu \end{aligned}$$

by the Monotone Convergence Theorem

$$\begin{aligned} &= \sup_N \sum_{n=1}^{N} \int |f_n| d\mu \\ &= \sum_{n=1}^{\infty} \int |f_n| d\mu < \infty \end{aligned} \tag{3.3}$$

This means that $N = \{x; \sum_{n=1}^{\infty} |f_n(x)| = \infty\}$ must be a null set. Or, we could say that $\sum_{n=1}^{\infty} |f_n| < \infty$ almost everywhere. So, for almost all x the series $\sum_{n=1}^{\infty} f_n(x)$ converges absolutely. Let $s(x)$ be the sum of this series on N^c and zero on N. Then s is a measurable function. We clearly have that $|s(x)| \leq \sum_{n=1}^{\infty} |f_n(x)|$ and (3.3) shows that $\int |s| d\mu < \infty$. So, s is in L^1. It remains to show that $s = \sum_{n=1}^{\infty} f_n$ in L^1. Let $\epsilon > 0$ and choose M so large that $\sum_{n=M+1}^{\infty} \|f_n\|_1 < \epsilon$. Then, we have

$$\int \sum_{n=M+1}^{\infty} |f_n| d\mu = \int \sup_N \sum_{n=M+1}^{N} |f_n| d\mu$$

$$= \sup_N \int \sum_{n=M+1}^{N} |f_n| d\mu$$

by the Monotone Convergence Theorem

$$= \sup_N \sum_{n=M+1}^{N} \int |f_n| d\mu$$

$$= \sum_{n=M+1}^{\infty} \int |f_n| d\mu < \epsilon \tag{3.4}$$

just as before, and it follows from (3.4) that

$$\int \Big| s - \sum_{n=1}^{m} f_n \Big| d\mu = \int \Big| \sum_{n=m+1}^{\infty} f_n \Big| d\mu \leq \int \sum_{n=m+1}^{\infty} |f_n| d\mu \leq \int \sum_{n=M+1}^{\infty} |f_n| d\mu < \epsilon$$

for $m \geq M$. This asserts that we have convergence in L^1 norm. ∎

COROLLARY 48 *Let $(X, \mathcal{M}, \mu)$ be a measure space. Then $L^1(X, \mathcal{M}, \mu)$ is a complete normed space.*

Proof. We exploit the idea of "rapid convergence". Let (f_n) be a Cauchy sequence in $L^1(X, \mathcal{M}, \mu)$. Then, there exists n_k such that

$$p, q \geq n_k \qquad \Longrightarrow \qquad \|f_p - f_q\| < 2^{-k} \tag{3.5}$$

We have (taking $p = n_{k+1}$ and $q = n_k$ in (3.5)) that $\sum_{k=1}^{\infty} \|f_{n_{k+1}} - f_{n_k}\| < \infty$, so that the series $\sum_{k=1}^{\infty} \left(f_{n_{k+1}} - f_{n_k}\right)$ converges in L^1 norm say to a function s.

Let $f = s + f_{n_1}$. Then, it is clear that $f_{n_k} \longrightarrow f$ as $k \longrightarrow \infty$. Now we go back to the Cauchy condition to capture the convergence of the original sequence. So, given $\epsilon > 0$ we find k so large that $2^{-k} < \epsilon/2$ and $\|f - f_{n_k}\| < \epsilon/2$. Then we find taking $q = n_k$ in (3.5) that

$$p \geq n_k \qquad \Longrightarrow \qquad \|f_p - f_{n_k}\| < \epsilon/2 \Longrightarrow \qquad \|f_p - f\| < \epsilon$$

thus showing convergence. ∎

The same proofs work for L^p when $1 < p < \infty$, but the proofs are not as clean. They involve a few intermediate steps.

PROPOSITION 49 *Let $(X, \mathcal{M}, \mu)$ be a measure space, $1 \leq p < \infty$ and let $f_n \in L^p(X, \mathcal{M}, \mu)$ be such that $\sum_{n=1}^{\infty} \|f_n\|_p < \infty$. Then the series $\sum_{n=1}^{\infty} f_n$ converges in L^p.*

Proof. We have

$$\begin{aligned}
\left\| \sum_{n=1}^{\infty} |f_n| \right\|_p^p &\leq \int \left(\sum_{n=1}^{\infty} |f_n| \right)^p d\mu \\
&= \int \sup_N \left(\sum_{n=1}^{N} |f_n| \right)^p d\mu \\
&= \sup_N \int \left(\sum_{n=1}^{N} |f_n| \right)^p d\mu
\end{aligned}$$

Now take pth roots to get

$$\begin{aligned}
\left\| \sum_{n=1}^{\infty} |f_n| \right\|_p &= \sup_N \left\| \sum_{n=1}^{N} |f_n| \right\|_p \\
&\leq \sup_N \sum_{n=1}^{N} \| \, |f_n| \, \|_p \\
&= \sum_{n=1}^{\infty} \| \, |f_n| \, \|_p < \infty
\end{aligned}$$

using the extended version of Minkowski's Inequality. We can now rewrite this as

$$\int \left(\sum_{n=1}^{\infty} |f_n| \right)^p d\mu < \infty$$

and it implies as before that $\sum_{n=1}^{\infty} |f_n| < \infty$ almost everywhere. So, for almost all x the series $\sum_{n=1}^{\infty} f_n(x)$ converges absolutely. Let $s(x)$ be defined as in the L^1 case. Then s is a measurable function satisfying $|s(x)| \le \sum_{n=1}^{\infty} |f_n(x)|$. We see that $\int |s|^p d\mu < \infty$. So, s is in L^p. Let $\epsilon > 0$ and choose M so large that $\sum_{n=M}^{\infty} \|f_n\|_p < \epsilon$. Then, we have, essentially repeating the steps above

$$\begin{aligned} \left\| \sum_{n=M+1}^{\infty} |f_n| \right\|_p &= \sup_N \| \sum_{n=M+1}^{N} |f_n| \, \|_p \\ &\le \sup_N \sum_{n=M+1}^{N} \| \, |f_n| \, \|_p \\ &= \sum_{n=M+1}^{\infty} \| \, |f_n| \, \|_p < \epsilon \end{aligned}$$

and it follows that

$$\left\| s - \sum_{n=1}^{M} f_n \right\|_p \le \left\| \sum_{n=M+1}^{\infty} |f_n| \right\|_p < \epsilon$$

showing that $s = \sum_{n=1}^{\infty} f_n$ where convergence is taken in L^p norm. ∎

COROLLARY 50 *Let $(X, \mathcal{M}, \mu)$ be a measure space and $1 \le p < \infty$. Then $L^p(X, \mathcal{M}, \mu)$ is a complete normed space.*

The proof is as for Corollary 48 above and we omit it.

PROPOSITION 51 *Let $(X, \mathcal{M}, \mu)$ be a measure space, let $f_n \in L^\infty(X, \mathcal{M}, \mu)$ be such that $\sum_{n=1}^{\infty} \|f_n\|_\infty < \infty$. Then the series $\sum_{n=1}^{\infty} f_n$ converges in L^∞.*

COROLLARY 52 *Let $(X, \mathcal{M}, \mu)$ be a measure space. Then $L^\infty(X, \mathcal{M}, \mu)$ is a complete normed space.*

Proof of the Proposition. We simply choose g_n to be a version of f_n which satisfies $\sup |g_n| \leq \|f_n\|_\infty$. Then $\sum_{n=M+1}^\infty |g_n| \leq \sum_{n=M+1}^\infty \|f_n\|_\infty$ for $M \in \mathbb{Z}^+$. We set $s(x) = \sum_{n=1}^\infty g_n(x)$ and the rest of the proof follows by the same methods as found earlier in this section. ■

3.2 L^2 as an inner product space

The space L^2 holds a very special position among the L^p spaces because it can be given the structure of an inner product space.

THEOREM 53 *The form*

$$\langle f, g \rangle = \int \overline{f} g d\mu$$

defines an inner product on $L^2(X, \mathcal{M}, \mu)$ which is compatible with the L^2 norm.

The proof is completely straightforward, the key point being that the associated norm of the inner product is just the L^2 norm.

$$\langle f, f \rangle = \int \overline{f} f d\mu = \int |f|^2 d\mu = \|f\|_2^2.$$

A complete inner product space is called a ***Hilbert space*** .

3.3 Dense subsets of L^p

PROPOSITION 54 *Let $1 \leq p < \infty$. Then the bounded functions carried on sets of finite measure are dense in L^p.*

Proof. There are two separate ideas here. The first is to show how to approximate a L^p function by a bounded function. Let $f \in L^p$. Define

$$f_n(x) = \begin{cases} f(x) & \text{if } |f(x)| \leq n, \\ 0 & \text{if } |f(x)| > n. \end{cases}$$

Then $|f - f_n| \leq |f|$ for all $n \in \mathbb{N}$. In particular, the functions $h_n = |f - f_n|^p \leq |f|^p$ are dominated by a single integrable function, namely $|f|^p$. Since f takes finite values almost everywhere, $f_n \longrightarrow f$ almost everywhere and therefore $h_n \longrightarrow 0$ almost everywhere. Therefore, by the Dominated Convergence Theorem, we find that $\|f - f_n\|^p = \int h_n d\mu \longrightarrow 0$. This means that there is a bounded function g as close as we like to f and we can arrange that $|g| \leq |f|$.

The first idea was to truncate the function where it was large. To approximate a function by another function which is carried on a set of finite measure we truncate the function where it is small. Let's start again with a function f in L^p. Now define

$$f_n(x) = \begin{cases} f(x) & \text{if } |f(x)| > n^{-1}, \\ 0 & \text{if } |f(x)| \leq n^{-1}. \end{cases}$$

Again, we can set $h_n = |f - f_n|^p \leq |f|^p$ and these functions are dominated by a single integrable function. The f_n are carried by a set of finite measure. This is a consequence of the Tchebychev Inequality. Let $A_n = \{x; |f(x)| > n^{-1}\}$. Then

$$\int \frac{1}{n^p} \mathbb{1}_{A_n} d\mu \leq \int |f|^p \mathbb{1}_{A_n} d\mu \leq \int |f|^p d\mu = \|f\|_p^p.$$

The left-hand inequality holds because $\dfrac{1}{n} < |f(x)|$ for $x \in A_n$. We clearly have $\mu(A_n) \leq n^p \|f\|_p^p$, so A_n definitely has finite measure. We have that $f_n \longrightarrow f$ pointwise. If $f(x) = 0$, then $f_n(x) = 0$ for all n. On the other hand, if $f(x) \neq 0$ then eventually $f_n(x) = f(x)$. As before, we find that $\|f - f_n\|^p = \int h_n d\mu \longrightarrow 0$.

To complete the proof, we simply need to observe that these two ideas do not interfere with each other. If we first truncate the function to make it bounded, then the second truncation where the function is small does not affect the boundedness. ■

PROPOSITION 55 *If $1 \leq p < \infty$, then simple functions s of the type $s = \sum_{k=1}^n a_k \mathbb{1}_{A_k}$ with $a_k \in \mathbb{C}$, $A_k \in \mathcal{M}$ and $\mu(A_k) < \infty$ are dense in L^p.*

Proof. Let $f \in L^p$ and $\epsilon > 0$. We need to find a function s of the required type with $\|f - s\|_p < \epsilon$. By Proposition 54, we can assume that f is bounded and carried on a set of finite measure A. For $n \in \mathbb{N}$ define

$$\beta_n(x + iy) = 2^{-n}(\lfloor 2^n x \rfloor + i \lfloor 2^n y \rfloor), \qquad x, y \text{ real}$$

Then $\beta_n : \mathbb{C} \longrightarrow \mathbb{C}$ is a Borel map. Also $|z - \beta_n(z)| \le 2^{-n}\sqrt{2}$ and $\beta_n(0) = 0$. We find that $\beta_n \circ f \longrightarrow f$ uniformly as $n \longrightarrow \infty$. Also $\beta_n \circ f$ is carried by the set A and takes only finitely many values. Hence it is a measurable step function of the required form. The uniform convergence on the set of finite measure A implies L^p convergence.

$$\int |f - \beta_n \circ f|^p d\mu = \int |f - \beta_n \circ f|^p \mathbb{1}_A d\mu \le \mu(A) \sup_{x \in X} |f(x) - \beta_n \circ f(x)|^p.$$

■

In the same vein we also have

PROPOSITION 56 *Simple functions s of the type $s = \sum_{k=1}^n a_k \mathbb{1}_{A_k}$ with $a_k \in \mathbb{C}$, $A_k \in \mathcal{M}$ are dense in L^∞.*

One cannot impose that $\mu(A_k)$ is finite unless one also knows that $\mu(X)$ is finite.

We now work specifically on $L^p(\mathbb{R}, \mathcal{L}, \nu)$ where $\mathcal{L}$ is the Lebesgue σ-field of $\mathbb{R}$ and ν is Lebesgue measure. Many of the results that we prove can be extended to other similar situations. Before we start, let's just point out that $L^p(\mathbb{R}, \mathcal{L}, \nu)$ is the same as $L^p(\mathbb{R}, \mathcal{B}_{\mathbb{R}}, \nu)$. In other words, every Lebesgue measurable function has a Borel version. We leave this as an exercise.

We denote by $C_c(\mathbb{R})$, the space of continuous functions of compact support on $\mathbb{R}$. It may seem strange that we can view $C_c(\mathbb{R})$ as a linear subspace of $L^p(\mathbb{R}, \mathcal{L}, \nu)$ because the first space consists of functions and the second consists of equivalence classes of functions, but there is actually no difficulty here. The obvious "inclusion" mapping $C_c(\mathbb{R}) \longrightarrow L^p(\mathbb{R}, \mathcal{L}, \nu)$ is easily seen to be injective. If a function in $C_c(\mathbb{R})$, maps to the zero element of $L^p(\mathbb{R}, \mathcal{L}, \nu)$, then it is in fact identically zero. A continuous function that vanishes almost everywhere must vanish identically. To see this, suppose that g is continuous on $\mathbb{R}$ and that $g(x) \neq 0$. Then the set $U = \{y; |g(y)| > \frac{1}{2}|g(x)|\}$ is an open subset of $\mathbb{R}$ containing x. So, U contains an interval and consequently has positive measure. But g does not vanish on U, so g is not almost everywhere zero.

3.4 Duality between L^p and $L^{p'}$

In this section we are going to establish a dual-type result for L^p. To establish the full duality theory between L^p and $L^{p'}$ is beyond the scope of this course. Nevertheless, dual-type arguments are very useful. Here is the theorem that we would like to prove.

THEOREM 57 *Let $1 \leq p \leq \infty$. Let $(X, \mathcal{M}, \mu)$ be a σ-finite measure space and $f : X \longrightarrow [0, \infty]$ be measurable function and such that*

$$\int f g d\mu \leq \|g\|_{p'}$$

for every positive measurable function g. Then $\|f\|_p \leq 1$.

Proof. The case $p = 1$ comes immediately from $g = \mathbb{1}$. The case $p = \infty$ will be proved separately. We therefore assume that $1 < p < \infty$. Let us prove the result under the additional hypothesis that f is bounded and carried on a set of finite measure. Then let $g(x) = (f(x))^{p-1}$. Then we find that $g \in L^{p'}$ and $\|g\|_{p'} = \|f\|_p^{p-1}$. So

$$\|f\|_p^p = \int f^p d\mu = \int f g d\mu \leq \|g\|_{p'} = \|f\|_p^{p-1}.$$

Since we know that $\|f\|_p < \infty$, we deduce that $\|f\|_p \leq 1$.

Next we remove the condition that f is bounded and carried on a set of finite measure. So, for general f find a sequence f_n bounded and carried on sets of finite measure, such that $f_n \uparrow f$ pointwise. For example, we can take $f_n(x) = \mathbb{1}_{X_n}(x) \min(n, f(x))$ where X_n are measurable subsets increasing to X. Clearly

$$\int f_n g d\mu \leq \int f g d\mu \leq \|g\|_{p'},$$

so $\|f_n\|_p \leq 1$. Finally, applying the Monotone Convergence Theorem

$$\int |f_n|^p d\mu \uparrow \int |f|^p d\mu$$

gives the desired conclusion.

The case $p = \infty$, still needs special attention. Let $t > 1$ and let $A_t = \{x; f(x) \geq t\}$. Let $g = \mathbb{1}_{A_t \cap X_n}$ and obtain

$$t\mu(A_t \cap X_n) \leq \int f g d\mu \leq \|g\|_1 = \mu(A_t \cap X_n).$$

The only way out is that $\mu(A_t \cap X_n) = 0$. Letting n tend to infinity, we get $\mu(A_t) = 0$. Now take a sequence of t's decreasing to 1 to see that $\{x; f(x) > 1\}$ is a null set. This says that $\|f\|_\infty \leq 1$. ∎

Note that Theorem 57 can be proved with the σ-finiteness assumption removed and the hypotheses $1 \leq p < \infty$ and $f < \infty$ μ-a.e. added. However the strong duality theorem requires σ-finiteness.

3.5 Interplay between Measure and Topology

In this section, we look at some results that are special because they depend heavily on the topology of the underlying space. The following definition is nonstandard, but very useful for discussing regularity in a course at this level.

DEFINITION *Let X be a metric space. The X is LCSC (locally compact and σ-compact) if there is a chain*

$$K_1 \subseteq \Omega_1 \subseteq K_2 \subseteq \Omega_2 \subseteq K_3 \subseteq \Omega_3 \subseteq \cdots \subseteq X$$

where

- K_n *is compact for* $n \in \mathbb{N}$.
- Ω_n *is open for* $n \in \mathbb{N}$.
- $\displaystyle\bigcup_{n=1}^{\infty} K_n = X$.

Clearly $\mathbb{R}$ is LCSC. It suffices to take

$$K_n = [-2n, 2n] \text{ and } \Omega_n =]-2n-1, 2n+1[.$$

Also any compact metric space is LCSC. We are also interested in Borel measures μ on X with the property that

$$\mu(K) < \infty \text{ for all } K \text{ compact } \subseteq X. \tag{3.6}$$

Lebesgue measure is an example of such a Borel measure (when restricted to the Borel subsets).

THEOREM 58 *Let X be a LCSC metric space and μ a Borel measure on X with the property (3.6). Then μ is* ***regular*** *in the sense that if $\epsilon > 0$ and B is a Borel subset of X with $\mu(B) < \infty$, then there exists K compact $\subseteq B \subseteq U$ open, such that $\mu(U \setminus K) < \epsilon$.*

Proof. This is not an easy proof. We work first under the additional assumption that X is compact. A Borel subset B (now necessarily of finite measure) is said to be ***approximable*** if for every $\epsilon > 0$, there exists K compact and U open, such that $K \subseteq B \subseteq U$ and $\mu(U \setminus K) < \epsilon$. Let $\mathcal{A}$ be the collection of all approximable Borel sets. Then we aim to show two facts.

- Every open subset is in $\mathcal{A}$.
- $\mathcal{A}$ is a λ-system.

Since the collection of open subsets is a π-system, the result follows from Dynkin's π-λ Theorem.

The first assertion is a consequence of the fact that every open subset is an F_σ. To see this, we write an arbitrary open subset U as

$$U = \bigcup_{k=1}^{\infty} \{x; \operatorname{dist}_{U^c}(x) \geq 2^{-k}\}$$

so that

$$\sup_{k=1}^{\infty} \mu\Big(\{x; \operatorname{dist}_{U^c}(x) \geq 2^{-k}\}\Big) = \mu(U) < \infty.$$

In the second assertion, it is routine to show that (λ_1) and (λ_2) hold. The only tricky part is to show (λ_3) that if (A_j) is a sequence of disjoint approximable sets, then the union is also approximable. Let $\epsilon > 0$. Since $\displaystyle\sum_{j=1}^{\infty} \mu(A_j) \leq \mu(X) < \infty$, there exists J such that $\displaystyle\sum_{j=J+1}^{\infty} \mu(A_j) < \tfrac{1}{2}\epsilon$. We find L_j compact and U_j open such that $L_j \subseteq A_j \subseteq U_j$ and $\mu(U_j \setminus L_j) < 2^{-1-j}\epsilon$. It now suffices to take $L = \bigcup_{j=1}^{J} L_j$ compact and $U = \bigcup_{j=1}^{\infty} U_j$ open. Clearly $L \subseteq \bigcup_{j=1}^{\infty} A_j \subseteq U$. Now

$$\bigcup_{j=1}^{\infty} U_j \setminus \bigcup_{j=1}^{J} L_j \subseteq \bigcup_{j=1}^{\infty}(U_j \setminus L_j) \cup \bigcup_{j=J+1}^{\infty} A_j$$

since

$$\bigcup_{j=1}^{\infty} U_j \subseteq \bigcup_{j=1}^{\infty}(U_j \setminus L_j) \cup \bigcup_{j=1}^{\infty} L_j \subseteq \bigcup_{j=1}^{\infty}(U_j \setminus L_j) \cup \bigcup_{j=1}^{J} L_j \cup \bigcup_{j=J+1}^{\infty} A_j.$$

Therefore

$$\mu(U \setminus L) \leq \sum_{j=1}^{\infty} \mu(U_j \setminus L_j) + \sum_{j=J+1}^{\infty} \mu(A_j) < \tfrac{1}{2}\epsilon + \tfrac{1}{2}\epsilon = \epsilon.$$

This completes the proof in the compact case.

Now we tackle the general case where X is LCSC. Let B be a Borel subset of X with $\mu(B) < \infty$. Let $\epsilon > 0$. Then we can find n so large that $\mu(B \cap K_n^c) < \frac{1}{2}\epsilon$. Applying the result for the compact case, we can also find L compact with $L \subseteq B \cap K_n \subseteq B$ and $\mu((B \cap K_n) \setminus L) < \frac{1}{2}\epsilon$. It follows that $\mu(B \setminus L) < \epsilon$ finishing the verification of the inner regularity.

For the outer regularity, we first observe that we can approximate K_n from the outside by open sets. Let $\epsilon > 0$ then since K_n is approximable in K_{n+1} there exists W open in K_{n+1} such that $\mu(W \setminus K_n) < \epsilon$ and $K_n \subseteq W$. From the characterization of relatively open subsets, there exists $\widetilde{W}$ open in X such that $\widetilde{W} \cap K_{n+1} = W$. But $K_n \subseteq \Omega_n$ open $\subseteq K_{n+1}$, so that $\widetilde{W} \cap \Omega_n$ is open in X and $K_n \subseteq \widetilde{W} \cap \Omega_n = W \cap \Omega_n \subseteq W$. This completes the claim that K_n is approximable.

So for $\epsilon > 0$, we can find V_n open $\supseteq K_n$ with $\mu(V_n \setminus K_n) < 2^{-1-n}\epsilon$. Now apply the established regularity of $B \cap K_n$ to find a relatively open subset W_n of K_n such that $\mu(W_n \setminus (K_n \cap B)) < 2^{-1-n}\epsilon$ and $W_n \supseteq K_n \cap B$. We can find an open subset $\widetilde{W_n}$ (open in X) such that $\widetilde{W_n} \cap K_n = W_n$. Then define $U_n = \widetilde{W_n} \cap V_n$. This is open in X and

$$\mu(U_n \setminus (K_n \cap B)) \leq \mu(W_n \setminus (K_n \cap B)) + \mu(V_n \setminus K_n) < 2^{-n}\epsilon$$

Finally, letting $U = \bigcup_{n=1}^{\infty} U_n$ gives $(\mu \times \nu)(U \setminus B) < \epsilon$ and $B \subseteq U$ as required.■

A detailled analysis of the proof shows the following corollary.

COROLLARY 59 *Let X be a LCSC metric space and μ a Borel measure on X with the property (3.6). For every $\epsilon > 0$ and every Borel subset B of X, there exists E closed $\subseteq B \subseteq U$ open, such that $\mu(U \setminus E) < \epsilon$.*

PROPOSITION 60 *Let X be a LCSC metric space and μ a regular Borel measure on X. For $1 \leq p < \infty$, $C_c(X)$ is dense in $L^p(X, \mathcal{B}_X, \mu)$.*

Proof. We use Proposition 55. It's enough to show that if $A \in \mathcal{B}_X$ is of finite measure, then $\mathbb{1}_A$ can be approximated in L^p norm by continuous functions of compact support.

The first step is to see that we can assume without loss of generality that there exists n such that $A \subseteq K_n$. We have $\mu(A \cap K_n) \uparrow \mu(A)$ as $n \to \infty$ since $\bigcup_{n=1}^{\infty} A \cap K_n = A$. Further, since $\mu(A) < \infty$, we can deduce that $\mu(A \cap K_n^c)$ is as small as we please for n large enough. Since $1 \le p < \infty$, this implies that $\mathbb{1}_{A \cap K_n^c}$ has small L^p norm. The remainder of A, namely $A \cap K_n$ is a subset of K_n.

Now, let $\epsilon > 0$. Then, using the regularity of μ, we can find a compact subset C of X and an open subset U of X such that $C \subseteq A \subseteq U$ and $\mu(U \setminus C) < \epsilon$. Let $V = U \cap \Omega_{n+1}$. Then V is open with compact closure and $C \subseteq A \subseteq V$ and $\mu(V \setminus C) < \epsilon$. Now, find a continuous function φ equal to 1 on C and equal to 0 on $X \setminus V$. This is a consequence of the Tietze Extension Theorem, or it can be done by combining distance functions

$$\varphi(x) = \frac{\mathrm{dist}_{V^c}(x)}{\mathrm{dist}_{V^c}(x) + \mathrm{dist}_C(x)}.$$

If it's done this way, its easy to see that φ takes values in $[0, 1]$ otherwise, this has to be arranged. Now $|\mathbb{1}_A - \varphi| \le 1$ and $\mathbb{1}_A - \varphi$ is carried on the set $V \setminus C$ of measure less than ϵ, so we have $\|\mathbb{1}_A - \varphi\| < \epsilon^{1/p}$. The proof is complete. Note that the only purpose of introducing the set V is to ensure that φ has compact support. ■

COROLLARY 61 *For $1 \le p < \infty$, translation is continuous on $L^p(\mathbb{R}, \mathcal{L}, \nu)$.*

In fact, we define the translation operator T_t on $L^p(\mathbb{R}, \mathcal{L}, \nu)$ for $t \in \mathbb{R}$. The definition is

$$(T_t(f))(x) = f(x - t).$$

It is obvious that T_t is an isometric linear operator on L^p for $1 \le p \le \infty$. What we are asserting here is that if $1 \le p < \infty$, then for a fixed function $f \in L^p(\mathbb{R}, \mathcal{L}, \nu)$, we have that $T_t(f) \longrightarrow f$ in L^p norm as $t \longrightarrow 0$ in $\mathbb{R}$.

Proof. Let $\epsilon > 0$ and $f \in L^p(\mathbb{R}, \mathcal{L}, \nu)$. Find $\varphi \in C_c(\mathbb{R})$ such that $\|f - \varphi\|_p < \epsilon/3$. Then, also $\|T_t(f) - T_t(\varphi)\|_p = \|T_t(f - \varphi)\|_p = \|f - \varphi\|_p < \epsilon/3$. So, it is enough to show that for t small, we have $\|T_t(\varphi) - \varphi\|_p < \epsilon/3$. Since φ is both continuous and compactly supported, it is uniformly continuous. Let the support of φ be contained in $[-n, n]$. We define

$$\kappa = \frac{\epsilon}{3(2n+2)^{1/p}} > 0.$$

Now, there exists $\delta_1 > 0$ such that $|t| < \delta_1$ implies that $\|T_t(\varphi) - \varphi\|_\infty < \kappa$. This is just the uniform continuity of φ. Let us set $\delta = \min(1, \delta_1) > 0$ and then the support of $T_t(\varphi) - \varphi$ is in the interval $[-n - 1, n + 1]$. It follows that

$$\|T_t(\varphi) - \varphi\|_p^p \le \|T_t(\varphi) - \varphi\|_\infty^p (2n + 2)$$

and the desired result follows. ■

It is of course false that the operator T_t converges to the identity operator in the operator norm on L^p. This would be something completely different. It is also false that translation is continuous on L^∞. To see this, let $f = \mathbb{1}_{[0,\infty[}$. Then, for $t > 0$, we have

$$T_t(f) - f = \mathbb{1}_{[t,\infty[} - \mathbb{1}_{[0,\infty[} = -\mathbb{1}_{[0,t[}$$

and so $\|T_t(f) - f\|_\infty = 1$ for all $t > 0$ no matter how small.

THEOREM 62 (LUSIN'S THEOREM) *Let X be a LCSC metric space and μ a regular Borel measure on X. Let f be a Borel measurable complex-valued function on X, zero outside a set of finite μ measure. Let $\epsilon > 0$. Then there is a function $g \in C_c(X)$, such that $\mu(\{x; f(x) \neq g(x)\}) < \epsilon$.*

In particular, Lusin's Theorem holds when $X = \mathbb{R}$ and when μ is Lebesgue measure.

Proof. The first step is to see that we can reduce to the specific case where X is compact. Let $Y = \{x; x \in X, f(x) \neq 0\}$. Then Y has finite measure, but it may not be contained in a compact set. However $\mu(Y \cap K_n) \uparrow \mu(Y)$ as $n \to \infty$. Since $\mu(Y) < \infty$, we can deduce that $\mu(Y \cap K_n^c) < \frac{1}{3}\epsilon$ for n large enough. We leave the reader to check that the restriction of μ to K_n is still a regular Borel measure and, since we are assuming the result in the compact case, there is a continuous function $h : K_n \longrightarrow \mathbb{C}$ such that $\mu(\{x \in K_n; f(x) \neq h(x)\}) < \frac{1}{3}\epsilon$. Now apply the regularity of μ to K_n itself. This guarantees the existence of Ω open with $\mu(\Omega \setminus K_n) < \frac{1}{3}\epsilon$. Furthermore, we can always assume that $\Omega \subseteq \Omega_n$. Now extend h to $K_n \cup \Omega^c$ by setting h to be zero on Ω^c. Then the extended h is also continuous. We can now extend h to a continuous function g on the whole of X by the Tietze Extension Theorem. Since $\Omega \subseteq \Omega_n \subset K_{n+1}$, we see that g has compact support. It is easy to see that f and g agree except on the union of three sets each of measure controlled by $\frac{1}{3}\epsilon$ and the result follows.

So now we can assume that X is compact. Take the function f, split it into real and imaginary parts, then nonnegative and nonpositive parts. Without loss of generality, we can assume that f takes values in $[0, \infty[$. But there exists n so large that $f^{-1}([n, \infty[)$ has small measure, so we can assume that f takes values in $[0, n[$. After scaling, we can assume that f takes values in $[0, 1[$. Now write, corresponding to the binary expansion of $f(x)$,

$$f = \sum_{n=1}^{\infty} 2^{-n} \mathbb{1}_{A_n}$$

where A_n are Borel sets. Exercise — show that $A_n \in \mathcal{B}_X$ using induction on n.

Now approximate each A_n from inside and out. $K_n \subseteq A_n \subseteq U_n$, where K_n is compact, U_n is open and $\mu(U_n \setminus K_n) < \epsilon 2^{-n}$. Find g_n continuous, equal to 1 on K_n, 0 off U_n and $0 \leq g_n \leq 1$ globally. Then set

$$g = \sum_{n=1}^{\infty} 2^{-n} g_n$$

a continuous function, because it is a uniform limit of continuous functions (use the M-test). Finally, f and g disagree only on $\bigcup_{n=1}^{\infty}(U_n \setminus K_n)$ which has measure $< \epsilon$. ∎

LEMMA 63 *Let $f \in L^1(X, \mathcal{F}, \mu)$ and $f \geq 0$. Let $\epsilon > 0$. Then there exists $\delta > 0$ such that $\int \mathbb{1}_A f d\mu < \epsilon$ whenever $A \in \mathcal{F}$ and $\mu(A) < \delta$.*

Proof. Suppose not. The there exists $\epsilon > 0$ such that the desired conclusion fails for every $\delta > 0$ and therefore for $\delta = 2^{-n}$. So, there is a set $A_n \in \mathcal{F}$ such that $\int \mathbb{1}_{A_n} f d\mu \geq \epsilon$ and $\mu(A_n) < 2^{-n}$. So, now set $B_n = \bigcup_{m=n}^{\infty} A_m$. Then $\mu(B_n) < 2^{1-n}$, the B_n are decreasing with n and $\int \mathbb{1}_{B_n} f d\mu \geq \epsilon$ since $B_n \supseteq A_n$. But now let $B = \bigcap_{n=1}^{\infty} B_n$. Then $\mu(B) = 0$, $\mathbb{1}_{B_n} f \downarrow \mathbb{1}_B f$ and the function $\mathbb{1}_{B_1} f$ is integrable. So, by the Dominated Convergence Theorem, we have $0 = \int \mathbb{1}_B f d\mu \geq \epsilon$ a contradiction. ∎

We can now tackle the duality situation in which we are interested. First we need a definition.

DEFINITION *Let X be a LCSC metric space and $g : X \longrightarrow \mathbb{C}$ be a continuous function. Then g tends to zero at infinity if for all $\epsilon > 0$ there exists a compact subset K of X such that $|g(x)| < \epsilon$ for all $x \in X \setminus K$.*

With a little thought, it is easy to see that K can always be chosen to be one of the K_n. (First step in the proof is to write $K \subseteq \displaystyle\bigcup_{n=1}^{\infty} \Omega_n$.) Another key observation is that if g is a continuous function tending to zero at infinity, then necessarily g is bounded. (Take $\epsilon = 1$ and "collect" K. Then g is bounded on K since g is continuous and K is compact and g is bounded by 1 off K.

The space of all continuous functions that tend to zero at infinity is denoted $C_0(X)$. We use the uniform norm on this space. It is easy to show that a uniform limit of functions in $C_0(X)$ is also in $C_0(X)$. It follows that $C_0(X)$ is a complete normed space with the uniform norm. Note that if X is in fact compact, then the "tends to zero" condition is void and so $C_0(X) = C(X)$, the space of all continuous functions on X.

We next consider the space of continuous linear forms on $C_0(X)$. These are continuous linear mappings from $C_0(X)$ to the base field (field of scalars). If we are considering real-valued functions the field of scalars will be $\mathbb{R}$. In the case of complex valued functions, it will be $\mathbb{C}$. The norm of such a form u is the operator norm, given by

$$\|u\|_{C_0(X)'} = \sup_{\substack{g \in C_0(X) \\ \|g\|_\infty \le 1}} |u(g)|.$$

Before proceeding, we mention the following theorem which is outside the scope of this course.

THEOREM 64 *Let X be a LCSC space and u a positive continuous linear form on $C_0(X)$. Positive in this context means*

$$g(x) \ge 0 \text{ for all } x \in X \Longrightarrow u(g) \ge 0.$$

Then there exists a regular Borel measure μ on X such that $u(g) = \displaystyle\int g(x) d\mu(x)$ for all $g \in C_0(X)$.

We will need the following result later.

PROPOSITION 65 *Let X be a LCSC space, μ a regular Borel measure on X. Let $f \in L^1(X, \mathcal{B}_X, \mu)$. Then we may define a continuous linear form u_f on $C_0(X)$ by*

$$u_f(g) = \int f(x) g(x) d\mu(x) \text{ for } g \in C_0(X).$$

Furthermore, $\|u_f\|_{C_0(X)'} = \|f\|_1$. In particular, if $u_f = 0$, then $f = 0$ μ-a.e.

Proof. It is clear that u_f is a continuous linear form on $C_0(X)$ and also that $\|u_f\|_{C_0(X)'} \le \|f\|_1$. The real content of the proposition is that $\|f\|_1 \le \|u_f\|_{C_0(X)'}$. By Lemma 63, given $\epsilon > 0$, there exists $\delta > 0$ such that B Borel, $\mu(B) < \delta$ implies $\displaystyle\int_B |f(x)| d\mu(x) < \epsilon$. Now, according to Lusin's Theorem, there exists

$g \in C_c(X) \subseteq C_0(X)$ such that $g(x) = \overline{\text{sgn}(f(x))}$ except for $x \in B$, where B is some Borel set with $\mu(B) < \delta$. Now let

$$\varphi(z) = \begin{cases} z & \text{if } |z| \leq 1, \\ \text{sgn}(z) & \text{if } |z| > 1. \end{cases}$$

Note that $\varphi : \mathbb{C} \longrightarrow \mathbb{C}$ is continuous. We put $h = \varphi \circ g$. Then $h \in C_c(X) \subseteq C_0(X)$ such that $h(x) = \overline{\text{sgn}(f(x))}$ except for $x \in B$ and $|h(x)| \leq 1$ for all $x \in X$. Therefore

$$\left| \int h(x) f(x) d\mu(x) \right| = |u_f(h)| \leq \|u_f\|_{C_0(X)'} \|h\|_\infty = \|u_f\|_{C_0(X)'}.$$

On the other hand

$$\left| \int |f| d\mu - \int h f d\mu \right| = \left| \int \Big(|f| - hf\Big) d\mu \right| = \left| \int_B \Big(|f| - hf\Big) d\mu \right|$$

since $|f| = hf$ off B,

$$\leq 2 \int_B |f| d\mu < 2\epsilon$$

It follows that

$$\int |f| d\mu \leq \|u_f\|_{C_0(X)'} + 2\epsilon$$

and since ϵ is an arbitrary positive number, the result follows.

4

Products of Measure Spaces

In this chapter we look at products of two measure spaces. Everything that we do here generalizes to finite products of measure spaces. First of all, we should look at measurable spaces. If $(X, \mathcal{S})$ and $(Y, \mathcal{T})$ are two measurable spaces, we say that a ***measurable rectangle*** is a subset $S \times T$ of $X \times Y$ where $S \in \mathcal{S}$ and $T \in \mathcal{T}$. The σ-field of $X \times Y$ generated by the measurable rectangles will be denoted $\mathcal{S} \otimes \mathcal{T}$. Many authors use the notation $\mathcal{S} \times \mathcal{T}$ for this, but strictly speaking this is not correct, $\mathcal{S} \times \mathcal{T}$ should really denote the measurable rectangles. It is fairly clear that the measurable rectangles form a π-system, so one can assert from Dynkin's π-λ Theorem that $\mathcal{S} \otimes \mathcal{T}$ is the smallest λ-system containing $\mathcal{S} \times \mathcal{T}$.

EXAMPLE If $X = Y = \mathbb{R}$ and $\mathcal{S} = \mathcal{T} = \mathcal{B}_{\mathbb{R}}$, then $\mathcal{S} \otimes \mathcal{T} = \mathcal{B}_{\mathbb{R}^2}$. To see that $\mathcal{B}_{\mathbb{R}^2} \subseteq \mathcal{B}_{\mathbb{R}} \otimes \mathcal{B}_{\mathbb{R}}$, recall that every open subset of $\mathbb{R}^2$ is a countable union of open rectangles $J \times K$ where J, K are open intervals in $\mathbb{R}$. This shows that every open subset of $\mathbb{R}^2$ lies in the σ-field $\mathcal{B}_{\mathbb{R}} \otimes \mathcal{B}_{\mathbb{R}}$. The inclusion now follows from the definition of $\mathcal{B}_{\mathbb{R}^2}$. The other direction is easier, but more involved. One starts from

$$A, B \text{ open} \implies A \times B \text{ open} \implies A \times B \in \mathcal{B}_{\mathbb{R}^2}.$$

Now let A be a fixed open set and show that

$$\{B; B \subseteq \mathbb{R}, A \times B \in \mathcal{B}_{\mathbb{R}^2}\} \text{ is a } \sigma\text{-field on } \mathbb{R} \text{ containing the open sets.}$$

It follows that

$$A \text{ open }, B \text{ borel} \implies A \times B \in \mathcal{B}_{\mathbb{R}^2}.$$

Then, fix B borel and show that

$$\{A; A \subseteq \mathbb{R}, A \times B \in \mathcal{B}_{\mathbb{R}^2}\} \text{ is a } \sigma\text{-field on } \mathbb{R} \text{ containing the open sets.}$$

We may deduce that

$$A, B \text{ borel } \implies A \times B \in \mathcal{B}_{\mathbb{R}^2}.$$

Finally, since $\mathcal{B}_{\mathbb{R}^2}$ is a σ-field , $\mathcal{B}_{\mathbb{R}} \otimes \mathcal{B}_{\mathbb{R}} \subseteq \mathcal{B}_{\mathbb{R}^2}$. □

4.1 The product σ-field

Much of what we do in this chapter is done with slices. If $E \subseteq X \times Y$, then we denote $E_x = \{y; (x, y) \in E\} \subseteq Y$ for every $x \in X$. We also denote $E^y = \{x; (x, y) \in E\} \subseteq X$ for every $y \in Y$.

LEMMA 66 *If $E \in \mathcal{S} \otimes \mathcal{T}$ then $E_x \in \mathcal{T}$ for all $x \in X$ and $E^y \in \mathcal{S}$ for all $y \in Y$.*

Proof. Fix $x \in X$. Now consider all subsets E of $X \times Y$ such that $E_x \in \mathcal{T}$. Call the collection of such subsets $\mathcal{A}$. Then $X \times Y \in \mathcal{A}$ and it is clear that $\mathcal{A}$ is closed under complementation since $(E^c)_x = (E_x)^c$. Also we have

$$\left(\bigcup_{j=1}^{\infty} E_j\right)_x = \bigcup_{j=1}^{\infty} (E_j)_x$$

and this shows that $\mathcal{A}$ is closed under countable unions. Hence $\mathcal{A}$ is a σ-field . Since $\mathcal{A}$ clearly contains the measurable rectangles, it also must contain $\mathcal{S} \otimes \mathcal{T}$. This proves the first assertion. The proof of the second assertion is exactly similar. ■

It is also possible to slice functions. We use the corresponding notation. If f is a mapping defined on $X \times Y$, then $f_x(y) = f(x, y) = f^y(x)$. This definition extends the idea of slicing sets in the sense that $(\mathbb{1}_A)_x = \mathbb{1}_{A_x}$ and $(\mathbb{1}_A)^y = \mathbb{1}_{A^y}$.

We can now state the analogue of Lemma 66

LEMMA 67 *Let Z be a metric space and suppose that $f : X \times Y \longrightarrow Z$ is $\mathcal{S} \otimes \mathcal{T}$-measurable. Then for each fixed $x \in X$, f_x is $\mathcal{T}$ measurable and for each fixed $y \in Y$, f^y is $\mathcal{S}$ measurable.*

Proof. The proof is pretty straightforward. Here Z can be either a metric space or a measurable space. In the former case, we work with the Borel σ-field on Z. Let V be a measurable set in Z, then it is easy to see that

$$(f^{-1}(V))_x = f_x^{-1}(V).$$

Since f is measurable, $f^{-1}(V)$ is in $\mathcal{S} \otimes \mathcal{T}$ and the slice $(f^{-1}(V))_x$ is in $\mathcal{T}$ by Lemma 66. So, $f_x^{-1}(V) \in \mathcal{T}$. This just says that f_x is measurable. Similarly for f^y. ■

This is pretty much as far as we can get with measurable spaces. From this point on, we assume that we have measure spaces $(X, \mathcal{S}, \mu)$ and $(Y, \mathcal{T}, \nu)$. In addition, we assume that these spaces are σ-finite.

LEMMA 68 *Let $A \in \mathcal{S} \otimes \mathcal{T}$, then $x \longrightarrow \nu(A_x)$ is $\mathcal{S}$-measurable and $y \longrightarrow \mu(A^y)$ is $\mathcal{T}$-measurable (as functions taking values in $[0, \infty]$).*

Proof. We show the first assertion under the additional assumption that $\nu(Y)$ is finite. This assumption will be removed later. Let us define $\mathcal{M}$ to be the collection of subsets A of $X \times Y$ such that $x \longrightarrow \nu(A_x)$ is $\mathcal{S}$-measurable. Obviously a measurable rectangle is necessarily in $\mathcal{M}$.

Since $\mathcal{S} \times \mathcal{T}$ is a π-system, by the Dynkin π-λ Theorem it will suffice to show that $\mathcal{M}$ is a λ-system. First, $X \times Y \in \mathcal{M}$ is clear since already $X \times Y \in \mathcal{S} \times \mathcal{T}$. For complementation, we have $(A^c)_x = (A_x)^c$. It follows that $\mathcal{M}$ is closed under complementation since

$$\nu(A_x) + \nu((A^c)_x) = \nu(Y).$$

and $\nu(Y) < \infty$. Finally, if $A = \bigcup_{j=1}^{\infty} A_j$ is a disjoint union, then we have

$$\nu(A_x) = \nu\left(\bigcup_{j=1}^{\infty} (A_j)_x\right) = \sum_{j=1}^{\infty} \nu((A_j)_x)$$

and a sum of a series of measurable functions is measurable. This settles the issue for the case of finite measure spaces. In the σ-finite case, we have measurable subsets $(Y_n)_{n=1}^{\infty}$ of finite measure with union Y. We can assume that this sequence is increasing. We cut on ν on Y_n, that is, we construct new measures $\nu_n(T) =$

$\nu(Y_n \cap T)$ on the measurable space $(Y, \mathcal{T})$. From the finite measure case using the fact that ν_n is a finite measure, we find that

$$x \mapsto \nu(((X \times Y_n) \cap A)_x) = \nu(Y_n \cap A_x) = \nu_n(A_x)$$

is $\mathcal{S}$-measurable. But as n increases, the quantities $\nu(((X \times Y_n) \cap A)_x)$ increase to $x \mapsto \nu(A_x)$. ■

The next step is to show that the order of processing is irrelevant. Let us define $\varphi_A(x) = \nu(A_x)$ and $\psi_A(y) = \mu(A^y)$.

LEMMA 69 *We have*

$$\int \varphi_A d\mu = \int \psi_A d\nu. \tag{4.1}$$

Proof. The proof follows the same line as above. We first assume that we are working on finite measure spaces. We let $\mathcal{M}$ be the collection of subsets A of $X \times Y$ such that (4.1) holds. It is easy to see that (69) holds for $A \in \mathcal{S} \times \mathcal{T}$ and we will again use the Dynkin π-λ Theorem. We need to show that $\mathcal{M}$ is a π-system. Again $X \times Y \in \mathcal{M}$ is clear since already $X \times Y \in \mathcal{S} \times \mathcal{T}$. For complements, we have from the proof of Lemma 68 that

$$\varphi_A + \varphi_{A^c} = \nu(Y)\mathbb{1}_X \qquad \text{and } \psi_A + \psi_{A^c} = \mu(X)\mathbb{1}_Y$$

and it follows that (4.1) implies $\displaystyle\int \varphi_{A^c} d\mu = \int \psi_{A^c} d\nu$ since the measure spaces are finite. Now let $A_j \in \mathcal{M}$ be disjoint with $A = \bigcup_{j=1}^{\infty} A_j$ and again from the proof of Lemma 68 we get

$$\varphi_A(x) = \nu(A_x) = \nu\left(\bigcup_{j=1}^{\infty} (A_j)_x\right) = \sum_{j=1}^{\infty} \nu((A_j)_x) = \sum_{j=1}^{\infty} \varphi_{A_j}(x)$$

and a similar statement for the ψs, whence

$$\begin{aligned}
\int \varphi_A d\mu &= \int \sum_{j=1}^{\infty} \varphi_{A_j} d\mu = \sum_{j=1}^{\infty} \int \varphi_{A_j} d\mu \\
&= \sum_{j=1}^{\infty} \int \psi_{A_j} d\nu = \int \sum_{j=1}^{\infty} \psi_{A_j} d\nu = \int \psi_A d\nu.
\end{aligned}$$

This settles the finite measure case.

For the general case, we find increasing measurable subsets X_n and Y_n as before and use

$$\int \varphi_A d\mu = \sup_n \int \varphi_{A\cap(X_n\times Y_n)} d\mu = \sup_n \int \psi_{A\cap(X_n\times Y_n)} d\nu \int \psi_A d\nu$$

using two applications of the Monotone Convergence Theorem. ■

Now, at long last we can define for $A \in \mathcal{S}\otimes\mathcal{T}$

$$\mu\times\nu(A) = \int \varphi_A d\mu = \int \psi_A d\nu.$$

PROPOSITION 70 *Let $(X, \mathcal{S}, \mu)$ and $(Y, \mathcal{T}, \nu)$ be σ-finite measure spaces. Then the set function $\mu\times\nu$ is a measure on $(X\times Y, \mathcal{S}\otimes\mathcal{T})$.*

Proof. It is obvious that $\mu\times\nu(\emptyset) = 0$. It remains to check the countable additivity. Let $A_j \in \mathcal{S}\otimes\mathcal{T}$ be disjoint sets for $j = 1, 2, \ldots$ and let A be their union. Then the sets $(A_j)_x$ are also disjoint for each x and their union is A_x. So it follows that

$$\varphi_A(x) = \nu(A_x) = \nu\left(\bigcup_{j=1}^{\infty}(A_j)_x\right) = \sum_{j=1}^{\infty}\nu((A_j)_x) = \sum_{j=1}^{\infty}\varphi_{A_j}(x).$$

Integrating with respect to μ and using the Monotone Convergence Theorem we get

$$\mu\times\nu(A) = \int \varphi_A d\mu = \int \left\{\sum_{j=1}^{\infty}\varphi_{A_j}\right\} d\mu = \sum_{j=1}^{\infty}\int \varphi_{A_j} d\mu = \sum_{j=1}^{\infty}\mu\times\nu(A_j).$$

Thus $\mu\times\nu$ is a measure. ■

4.2 The Monotone Class Approach to Product Spaces*

Everything that we have done in the previous section has been with the Dynkin π-λ Theorem. It seems that this is the easiest way to proceed. Most mathematical textbooks however approach this material using monotone classes and omit to mention the following technical lemma which is needed to make the Monotone Class Theorem work.

LEMMA 71 *Let $\mathcal{F}$ be the field generated by $\mathcal{S} \times \mathcal{T}$. Then every $M \in \mathcal{F}$ can be written as a finite disjoint union of measurable rectangles.*

Proof. By Lemma 5, and bearing in mind that the whole space $X \times Y$ is itself a measurable rectangle, there exists $N \in \mathbb{N}$ such that $M = M_N$ and the sequence $(M_n)_{n=1}^N$ is defined inductively by one of the following

- $M_n = S_n \times T_n$ with $S_n \in \mathcal{S}$ and $T_n \in \mathcal{T}$.
- $M_n = M_{p_n} \setminus M_{q_n}$ with $1 \leq p_n, q_n < n$.
- $M_n = M_{p_n} \cup M_{q_n}$ with $1 \leq p_n, q_n < n$.

Obviously, the first option is used at most N times. In the instances where the first option is not used, define $S_n = X$, $T_n = Y$. Now consider for each subset $Z \subseteq \{1, 2, 3, \ldots, N\}$, the sets

$$S_Z = \bigcap_{n=1}^N S_{n,\epsilon(n)}, \qquad T_Z = \bigcap_{n=1}^N T_{n,\epsilon(n)}$$

where

$$\epsilon(n) = \begin{cases} 1 & \text{if } n \in Z, \\ 0 & \text{if } n \notin Z. \end{cases}$$

and where $S_{n,1} = S_n$, $S_{n,0} = X \setminus S_n$, $T_{n,1} = T_n$ and $T_{n,0} = Y \setminus T_n$. A simple induction proof now shows that each M_n is a union of some subcollection of the 4^N *disjoint* sets $S_Z \times T_W$ as Z and W run over $\{1, 2, 3, \ldots, N\}$. ∎

It follows from the Monotone Class Theorem that the smallest monotone class containing $\mathcal{F}$ is $\mathcal{S} \otimes \mathcal{T}$ and the results in this section can be proved using this fact. We leave the details to the reader.

4.3 Fubini's Theorem

In this section we look at the product integral and compare it with iterated integrals. Let's start by setting up the iterated integrals. We assume throughout this section that $(X, \mathcal{S}, \mu)$ and $(Y, \mathcal{T}, \nu)$ are σ-finite measure spaces.

LEMMA 72 *Let $f : X \times Y \longrightarrow [0, \infty]$ be $\mathcal{S} \otimes \mathcal{T}$ measurable. Then define*

$$\varphi(x) = \int f(x, y) d\nu(y).$$

Then φ is $\mathcal{S}$ measurable on X. Obviously there is an analogous statement if we integrate first over x.

Proof. Let $A \in \mathcal{S} \otimes \mathcal{T}$ and $f = \mathbb{1}_A$. Then the result follows from Lemma 68. It therefore also follows for $\mathcal{S} \otimes \mathcal{T}$ measurable nonnegative simple functions. But we can find a sequence of such functions (f_n) increasing pointwise to f. Let $\varphi_n(x) = \int f_n(x, y) d\nu(y)$. Then φ_n is $\mathcal{S}$ measurable and by the Monotone Convergence Theorem, $\varphi \uparrow \varphi$ pointwise on X. It follows that φ is $\mathcal{S}$ measurable. ∎

Let us define also $\psi(y) = \displaystyle\int f(x, y) d\mu(x)$. Then we have

THEOREM 73 (TONELLI'S THEOREM) *Let $f : X \times Y \longrightarrow [0, \infty]$ be $\mathcal{S} \otimes \mathcal{T}$ measurable. Then*

$$\int \varphi(x) d\mu(x) = \int f(x, y) d(\mu \times \nu)(x, y) = \int \psi(y) d\nu(y).$$

Proof. Same proof as above. If $A \in \mathcal{S} \otimes \mathcal{T}$ and $f = \mathbb{1}_A$, then the result follows from Lemma 69. It therefore also follows for $\mathcal{S} \otimes \mathcal{T}$ measurable nonnegative simple functions. Now take a sequence of such functions (f_n) increasing pointwise to f and pass to the limit using the Monotone Convergence Theorem 5 times! ∎

Notice that the σ-finiteness is used to define the product measure $\mu \times \nu$. But we can ask whether

$$\int \varphi(x) d\mu(x) = \int \psi(y) d\nu(y).$$

might hold in general, without the σ-finiteness assumption. The answer is that it does not. Let $X = Y = [0, 1]$. Let $\mathcal{S}$ be the Lebesgue field of $[0, 1]$. Let $\mathcal{T}$ be the collection of all subsets of $[0, 1]$. Let μ be Lebesgue measure on $[0, 1]$ and let ν be the counting measure. Finally, let Δ be the diagonal set. It's easy to prove that $\Delta \in \mathcal{S} \otimes \mathcal{T}$. Now

$$\int \left\{ \int \mathbb{1}_\Delta(x, y) d\nu(y) \right\} d\mu(x) = 1$$

because each inner integral evaluates to 1. On the other hand

$$\int \left\{ \int \mathbb{1}_\Delta(x,y) d\mu(x) \right\} d\nu(y) = 0$$

because each inner integral evaluates to 0. This does not contradict Tonelli's Theorem because the counting measure is not σ-finite on an uncountable space such as $[0,1]$.

A special case of Tonelli's Theorem is

$$\sum_{j=1}^\infty \sum_{k=1}^\infty a_{j,k} = \sum_{k=1}^\infty \sum_{j=1}^\infty a_{j,k}$$

for $a_{j,k} \geq 0$.

So much for the unsigned case. We know already from iterated infinite sums that for signed series some additional hypotheses are going to be needed if the order of integration will not matter. See §2.8 in the notes for MATH 255.

Let $f : X \times Y \longrightarrow \mathbb{R}$ be $\mathcal{S} \otimes \mathcal{T}$ measurable. Then

$$\iint |f(x,y)| d\nu(y) d\mu(x) = \int |f| d(\mu \times \nu) = \iint |f(x,y)| d\mu(x) d\nu(y). \tag{4.2}$$

THEOREM 74 (FUBINI'S THEOREM) *If any one of the three quantities in* (4.2) *is finite, then*

$$\iint f(x,y) d\nu(y) d\mu(x) = \int f d(\mu \times \nu) = \iint f(x,y) d\mu(x) d\nu(y). \tag{4.3}$$

The precise meaning of the iterated integrals will become clear in the proof.

Proof. By symmetry, it suffices to establish the left-hand equality in (4.3). We write $f = f_+ - f_-$ in the usual way and define $\varphi_\pm(x) = \int f_\pm(x,y) d\nu(y)$. We have

$$\int \varphi_\pm(x) d\mu(x) = \iint f_\pm(x,y) d\nu(y) d\mu(x) \leq \iint |f(x,y)| d\nu(y) d\mu(x) < \infty.$$

This means that for μ-almost all x, both $\varphi_+(x)$ and $\varphi_-(x)$ are finite. So, for μ-almost all x, the integral $\int f(x,y) d\nu(y)$ exists and equals $\varphi_+(x) - \varphi_-(x)$. So, let us define

$$\varphi(x) = \begin{cases} \int f(x,y) d\nu(y) & \text{if this integral exists} \\ 0 & \text{otherwise} \end{cases}$$

So, $\varphi(x) = \varphi_+(x) - \varphi_-(x)$ for μ-almost all x and indeed $|\varphi(x)| \leq \varphi_+(x) + \varphi_-(x)$ for all x. So $\int |\varphi(x)| d\mu(x) \leq 2 \iint |f(x,y)| d\nu(y) d\mu(x) < \infty$. This means that $\int \varphi(x) d\mu(x)$ is defined and

$$\begin{aligned}
\int \varphi(x) d\mu(x) &= \int \varphi_+(x) d\mu(x) - \int \varphi_-(x) d\mu(x) \\
&= \iint f_+(x,y) d\nu(y) d\mu(x) - \iint f_-(x,y) d\nu(y) d\mu(x) \\
&= \int f_+(x,y) d(\mu \times \nu)(x,y) - \int f_-(x,y) d(\mu \times \nu)(x,y) \\
&= \int f(x,y) d(\mu \times \nu)(x,y)
\end{aligned}$$

as required. ■

4.4 Estimates on Homogenous kernels*

As a demonstration of duality at work, we will prove the following result for the integral operator

$$Tf(x) = \int_0^\infty K(x,y) f(y) dy$$

where x runs over $]0, \infty[$. We assume that the kernel function K is nonnegative and satisfies the homogeneity condition

$$K(tx, ty) = t^{-1} K(x,y).$$

We will also need to know that K is Lebesgue measurable on the positive quadrant and that it is sufficiently regular for us invoke some change of variable arguments from the Riemann Theory. We further define

$$\begin{aligned}
C_p = \int_0^\infty t^{-\frac{1}{p'}} K(t,1) dt &= \int_0^\infty t^{-\frac{1}{p'}} K(1, t^{-1}) t^{-1} dt \\
&= \int_0^\infty s^{\frac{1}{p'}} K(1,s) s^{-1} ds \\
&= \int_0^\infty s^{-\frac{1}{p}} K(1,s) ds
\end{aligned}$$

and we will assume that C_p is finite.

Now comes a remarkable idea. For f and g nonnegative functions we have using Hölder's Inequality and Tonelli's Theorem

$$\begin{aligned}
&\int_0^\infty\int_0^\infty K(x,y)g(x)f(y)dxdy \\
&= \int_0^\infty\int_0^\infty \left(\frac{x}{y}\right)^{-\frac{1}{pp'}} K(x,y)^{\frac{1}{p}} f(y) \left(\frac{y}{x}\right)^{-\frac{1}{pp'}} K(x,y)^{\frac{1}{p'}} g(x)dxdy \\
&\leq \Big\{ \int_0^\infty\int_0^\infty \left(\frac{x}{y}\right)^{-\frac{1}{p'}} K(x,y)f(y)^p dxdy\Big\}^{\frac{1}{p}} \Big\{ \int_0^\infty\int_0^\infty \left(\frac{y}{x}\right)^{-\frac{1}{p}} K(x,y)g(x)^{p'} dxdy\Big\}^{\frac{1}{p'}}
\end{aligned}$$

We now have

$$\begin{aligned}
&\int_0^\infty\int_0^\infty \left(\frac{x}{y}\right)^{-\frac{1}{p'}} K(x,y)f(y)^p dxdy \\
&= \int_{y=0}^\infty \left\{ \int_{x=0}^\infty K\left(\frac{x}{y},1\right)\left(\frac{x}{y}\right)^{-\frac{1}{p'}} y^{-1}dx \right\} f(y)^p dy \\
&= C_p\|f\|_p^p
\end{aligned}$$

and similarly

$$\int_0^\infty\int_0^\infty \left(\frac{y}{x}\right)^{-\frac{1}{p}} K(x,y)g(x)^{p'} dxdy = C_p\|g\|_{p'}^{p'}$$

resulting in

$$\int_0^\infty\int_0^\infty K(x,y)g(x)f(y)dxdy \leq C_p\|f\|_p\|g\|_{p'}$$

It follows from this by duality that $\|Tf\|_p \leq C_p\|f\|_p$ for f nonnegative and the same inequality for signed or complex f then follows.

Some specific cases of interest are

- $Tf(x) = x^{-1}\int_0^x f(y)dy$, $C_p = \dfrac{p}{p-1}$.
- $Tf(x) = \int_x^\infty y^{-1}f(y)dy$, $C_p = p$.
- $Tf(x) = \int_0^\infty (x+y)^{-1}f(y)dy$, $C_p = \pi\text{cosec}\left(\dfrac{\pi}{p}\right)$.

4.5 Uniqueness of Translation Invariant Measures

In this section, we show the uniqueness of translation invariant measures.

PROPOSITION 75 *Let μ and ν be nonzero translation invariant Borel measures on $\mathbb{R}^d$ which assign finite values to compact sets. Then μ and ν are scalar multiples of one another.*

Proof. Let C be a Borel set with $0 < \mu(C)$. Write $\mathbb{R}^d$ as a union of countably many cubes Q of side one. Then choosing $B = C \cap Q$ for suitable Q, we can assume that $0 < \mu(B) < \infty$ and $\nu(B) < \infty$. Let $g = \mathbb{1}_B$ and let $t = \int g(-x)d\nu(x)$. Then we have for $f = \mathbb{1}_A$ with A Borel,

$$\mu(B) \int f(x)d\nu(x) = \iint g(y)f(u)d\nu(u)d\mu(y)$$

$$= \iint f(x+y)g(y)d\nu(x)d\mu(y)$$

by changing variables $u = x + y$ in the inner integral

$$= \iint f(x+y)g(y)d\mu(y)d\nu(x)$$

by Tonelli's Theorem

$$= \iint f(u)g(u-x)d\mu(u)d\nu(x)$$

by changing variables $u = x + y$ in the inner integral

$$= \iint f(u)g(u-x)d\nu(x)d\mu(u)$$

again by Tonelli's Theorem

$$= \iint f(u)g(-y)d\nu(y)d\mu(u)$$

by changing variables $y = x - u$ in the inner integral

$$= t \int f(u)d\mu(u)$$

Thus $\mu(B)\nu(A) = t\mu(A)$. Putting $A = B$ gives $\nu(B) = t$, so that $t < \infty$. If $t = 0$, then ν vanishes identically which is not allowed. Hence $0 < t < \infty$ and each measure is a positive multiple of the other. ■

4.6 Infinite products of probability spaces**

Results for the product of two measure spaces generalize without problem to finite products. Infinite products of measure spaces are more problematic in general, but are important in probability theory because they are needed to set up sequences of independent trials. We develop the basics of this theory here. Let $(\Omega_k, \mathcal{M}_k, \mu_k)$ be a probability space for $k \in \mathbb{N}$. We let $\Omega = \prod_{k=1}^{\infty} \Omega_k$ the full infinite cartesian product of the Ω_k. An element of Ω is then a sequence $(\omega_k)_{k=1}^{\infty}$ with $\omega_k \in \Omega_k$ for every k. The next concept we need is that of a ***cylinder set*** at level q which is a set $C = Q \times \prod_{k=q+1}^{\infty} \Omega_k$ where $Q \in \mathcal{M}_1 \otimes \mathcal{M}_2 \otimes \cdots \otimes \mathcal{M}_q$. The family of cylinder sets at level q is a σ-field $\mathcal{C}_q$ in some sense isomorphic to $\mathcal{M}_1 \otimes \mathcal{M}_2 \otimes \cdots \otimes \mathcal{M}_q$ (although the two gadgets live on different sets). The family $\mathcal{C} = \bigcup_{q=1}^{\infty} \mathcal{C}_q$ is the collection of all cylinder sets and is a field on Ω, but usually not a σ-field. By patching the probabilities together, we can define a finitely additive set function μ on $\mathcal{C}$

$$\mu\left(Q \times \prod_{k=q+1}^{\infty} \Omega_k\right) = \mu_1 \times \cdots \times \mu_q(Q).$$

First of all, we need to check that this is well-defined, i.e. we would get the same answer if $Q \times \prod_{k=q+1}^{\infty} \Omega_k$ were considered as a cylinder set at a level higher than q. This works because we are dealing with probability spaces. To check the finite additivity we need only consider finitely many cylinder sets and they can all be defined as cylinder sets at a fixed level, say $\tilde{q}$. The corresponding values of μ are effectively given by the measure $\mu_1 \times \cdots \times \mu_{\tilde{q}}$.

THEOREM 76 *The finitely additive set function μ is in fact a premeasure on $\mathcal{C}$ and therefore extends to a measure on the σ-field generated by $\mathcal{C}$.*

Sketch proof. It is not too difficult to see that given that μ is finitely additive on $\mathcal{C}$, the following condition guarantees that μ is in fact countably additive on $\mathcal{C}$. Let $C_k \in \mathcal{C}$ be decreasing and such that $\bigcap_{k=1}^{\infty} C_k = \emptyset$, then $\lim_{k\to\infty} \mu(C_k) = 0$. Obviously, if the levels of the sets C_k are bounded, then the result is clear, so we can assume that the levels are unbounded. Therefore, after making some adjustments, we see that it suffices to show the following:

> Let $C_k \in \mathcal{C}_k$ be decreasing for $k \in \mathbb{N}$ with $\bigcap_{k=1}^{\infty} C_k = \emptyset$, then $\lim_{k\to\infty} \mu(C_k) = 0$.

We prove this assertion by contradiction. So let $t > 0$ be such that $\mu(C_k) \geq t$ for all k. We look at the distributions on the first coordinate

$$f_{1,k}(\omega_1) = \int \mathbb{1}_{C_k}(\omega_1, \ldots, \omega_k) d\mu_2(\omega_2) d\mu_3(\omega_3) \cdots d\mu_k(\omega_k)$$

We have $\int f_{1,k}(\omega_1) d\mu_1(\omega_1) = \mu(C_k) \geq t$. Since the $f_{1,k}$ are clearly decreasing (since the C_k are), we get by dominated convergence

$$\int \inf_{k=1}^{\infty} f_{1,k}(\omega_1) d\mu_1(\omega_1) \geq t$$

It now follows that there exists $\tilde{\omega}_1$ such that $\inf_{k=1}^{\infty} f_{1,k}(\tilde{\omega}_1) \geq t$. In particular, $\mathbb{1}_{C_1}(\tilde{\omega}_1) = f_{1,1}(\tilde{\omega}_1) \geq t > 0$ and since C_1 is known to be a cylinder set at level 1, this means that

$$(\tilde{\omega}_1, \omega_2, \omega_3, \ldots) \in C_1$$

whatever the values of $\omega_2 \in \Omega_2, \omega_3 \in \Omega_3, \ldots$ The next step is to slice all these sets on $\tilde{\omega}_1$. Thus we arrive at

$$C_k(\tilde{\omega}_1) = \{(\omega_2, \omega_3, \ldots); (\tilde{\omega}_1, \omega_2, \omega_3, \ldots) \in C_k\} \subseteq \prod_{k=2}^{\infty} \Omega_k$$

for $k = 2, 3, \ldots$. We proceed to compute the distributions on the second coordinate

$$f_{2,k}(\omega_2) = \int \mathbb{1}_{C_k(\tilde{\omega}_1)}(\omega_2, \omega_3, \ldots, \omega_k) d\mu_3(\omega_3) d\mu_4(\omega_4) \cdots d\mu_k(\omega_k)$$

and we observe that

$$\begin{aligned} \int_{\Omega_2} f_{2,k}(\omega_2) d\mu_2(\omega_2) &= \int \mathbb{1}_{C_k(\tilde{\omega}_1)}(\omega_2, \omega_3, \ldots, \omega_k) d\mu_2(\omega_2) d\mu_3(\omega_3) \cdots d\mu_k(\omega_k) \\ &= f_{1,k}(\tilde{\omega}_1) \geq t \end{aligned}$$

Since the $f_{2,k}$ are decreasing, we get again by dominated convergence that

$$\int \inf_{k=2}^{\infty} f_{2,k}(\omega_2) d\mu_2(\omega_2) \geq t$$

It again follows that there exists $\tilde{\omega}_2$ such that $\inf_{k=2}^{\infty} f_{2,k}(\tilde{\omega}_2) \geq t$. In particular, $\mathbb{1}_{C_2(\tilde{\omega}_1)}(\tilde{\omega}_2) = f_{2,2}(\tilde{\omega}_2) \geq t > 0$ and we see that the point $\tilde{\omega}_2 \in \Omega_2$ has the property that

$$(\tilde{\omega}_2, \omega_3, \omega_4 \ldots) \in C_2(\tilde{\omega}_1), \text{or equivalently } (\tilde{\omega}_1, \tilde{\omega}_2, \omega_3, \omega_4 \ldots) \in C_2$$

whatever the values of $\omega_3 \in \Omega_3, \omega_4 \in \Omega_4, \ldots$ We continue in this way ad infinitum (using the axiom of choice to make the sequence of choices that are necessary) and we find ultimately that

$$(\tilde{\omega}_1, \tilde{\omega}_2, \tilde{\omega}_3, \tilde{\omega}_4 \ldots) \in \bigcap_{k=1}^{\infty} C_k.$$

But this is a contradiction, because we are assuming that $\bigcap_{k=1}^{\infty} C_k = \emptyset$! Thus μ is a premeasure on $\mathcal{C}$, and an application of Carathéodory's Extension Theorem finishes the proof. ■

5

Hilbert Spaces

A Hilbert space is a complete inner product space. The key example of a Hilbert space is $L^2(X, \mathcal{M}, \mu)$ where $(X, \mathcal{M}, \mu)$ is a measure space. It's worth observing that you need a reasonable measure space for $L^2(X, \mathcal{M}, \mu)$ to be a useful concept. If for example μ takes only the value 0 and ∞, then you will have a great scarcity of L^2 functions.

Hilbert spaces are important because they have almost magical properties and are usually very easy to handle. This was observed in particular by John von Neumann who devised a scheme for developing some of the harder measure theory theorems using them. They are also extremely important in Physics, where they form the theoretical basis for Quantum Mechanics. John von Neumann also had a hand in this development.

PROPOSITION 77 *Let H be a Hilbert space (real or complex) and let $C \subseteq H$ be a closed convex subset. Let $x \in H$. Then there is a unique nearest point y of C to x.*

Proof. First of all, if $x \in C$, then we clearly have that $y = x$ is the unique nearest point of C to x. So, we can assume that $x \notin C$. Then, since C is closed, we have $\operatorname{dist}_C(x) > 0$. We can find a sequence (y_n) with $y_n \in C$ and $d(y_n, x) \downarrow \operatorname{dist}_C(x)$. Let $z_n = y_n - x$. We use the parallelogram identity

$$\|z_p - z_q\|^2 = 2\|z_p\|^2 + 2\|z_q\|^2 - 4\|\tfrac{1}{2}(z_p + z_q)\|^2. \tag{5.1}$$

This identity is valid in inner product spaces, but usually not for other norms. Now, given $\epsilon > 0$ there exists N such that $n \geq N$ implies

$$\|z_n\|^2 = \|y_n - x\|^2 \leq (\operatorname{dist}_C(y_n))^2 + \epsilon.$$

Also, since C is convex $\frac{1}{2}(y_p+y_q) \in C$ and so $\|\frac{1}{2}(z_p+z_q)\| = \|\frac{1}{2}(y_p+y_q)-x\| \geq \mathrm{dist}_C(x)$. Putting these facts in (5.1) we get

$$\begin{aligned}
\|y_p - y_q\|^2 &= \|z_p - z_q\|^2 \\
&= 2\|z_p\|^2 + 2\|z_q\|^2 - 4\|\tfrac{1}{2}(z_p + z_q)\|^2, \\
&\leq 2((\mathrm{dist}_C(y_n))^2 + \epsilon) + 2((\mathrm{dist}_C(y_n))^2 + \epsilon) - 4(\mathrm{dist}_C(y_n))^2, \\
&= 4\epsilon.
\end{aligned}$$

for $p, q \geq N$. So, (y_n) is a Cauchy sequence. Since H is complete and C is closed, the sequence must converge to some element $y \in C$. By continuity of the distance function we get $d(y, x) = \mathrm{dist}_C(x)$. That settles the existence. Now for the uniqueness. This is also a consequence of the parallelogram inequality. Let y_1 and y_2 both be nearest points of C to X. Then let $z_j = y_j - x$ for $j = 1, 2$. Then

$$0 \leq \|z_1 - z_2\|^2 = 2\|z_1\|^2 + 2\|z_2\|^2 - 4\|\tfrac{1}{2}(z_1 + z_2)\|^2 \leq 0.$$

by much the same reasoning as above. It follows that $y_1 = y_2$. ■

In fact, this defines a mapping $P_C : H \longrightarrow C$ called the metric projection onto C. We do not need the Lemma below, but it is an interesting fact.

LEMMA 78 *Let H be a Hilbert space (real or complex) and let $C \subseteq H$ be a closed convex subset. Then P_C satisfies $\|P_C(x_1) - P_C(x_2)\| \leq \|x_1 - x_2\|$ for all $x_1, x_2 \in H$.*

Proof. This result and the previous result are actually results about real Hilbert spaces. Every complex Hilbert space is actually also a real Hilbert space. This is achieved by

- Forgetting how to scalar multiply vectors by non-real complex numbers.
- Replacing the inner product with its real part and verifying that this is now a "real" inner product.

Let us denote $y_j = P_C(x_j)$ for $j = 1, 2$. Now if $y_1 = y_2$ there is nothing to prove. Otherwise, we see from the convexity of C that $(1-t)y_1 + ty_2 \in C$ for $0 \leq t \leq 1$. Thus we must have since y_1 is the nearest point of C to x_1 that

$$\|x_1 - \Big((1-t)y_1 + ty_2\Big)\|^2 \geq \|x_1 - y_1\|^2$$

for $0 \leq t \leq 1$. Expanding the norms in terms of the inner product and considering small positive values of t gives

$$\langle (x_1 - y_1), (y_2 - y_1) \rangle \leq 0.$$

This inequality expresses the fact that the angle subtended at y_1 between x_1 and y_2 is obtuse. Similarly we have

$$\langle (x_2 - y_2), (y_1 - y_2) \rangle \leq 0.$$

Now, we get

$$\begin{aligned}
\langle (x_1 - x_2), (y_1 - y_2) \rangle &= \langle (x_1 - y_1), (y_1 - y_2) \rangle + \langle (y_1 - y_2), (y_1 - y_2) \rangle \\
&\qquad + \langle (y_2 - x_2), (y_1 - y_2) \rangle \\
&\geq \langle (y_1 - y_2), (y_1 - y_2) \rangle \\
&= \|y_1 - y_2\|^2.
\end{aligned}$$

Next we apply the Cauchy-Schwarz inequality

$$\|y_1 - y_2\|^2 \leq \langle (x_1 - x_2), (y_1 - y_2) \rangle \leq \|x_1 - x_2\| \|y_1 - y_2\|.$$

Finally, since $\|y_1 - y_2\| > 0$ we can divide out to get $\|y_1 - y_2\| \leq \|x_1 - x_2\|$ as required. ■

5.1 Orthogonal Projections

Let H be a Hilbert space either real or complex. Let $S \subseteq H$. Then we define

$$S^\perp = \{x; x \in H,\ \langle s, x \rangle = 0,\ \text{for all } s \in S\}.$$

It is clear that $S^\perp$ is an intersection of closed linear subspaces of H and therefore it is a closed linear subspace of H.

THEOREM 79 *Let M be a closed linear subspace of H. Then we have $H = M \oplus M^\perp$. Furthermore. let P and Q be the linear projection operators onto M and $M^\perp$ associated with the direct sum. Then P and Q are norm decreasing and in fact, more generally we have $\|x\|^2 = \|P(x)\|^2 + \|Q(x)\|^2$ for all $x \in H$.*

Proof. Let $x \in H$. Now, since M is a closed linear subspace, it is *a fortiori* a closed convex set. Therefore, there exists a unique nearest point y of M to x. So, for every $u \in M$ and scalar t we have

$$\|x - (y - tu)\|^2 \geq \|x - y\|^2.$$

So,

$$\Re t\langle x - y, u\rangle + |t|^2\|u\|^2 \geq 0.$$

Dividing by $|t|$ and letting t tend to zero from all possible directions yields that $\langle x - y, u\rangle = 0$. So $x - y \in M^\perp$. This shows that $H = M + M^\perp$. It remains to show that the sum is direct. So, let $x \in M \cap M^\perp$ and then $\|x\|^2 = \langle x, x\rangle = 0$. This shows that $M \cap M^\perp = \{0_H\}$. The sum is direct. The equality $\|x\|^2 = \|P(x)\|^2 + \|Q(x)\|^2$ is just Pythagoras' Theorem. ■

We denote $S^{\perp\perp} = (S^\perp)^\perp$. This set has a neat characterization.

LEMMA 80 *Let H be a Hilbert space either real or complex. Let $S \subseteq H$. Let M be the closure of the linear span of S. Then $S^{\perp\perp} = M$.*

Proof. We start by proving the result in case that S is already a closed linear subspace, i.e. $M = S$. Then clearly we have $M \subseteq M^{\perp\perp}$. To establish the opposite inclusion, let $x \in M^{\perp\perp}$ and write $x = y + z$ where $y \in M$ and $z \in M^\perp$, according to the direct sum $H = M \oplus M^\perp$. Then

$$\|z\|^2 = \langle z, z\rangle = \langle x - y, z\rangle = \langle x, z\rangle - \langle y, z\rangle = 0 - 0 = 0$$

The term $\langle x, z\rangle$ vanishes since $x \in M^{\perp\perp}$ and $z \in M^\perp$ and $\langle y, z\rangle$ vanishes since $y \in M$ and $z \in M^\perp$. Hence $z = 0$ and it follows that $x = y \in M$.

Now for the general case we assume only that S is an arbitrary subset of H. It is clear that $S \subseteq M$, whence $S^\perp \supseteq M^\perp$, whence $S^{\perp\perp} \subseteq M^{\perp\perp} = M$. We need to establish the inclusion $M \subseteq S^{\perp\perp}$. It is evident that $S \subseteq S^{\perp\perp}$, and since $S^{\perp\perp}$ is a closed linear subspace of H it must contain the smallest closed linear subspace of H containing S which is by definition M. ■

5.2 Conditional Expectation Operators

As an example of orthogonal projections, we can look at conditional expectation operators. These arise when we have two nested σ-fields on the same set. So, let

$(X, \mathcal{F}, \mu)$ be a measure space and suppose that $\mathcal{G} \subseteq \mathcal{F}$ is also a σ-field. Then, $(X, \mathcal{G}, \mu|_{\mathcal{G}})$ is an equally good measure space. It is easy to see that $L^2(X, \mathcal{G}, \mu|_{\mathcal{G}})$ is a closed linear subspace of $L^2(X, \mathcal{F}, \mu)$ because $L^2(X, \mathcal{G}, \mu|_{\mathcal{G}})$ is complete and whenever a complete space is embedded isometrically in a larger metric space, it necessarily occurs as a closed subset. It should be pointed out, that trivialities can arise even when we might not expect them. For example, let $X = \mathbb{R}^2$, $\mathcal{F}$ the Borel σ-field of $\mathbb{R}^2$ and $\mathcal{G}$ the sets which depend only on the first coordinate. Then unfortunately $L^2(X, \mathcal{G}, \mu|_{\mathcal{G}})$ consists just of the zero vector.

The situation is very significant in probability theory, where the σ-fields $\mathcal{F}$ and $\mathcal{G}$ encode which events are available to different "observers" or to the same observer at different times". For example $\mathcal{G}$ might encode outcomes based on the first 2 rolls of the dice, while $\mathcal{F}$ might encode outcomes based on the first 4 rolls.

A useful example is the case where $X = [0, 1[$, $\mathcal{F}$ is the borel σ-field of X. Then partition $[0, 1[$ into n intervals and let $\mathcal{G}$ be the σ-field generated by these intervals. We take μ the linear measure on the interval. In this case $\mathbf{E}_{\mathcal{G}}$ will turn out to be the mapping which replaces a function with its average value on each of the given intervals.

Well, the orthogonal projection operator is denoted $\mathbf{E}_{\mathcal{G}}$. We view it as a map

$$\mathbf{E}_{\mathcal{G}} : L^2(X, \mathcal{F}, \mu) \longrightarrow L^2(X, \mathcal{G}, \mu) \subseteq L^2(X, \mathcal{F}, \mu).$$

We usually understand $\mathbf{E}_{\mathcal{G}}$ in terms of its properties. These are

1. $\mathbf{E}_{\mathcal{G}}(f) \in L^2(X, \mathcal{G}, \mu)$.

2. $\int (f - \mathbf{E}_{\mathcal{G}}(f)) g d\mu = 0$ whenever $g \in L^2(X, \mathcal{G}, \mu)$.

The probabilists will write this last condition as $\mathbf{E}(f - \mathbf{E}_{\mathcal{G}}(f))g = 0$ whenever $g \in L^2(X, \mathcal{G}, \mu)$, where $\mathbf{E}$ is the scalar-valued expectation.

To get much further we will need the additional assumption that $(X, \mathcal{G}, \mu)$ is σ-finite. So we are assuming the existence of an increasing sequence of sets $G_n \in \mathcal{G}$ with $X = \bigcup_{n=1}^{\infty} G_n$ and $\mu(G_n) < \infty$. As an exercise, the reader should check that $\mathbf{E}_{\mathcal{G}}(\mathbb{1}_G f) = \mathbb{1}_G \mathbf{E}_{\mathcal{G}}(f)$ for $G \in \mathcal{G}$. We do this by taking the inner product against every function in $L^2(X, \mathcal{G}, \mu)$ and using property (ii) above.

Next we claim that if $f \in L^2(X, \mathcal{F}, \mu)$ with $|f| \leq 1$, then $|\mathbf{E}_{\mathcal{G}}(f)| \leq 1$. To see this, let $G \in \mathcal{G}$ with $\mu(G) < \infty$, $t > 1$ and $g = \mathbb{1}_G \operatorname{sgn}(\mathbf{E}_{\mathcal{G}} f) \mathbb{1}_{\{|\mathbf{E}_{\mathcal{G}} f| > t\}}$. Then we have

$$t\mu\Big(\{|\mathbf{E}_{\mathcal{G}} f| > t\} \cap G\Big) = \int t \mathbb{1}_{\{|\mathbf{E}_{\mathcal{G}} f| > t\} \cap G} d\mu$$

$$\leq \int |\mathbf{E}_{\mathcal{G}} f| \mathbb{1}_{\{|\mathbf{E}_{\mathcal{G}} f|>t\}\cap G} d\mu$$

since $t < |\mathbf{E}_{\mathcal{G}} f|$ on the range of integration,

$$= \int (\mathbf{E}_{\mathcal{G}} f)\, \overline{\operatorname{sgn}(\mathbf{E}_{\mathcal{G}} f)} \mathbb{1}_{\{|\mathbf{E}_{\mathcal{G}} f|>t\}} \mathbb{1}_G d\mu$$

$$= \int (\mathbf{E}_{\mathcal{G}} f)\, g d\mu$$

by definition of g,

$$= \int f\, g d\mu$$

by definition of $\mathbf{E}_{\mathcal{G}} f$ and since $g \in L^2(X, \mathcal{G}, \mu)$,

$$\leq \int |g| d\mu = \mu\Big(\{|\mathbf{E}_{\mathcal{G}} f| > t\} \cap G\Big).$$

But since $\mu\Big(\{|\mathbf{E}_{\mathcal{G}} f| > t\} \cap G\Big)$ is finite, and $t > 1$, the only way out is that $\mu\Big(\{|\mathbf{E}_{\mathcal{G}} f| > t\} \cap G\Big) = 0$. Since this is true for all $t > 1$ and all $G \in \mathcal{G}$ with finite measure, it follows that $|\mathbf{E}_{\mathcal{G}} f| \leq 1$ μ-a.e. using the σ-finiteness of $\mathcal{G}$.

This gives us a way of extending the definition of conditional expectation to L^∞ functions. We define for $f \in L^\infty(X, \mathcal{F}, \mu)$,

$$\mathbf{E}_{\mathcal{G}} f(x) = \mathbf{E}_{\mathcal{G}} \mathbb{1}_{G_n} f(x) \qquad \forall x \in G_n$$

The apparent dependence of this definition on n is illusory because for $x \in G_n$

$$\mathbf{E}_{\mathcal{G}} \mathbb{1}_{G_{n+1}} f(x) = (\mathbb{1}_{G_n} \cdot \mathbf{E}_{\mathcal{G}} \mathbb{1}_{G_{n+1}} f)(x) = \mathbf{E}_{\mathcal{G}} \mathbb{1}_{G_n} \mathbb{1}_{G_{n+1}} f(x) = \mathbf{E}_{\mathcal{G}} \mathbb{1}_{G_n} f(x)$$

and indeed, as an exercise, the reader can show that the definition is independent of the choice of sequence G_n. The bottom line here is that $\mathbf{E}_{\mathcal{G}}$ is a norm decreasing map

$$\mathbf{E}_{\mathcal{G}} : L^\infty(X, \mathcal{F}, \mu) \longrightarrow L^\infty(X, \mathcal{G}, \mu).$$

Now let $1 \leq p < \infty$ and let $f \in V$ where V is the space of bounded $\mathcal{F}$-measurable simple functions carried by a subset $G \in \mathcal{G}$ with $\mu(G) < \infty$. In this case we will have that $\mathbf{E}_{\mathcal{G}} f$ is a bounded $\mathcal{G}$ measurable function still carried by the subset G. We will estimate the L^p norm of $\mathbf{E}_{\mathcal{G}} f$.

$$\int |\mathbf{E}_{\mathcal{G}} f|^p d\mu = \int (\mathbf{E}_{\mathcal{G}} f) g d\mu$$

where $g = |\mathbf{E}_{\mathcal{G}} f|^{p-1}\overline{\text{sgn}(\mathbf{E}_{\mathcal{G}} f)}$,

$$= \int f g d\mu$$

since g is $\mathcal{G}$-measurable and all functions are in the appropriate L^2 space,

$$\leq \|f\|_p \|g\|_{p'},$$

by Hölder's Inequality. On the other hand (at least in case $p > 1$)

$$\|g\|_{p'}^{p'} = \int |\mathbf{E}_{\mathcal{G}} f|^{\frac{p}{p-1}(p-1)} d\mu = \|\mathbf{E}_{\mathcal{G}} f\|_p^p,$$

leading to $\|g\|_{p'} \leq \|\mathbf{E}_{\mathcal{G}} f\|_p^{p-1}$. So, combining these inequalities yields

$$\|\mathbf{E}_{\mathcal{G}} f\|_p^p \leq \|f\|_p \cdot \|\mathbf{E}_{\mathcal{G}} f\|_p^{p-1} \tag{5.2}$$

We now obtain $\|\mathbf{E}_{\mathcal{G}} f\|_p \leq \|f\|_p$ because this is obvious if $\|\mathbf{E}_{\mathcal{G}} f\|_p = 0$ and if not, then we can divide out in (5.2) because we know that $\|\mathbf{E}_{\mathcal{G}} f\|_p < \infty$.

The inequality $\|\mathbf{E}_{\mathcal{G}} f\|_1 \leq \|f\|_1$ corresponding to $p = 1$ also holds and is even simpler to establish since we obtain directly

$$\int |\mathbf{E}_{\mathcal{G}} f| d\mu = \int f g d\mu \leq \|f\|_1 \|g\|_\infty = \|f\|_1.$$

We have obtained that $\mathbf{E}_{\mathcal{G}}$ is a linear mapping from V to $L^p(X, \mathcal{F}, \mu)$, norm decreasing for the L^p norm. Since V is dense in $L^p(X, \mathcal{F}, \mu)$ and $L^p(X, \mathcal{F}, \mu)$ is complete, we can extend this mapping to a norm decreasing linear mapping $\mathbf{E}_{\mathcal{G}} : L^p(X, \mathcal{F}, \mu) \longrightarrow L^p(X, \mathcal{F}, \mu)$ by uniform continuity. We naturally use the same notation for this mapping, although strictly speaking it is a different mapping. This gives a nice application of "abstract nonsense" ideas to a really quite practical situation. You can check that the extended mapping satisfies the expected conditions which are valid even in the case $p = \infty$ handled earlier.

- $\mathbf{E}_{\mathcal{G}}(f) \in L^p(X, \mathcal{G}, \mu)$ provided $f \in L^p(X, \mathcal{F}, \mu)$ and indeed we have $\|\mathbf{E}_{\mathcal{G}} f\|_p \leq \|f\|_p$.
- $\int (f - \mathbf{E}_{\mathcal{G}}(f)) g d\mu = 0$ whenever $g \in L^{p'}(X, \mathcal{G}, \mu)$.

This is a typical example of the von Neumann program at work by using Hilbert space methods as a foot in the door to get results that have no obvious connection to Hilbert space.

5.3 Linear Forms on Hilbert Space

THEOREM 81 *Let H be a Hilbert space and let L be a continuous linear map from H to the base field. Then there exists $z \in H$ such that $L(x) = \langle z, x\rangle$.*

If $L \equiv 0$ then we just take $z = 0_H$. So, we can assume that $\ker(L)$ is a closed proper linear subspace of H. Then $\ker(L)^\perp$ cannot be the zero subspace, because together with $\ker(L)$, $\ker(L)^\perp$ must span the whole of H. So, choose from $\ker(L)^\perp$ a unit vector u. Now let $x \in H$ and consider $y = L(x)u - L(u)x$. Then of course $L(y) = L(x)L(u) - L(u)L(x) = 0$ so that $y \in \ker(L)$. So, we must have

$$0 = \langle u, y\rangle = \langle u, L(x)u - L(u)x\rangle = L(x)\langle u, u\rangle - L(u)\langle u, x\rangle = L(x) - \langle \overline{L(u)}u, x\rangle.$$

Take $z = \overline{L(u)}u$ and we are done. ■

5.4 Orthonormal Sets

An orthonormal set is usually an indexed set $(e_\alpha)_{\alpha\in I}$ where I is the indexing set. The key property that it has to satisfy is

$$\langle e_\alpha, e_\beta\rangle = \begin{cases} 1 & \text{if } \alpha = \beta, \\ 0 & \text{if } \alpha \neq \beta. \end{cases}$$

Given a finite linearly independent set in an inner product space, one usually constructs an orthonormal set by using the Gram–Schmidt Orthogonalization Process.

THEOREM 82 *Let $(e_\alpha)_{\alpha\in I}$ be an orthonormal set. Then*

(i) *If $(c_\alpha) \in \ell^2$, then the series $\sum_{\alpha\in I} c_\alpha e_\alpha$ is a norm convergent unconditional sum and furthermore $\| \sum_{\alpha\in I} c_\alpha e_\alpha\|_H = \left\{\sum_{\alpha\in I} |c_\alpha|^2\right\}^{1/2}$.*

(ii) *If $x \in H$, then $\sum_{\alpha\in I} |\langle e_\alpha, x\rangle|^2 \leq \|x\|^2$.*

(iii) *If M is the closed linear span (i.e. the closure of the linear span) of $(e_\alpha)_{\alpha\in I}$, then we have*

$$P(x) = \sum_{\alpha\in I} \langle e_\alpha, x\rangle e_\alpha$$

where P is orthogonal projection on M.

Proof. We are using the notation ℓ^2 to stand for $L^2(I, \mathcal{P}_I, \gamma)$ where γ is the counting measure on I. Usually, I will be countable and that would make things a bit simpler. However, we make the effort to understand the situation in case that I is uncountable.

We can treat $\sum_{\alpha\in I} |c_\alpha|^2$ as $\int |c_\alpha|^2 d\gamma(\alpha)$ and as the integral of a nonnegative measurable function it is defined as a supremum of integrals of dominated measurable simple functions carried on sets of finite measure. For the counting measure, this means finite sets, so it is not difficult to see that

$$\sum_{\alpha\in I} |c_\alpha|^2 = \sup_{\substack{F \text{ finite} \\ F\subset I}} \sum_{\alpha\in F} |c_\alpha|^2.$$

Another important point is that if $\sum_{\alpha\in I} |c_\alpha|^2 < \infty$ then $c_\alpha = 0$ for all but countably many α. To see this, we let $t > 0$ and invoke the Tchebychev Inequality

$$\gamma(\{\alpha; |c_\alpha| > t\}) \le t^{-2} \sum_{\alpha\in I} |c_\alpha|^2.$$

In particular, $\{\alpha; |c_\alpha| > t\}$ is a finite set. Now take $t_k \downarrow 0$, then

$$\{\alpha; |c_\alpha| > 0\} = \bigcup_{k=1}^{\infty} \{\alpha; |c_\alpha| > t_k\}$$

a countable union of countable sets and hence countable.

The sum $\sum_{\alpha\in I} c_\alpha e_\alpha$ on the other hand is an uncountable Hilbert space valued sum. It cannot be interpreted as an integral over the counting measure, because we do not have a vector-valued integration theory (outside the scope of this course). It can be interpreted as an unconditional sum. By the statement

$$s = \sum_{\alpha\in I} v_\alpha$$

in the unconditional sense, we mean that for all $\epsilon > 0$, there exists F finite with $F \subseteq I$ such that

$$\left\| s - \sum_{\alpha\in G} v_\alpha \right\| < \epsilon$$

for every finite set G with $F \subseteq G \subseteq I$.

The partial sum of the series $\sum_{\alpha\in I} c_\alpha e_\alpha$ corresponding to a finite subset F of I is given by

$$s_F = \sum_{\alpha\in F} c_\alpha e_\alpha.$$

We get

$$\begin{aligned}
\|s_F\|^2 = \langle s_F, s_F\rangle &= \left\langle \sum_{\beta\in F} c_\beta e_\beta, \sum_{\alpha\in F} c_\alpha e_\alpha \right\rangle \\
&= \sum_{\beta\in F}\sum_{\alpha\in F} \overline{c_\beta} c_\alpha \langle e_\beta, e_\alpha\rangle \\
&= \sum_{\alpha\in F} \overline{c_\alpha} c_\alpha \\
&= \sum_{\alpha\in F} |c_\alpha|^2 \qquad (5.3)
\end{aligned}$$

Now, we are assuming that $\sum_{\alpha\in I} |c_\alpha|^2 < \infty$, so all but countably many of the c_α are zero. Let us enumerate $\{\alpha; c_\alpha \neq 0\}$ as $(\alpha_n)_{n=1}^\infty$. If only finitely many c_α are nonzero, then the result is straightforward. So build a sequence of partial sums $s_N = \sum_{n=1}^N c_{\alpha_n} e_{\alpha_n}$. It is easy to see that this is a Cauchy sequence in H because for $M > N$ we have by (5.3) that

$$\|s_M - s_N\|^2 = \sum_{n=N+1}^{M} |c_{\alpha_n}|^2$$

So, since H is complete, we have that $s_n \longrightarrow s$ as $n \longrightarrow \infty$ for some $s \in H$. Now, let $\epsilon > 0$. Choose N so large that two things happen

- $\|s - s_n\| < \epsilon$ for $n \geq N$.
- $\sum_{n=N+1}^\infty |c_{\alpha_n}|^2 < \epsilon^2$.

Take $F = \{\alpha_n, n = 1, 2, \ldots, N\}$. Then, if G is a finite subset of I with $G \supseteq F$ we have

$$\|s - s_G\| \leq \|s - s_F\| + \|s_F - s_G\| < \epsilon + \epsilon = 2\epsilon$$

since

$$\|s_F - s_G\|^2 \leq \sum_{n=N+1}^{\infty} |c_{\alpha_n}|^2 < \epsilon^2$$

and

$$\|s - s_F\| = \|s - s_N\| < \epsilon.$$

This establishes the unconditional convergence. For a discussion of unconditional convergence, see the notes for MATH 255. Observe however that the result stated there that an unconditionally convergent series of real numbers is necessarily absolutely convergent, does not generalize to series of vectors in a complete normed vector space.

Since the norm is continuous, we have

$$\|s\|^2 = \limsup_{n\to\infty} \|s_n\|^2 = \limsup_{n\to\infty} \sum_{j=1}^{n} |c_{\alpha_j}|^2 = \sum_{\alpha\in I} |c_\alpha|^2.$$

This completes the proof of (i).

Now for the second part. It is enough to show that $\sum_{\alpha\in F} |\langle e_\alpha, x\rangle|^2 \le \|x\|^2$ for every finite subset F of I. Let M_F be the linear span of $(e_\alpha)_{\alpha\in F}$. Then M_F is closed. This is because the mapping $(c_\alpha)_{\alpha\in F} \mapsto \sum_{\alpha\in F} c_\alpha e_\alpha$ is an isometric linear mapping from $\mathbb{C}^n$ onto M_F. Since $\mathbb{C}^n$ is complete, so is M_F and therefore M_F must be closed in H[1]. Then we have

$$\left\langle x - \sum_{\alpha\in F} \langle e_\alpha, x\rangle e_\alpha, e_\beta \right\rangle = \langle x, e_\beta\rangle - \sum_{\alpha\in F} \overline{\langle e_\alpha, x\rangle} \langle e_\alpha, e_\beta\rangle = 0.$$

So $x - \sum_{\alpha\in F} \langle e_\alpha, x\rangle e_\alpha$ is in $M_F^\perp$. So, we can write

$$x = \overbrace{\left(\sum_{\alpha\in F} \langle e_\alpha, x\rangle e_\alpha\right)}^{\in M_F} + \overbrace{\left(x - \sum_{\alpha\in F} \langle e_\alpha, x\rangle e_\alpha\right)}^{\in M_F^\perp}$$

Since there is only one way of splitting a vector up in a direct sum decomposition, it must be that $P_F(x) = \sum_{\alpha\in F} \langle e_\alpha, x\rangle e_\alpha$ where P_F is orthogonal projection on M_F. Since P_F is norm decreasing, it follows that $\|x\| \ge \|\sum_{\alpha\in F} \langle e_\alpha, x\rangle e_\alpha\|$. Squaring this inequality and using the orthogonality once again, gives the desired result.

$$\|x\|^2 \ge \|\sum_{\alpha\in F} \langle e_\alpha, x\rangle e_\alpha\|^2 = \sum_{\alpha\in F}\sum_{\beta\in F} \overline{\langle e_\alpha, x\rangle} \langle e_\beta, x\rangle \langle e_\alpha, e_\beta\rangle = \sum_{\alpha\in F} |\langle e_\alpha, x\rangle|^2.$$

[1] As an exercise, show that any finite dimensional linear subspace of a normed linear space is necessarily closed.

Finally, in the third part, we see that parts (i) and (ii) guarantee the convergence of the series $\sum_{\alpha\in I}\langle e_\alpha, x\rangle e_\alpha$ in H norm. It is a norm limit of finite linear combinations of the e_α, so it is in M. We will show that

$$x - \sum_{\alpha\in I}\langle e_\alpha, x\rangle e_\alpha \perp e_\beta$$

for all $\beta \in I$. The relevant inner product is just

$$\left\langle e_\beta, x - \sum_{\alpha\in I}\langle e_\alpha, x\rangle e_\alpha \right\rangle = \langle e_\beta, x\rangle - \sum_{\alpha\in I}\langle e_\alpha, x\rangle\langle e_\beta, e_\alpha\rangle = 0.$$

Taking linear combinations, $x - \sum_{\alpha\in I}\langle e_\alpha, x\rangle e_\alpha$ is orthogonal to all finite linear combinations of the e_β. But the orthogonal subspace of a vector is closed, so it contains the closure of the linear span of the e_β, i.e. M. We have

$$x = \overbrace{\left(\sum_{\alpha\in I}\langle e_\alpha, x\rangle e_\alpha\right)}^{\in M} + \overbrace{\left(x - \sum_{\alpha\in I}\langle e_\alpha, x\rangle e_\alpha\right)}^{\in M^\perp}$$

The result now follows. ∎

5.5 Orthonormal Bases

Let H be a Hilbert space. An orthonormal basis in H is a maximal orthonormal set. It turns out that in the finite dimensional case, orthonormal bases are simply linear bases that are also orthonormal. But, in the infinite dimensional case, orthonormal bases are never linear bases. First we need to address the question of existence or, more generally extension. In this setting, we'll simply work with *unindexed* sets.

LEMMA 83 *Every orthonormal set is contained in some orthonormal basis.*

The easy option would be to say that the proof is outside the scope of this course. In fact, it uses Zorn's Lemma named for Max Zorn, but in fact discovered by Kazimierz Kuratowski.

Zorn's lemma is equivalent to the axiom of choice, in the sense that either one together with the standard axioms of set theory is sufficient to prove the other.

It occurs in the proofs of several theorems of crucial importance, for instance the theorem that every vector space has a linear basis, the theorem that every field has an algebraic closure and that every ring has a maximal ideal. Some high-powered theorems in topology and functional analysis also use Zorn's Lemma. It is stated as follows.

LEMMA 84 (ZORN'S LEMMA) *Every non-empty partially ordered set in which every chain which is bounded above contains a maximal element.*

The terms are defined as follows. Suppose $(X, \leq)$ is a partially ordered set. Explicitly this means that $\leq$ is a relation on X satisfying the following axioms

- $x \leq x$ for all $x \in X$.
- $x, y \in X, x \leq y, y \leq x \Longrightarrow x = y$.
- $x, y, z \in X, x \leq y, y \leq z \Longrightarrow x \leq z$.

A subset C of X is chain if for any $x, y \in X$ we have either $x \leq y$ or $y \leq x$. A subset Y of X is bounded above if there exists $u \in X$ such that $y \leq u$ for all $y \in Y$. Note that u is an element of X and need not be an element of Y. A maximal element of X is an element $m \in X$ such that $x \in X$ and $m \leq x$ implies $x = m$.

Proof of Lemma 83. The proof uses the axiom of choice in the form of Zorn's Lemma. Let E be the given orthonormal set and consider $\mathcal{E}$ the partially ordered set of all orthonormal subsets of H which contain E and ordered by inclusion. We need to know that $\mathcal{E}$ has a maximal element. It suffices to show then by Zorn's Lemma, that every chain $\mathcal{C}$ in $\mathcal{E}$ possesses an upper bound in $\mathcal{E}$. So, let $\mathcal{C}$ be such a chain and define $F = \bigcup_{C \in \mathcal{C}} C$. We claim that $F \in \mathcal{E}$ and it is evident that $C \subseteq F$ for all $C \in \mathcal{C}$.

It suffices to check the claim. Obviously $F \supseteq E$, so it remains only to show that F is an orthonormal set. Let $f \in F$, then $f \in C$ for some $C \in \mathcal{C}$. So f must be a unit vector. Next choose two distinct vectors f_1 and f_2 in F. Then $f_1 \in C_1$ and $f_2 \in C_2$ for some $C_1, C_2 \in \mathcal{C}$. But $\mathcal{C}$ is a chain, so either $C_1 \subseteq C_2$ or $C_2 \subseteq C_1$. We suppose without loss of generality that the former holds. Then $f_2 \in C_2$ and $f_1 \in C_1 \subseteq C_2$. So, using the fact that C_2 is orthonormal and that $f_1 \neq f_2$ we have $f_1 \perp f_2$. We have just shown that F is orthonormal.

We conclude that $\mathcal{E}$ possesses a maximal element B. We clain that B is a maximal orthonormal set. Indeed if there were a larger one, then it would also contain E since already $E \subseteq B$ and hence would be in $\mathcal{E}$. But B is maximal in $\mathcal{E}$. ■

We need a theorem that characterizes orthonormal bases.

THEOREM 85 *Let $(e_\alpha)_{\alpha \in I}$ be an orthonormal set in a Hilbert space H. Then the following are equivalent.*

(i) $(e_\alpha)_{\alpha \in I}$ *is an orthonormal basis.*

(ii) *The closed linear span M of $(e_\alpha)_{\alpha \in I}$ is the whole of H.*

(iii) *The identity*

$$\sum_{\alpha \in I} |\langle e_\alpha, x \rangle|^2 = \|x\|^2$$

holds for all $x \in H$.

(iv) *The identity*

$$\sum_{\alpha \in I} \langle y, e_\alpha \rangle \langle e_\alpha, x \rangle = \langle y, x \rangle$$

holds for all $x, y \in H$.

Proof.
(i) $\Longrightarrow$ (ii). If not, then $M^\perp \neq \{0_H\}$. So, there is a unit vector $e \in M^\perp$. So, $\langle e, e \rangle = 1$, $\langle e, e_\alpha \rangle = 0$ for all $\alpha \in I$ and $\langle e_\alpha, e \rangle = 0$ for all $\alpha \in I$. Given that $(e_\alpha)_{\alpha \in I}$ is an orthonormal set, these are the conditions needed to ensure that $\{e\} \cup \{e_\alpha; \alpha \in I\}$ is also an orthonormal set. So the maximality of $(e_\alpha)_{\alpha \in I}$ is contradicted.

(ii) $\Longrightarrow$ (i). We are assuming that $M = H$ and that e is a vector which together with $(e_\alpha)_{\alpha \in I}$ still gives an orthonormal set. Then $\langle e, e_\alpha \rangle = 0$. Therefore $\langle e, v \rangle = 0$ where v is in the linear span of $(e_\alpha)_{\alpha \in I}$. Now, taking limits along a sequence of such v we have $\langle e, x \rangle = 0$ for all $x \in M$. So, in particular, this holds for $x = e$ and we obtain $1 = \langle e, e \rangle = 0$. This contradiction establishes the claim.

(ii) $\Longrightarrow$ (iii). By Theorem 82 item (iii), we have from the fact that orthogonal projection on M is the identity mapping

$$x = \sum_{\alpha \in I} \langle e_\alpha, x \rangle e_\alpha$$

and indeed, this is a norm convergent sum. Let s_F be a partial sum of this series corresponding to a finite subset F of I, then we have $s_F \longrightarrow x$ in the sense of unconditional convergence and a calculation gives

$$\|s_F\|^2 = \langle s_F, s_F \rangle = \sum_{\alpha \in I} \sum_{\beta \in I} \overline{\langle e_\alpha, x \rangle} \langle e_\beta, x \rangle \langle e_\alpha, e_\beta \rangle = \sum_{\alpha \in F} |\langle e_\alpha, x \rangle|^2$$

since $\langle e_\alpha, e_\beta \rangle = 1$ if $\alpha = \beta$ and 0 otherwise. Since $\|s_F\| \longrightarrow \|x\|$, we have the desired equality.

(ii) $\Longrightarrow$ **(iv)** follows by essentially the same argument.

(iv) $\Longrightarrow$ **(iii)** follows by setting $y = x$.

(iii) $\Longrightarrow$ **(i)**. Let e be a vector which together with $(e_\alpha)_{\alpha \in I}$ still gives an orthonormal set. Then we have

$$1 = \|e\|^2 = \sum_{\alpha \in F} |\langle e_\alpha, e \rangle|^2 = 0.$$

This contradiction completes the proof. ∎

There are some important consequences of this result and the existence of orthonormal bases.

COROLLARY 86

(i) *If H is a finite dimensional Hilbert space, then it is linearly isometric to d dimensional Euclidean space $\mathbb{R}^d$ or $\mathbb{C}^d$, depending on the field of scalars and where $d = \dim(H)$.*

(ii) *If H is infinite dimensional, but separable Hilbert space, then it is linearly isometric to ℓ^2 over the appropriate field of scalars.*

Proof. Statement (i) follows directly from Theorem 85. Statement (ii) follows the same route, but we need to be sure that there is a denumerable orthonormal basis. So, let $(e_\alpha)_{\alpha \in I}$ be an orthonormal basis and let $(x_j)_{j=1}^\infty$ be a sequence dense in H. Now we find for $\alpha \neq \beta$, $\|e_\alpha - e_\beta\|^2 = \|e_\alpha\|^2 + \|e_\beta\|^2 = 2$. For each $\alpha \in I$, choose a value $j = J(\alpha)$, such that $\|e_\alpha - x_{J(\alpha)}\| < 1/2$. Then, obviously J is one-to-one as a mapping $J : I \longrightarrow \mathbb{N}$. This is because $J(\alpha) = J(\beta)$ will imply

$$\sqrt{2} = \|e_\alpha - e_\beta\| \leq \|e_\alpha - x_{J(\alpha)}\| + \|e_\beta - x_{J(\beta)}\| < 1$$

So, I must be countable. It cannot be finite, so it is denumerable. ∎

6

Convergence of Functions

In this very brief chapter, we look at the different ways that functions can converge. Let $(f_n)_{n=1}^{\infty}$ be a sequence of measurable functions defined μ-almost everywhere on a measure space $(X, \mathcal{F}, \mu)$. Let f be a potential limit function also defined μ-almost everywhere. Then we can consider $f_n \longrightarrow f$ in a number of different senses.

- **Convergence in p-mean:** $\|f_n - f\|_p \longrightarrow 0$ as $n \longrightarrow \infty$. It is understood that $1 \leq p < \infty$.

- **Convergence μ-almost everywhere:** $f_n(x) \longrightarrow f(x)$ on the set $X \setminus N$ where $N \in \mathcal{F}$ and $\mu(N) = 0$. It is understood that N contains the sets where f_n and f are not defined. This kind of convergence is called **almost sure convergence** or **convergence with probability one** in the probabilistic setting.

- **Convergence in measure:** $\mu(\{x; |f(x) - f_n(x)| \geq t\}) \longrightarrow 0$ and $n \longrightarrow \infty$ for each $t > 0$ fixed. In the probabilistic setting, this is called **convergence in probability**. This is a rather weak type of convergence, but very heavily used.

One further type of convergence is also of importance in probability theory. This is convergence in distribution. It is equivalent to the pointwise convergence of the distribution functions of the random variables concerned. This is a different setting, because the f_n and f can be defined on *different* probability spaces.

LEMMA 87 *Convergence in mean implies convergence in measure.*

Proof. This is a consequence of Tchebychev's inequality.

$$\mu(\{x; |f(x) - f_n(x)| \geq t\}) \leq \frac{\|f_n - f\|_p^p}{t^p} \longrightarrow 0$$

as $n \longrightarrow \infty$ with $t > 0$ fixed. ■

THEOREM 88 (EGOROV'S THEOREM) *Let $(X, \mathcal{F}, \mu)$ be a finite measure space i.e. $\mu(X) < \infty$. Let $f_n \longrightarrow f$ μ-almost everywhere. Let $\delta > 0$. Then there is a set N with $\mu(N) < \delta$ and such that $f_n \longrightarrow f$ uniformly on $X \setminus N$. In particular, on a finite measure space, convergence almost everywhere implies convergence in measure.*

Proof. Let F be the set on which convergence fails. Let $\epsilon > 0$, and define $S_m(\epsilon) = \{x; |f_n(x) - f(x)| < \epsilon \text{ for all } n \geq m\}$. Clearly $S_m(\epsilon)$ increases with m and

$$X \setminus F \subseteq \bigcup_{m=1}^{\infty} S_m(\epsilon).$$

Therefore, $\mu(X) = \mu(X \setminus F) = \sup_{m=1}^{\infty} \mu(S_m(\epsilon))$ and, since we are in a finite measure space, $\inf_{m=1}^{\infty} \mu(X \setminus S_m(\epsilon)) = 0$. Now, let $\delta_k = 2^{-k}\delta$ and $\epsilon_k \downarrow 0$.

Working with $k = 1$, we can find m_1 such that

$$\mu(X \setminus S_{m_1}(\epsilon_1)) < \delta_1$$

For $k = 2$, we can find $m_2 > m_1$ such that

$$\mu(X \setminus S_{m_2}(\epsilon_2)) < \delta_2$$

For $k = 3$, we can find $m_3 > m_2$ such that

$$\mu(X \setminus S_{m_3}(\epsilon_3)) < \delta_3$$

$$\cdots\cdots\cdots$$

Let now $S = \bigcap_{k=1}^{\infty} S_{m_k}(\epsilon_k))$, then $\mu(X \setminus S) < \sum_{k=1}^{\infty} \delta_k = \delta$. We claim that convergence is uniform on S. Indeed, let $\epsilon > 0$. We find k such that $\epsilon_k < \epsilon$. Then $n \geq m_k$ implies that $|f(x) - f_n(x)| < \epsilon$ for $n \geq m_k$ and $x \in S$. We take $N = X \setminus S$. ■

Egorov's Theorem fails even in the σ-finite case. Let $f \equiv 0$ and $f_n = \mathbb{1}_{[n,\infty[}$ with μ Lebesgue measure on the line. Convergence holds everywhere, but for each n we have $|f - f_n| = 1$ on a set of infinite measure.

Finally, we remark that there are no obvious implications between convergence in mean and almost everywhere convergence. Either can hold without the other.

7

Fourier Series

Fourier came upon Fourier series in his search for a solution to the heat equation. The idea actually generalizes in a variety of ways and is central to a lot of advanced analysis. The simplest way of coming to grips with the subject is via orthonormal series.

The space on which the action takes place is the circle $\mathbb{T}$ which we can realize in several ways. One is as $\mathbb{R}/2\pi\mathbb{Z}$ which is the quotient group of the real line by the subgroup $\{2\pi n; n \in \mathbb{Z}\}$. In this model, functions on $\mathbb{T}$ are thought of as functions on $\mathbb{R}$ which are 2π-periodic, i.e.

$$f(x + 2n\pi) = f(x), \qquad \forall n \in \mathbb{Z}.$$

Another model of $\mathbb{T}$ is the interval $[0, 2\pi]$ with the endpoints identified and yet another is as the interval $[-\pi, \pi]$ with the endpoints identified. We work with the Lebesgue σ-field $\mathcal{L}$ on $\mathbb{T}$ and the normalized linear measure η. The normalization gives the circle $\mathbb{T}$ a measure of 1. So, we have say $d\eta(t) = \dfrac{1}{2\pi}dt$ on $[0, 2\pi]$. Everything hinges on the function $e_n(t) = e^{int}$ for $n \in \mathbb{Z}$.

LEMMA 89 *We have* $\displaystyle\int e_n d\eta = \begin{cases} 1 & \text{if } n = 0, \\ 0 & \text{if } n \neq 0. \end{cases}$

Proof. The case $n = 0$ follows from $e_0 = \mathbb{1}$. For the case $n \neq 0$, we use

$$\begin{aligned} \int e_n d\eta &= \frac{1}{2\pi} \int_0^{2\pi} \Big(\cos(nt) + i \sin(nt) \Big) dt \\ &= \frac{1}{2\pi} \left[\frac{\sin(nt)}{n} - i \frac{\cos(nt)}{n} \right]_0^{2\pi} = 0. \end{aligned}$$

We now have

LEMMA 90 *The set of functions $(e_n)_{n\in\mathbb{Z}}$ is an orthonormal set in $L^2(\mathbb{T},\mathcal{L},\eta)$.*

Proof. We have $\langle e_m, e_n\rangle = \int \overline{e_m} e_n d\eta = \int e_{n-m} d\eta = \begin{cases} 1 & \text{if } m = n, \\ 0 & \text{if } m \neq n. \end{cases}$ ∎

THEOREM 91 *In fact, $(e_n)_{n\in\mathbb{Z}}$ is an orthonormal basis in $L^2(\mathbb{T},\mathcal{L},\eta)$.*

Proof. We consider the set of functions

$$A = \{\sum_{k=0}^{n} a_k \cos(kt) + \sum_{k=1}^{n} b_k \sin(kt); n \in \mathbb{N}, a_k, b_k \in \mathbb{R}\}$$

It is fairly straightforward to verify that A is closed under pointwise linear combinations and pointwise products. To see this for the products, we need identities like

$$\cos(kt)\cos(\ell t) = \frac{1}{2}\Big(\cos((k-\ell)t) + \cos((k+\ell)t)\Big).$$

We leave the full justification to the reader. It is also clear that the functions of A separate the points of $\mathbb{T}$. So, by the Stone–Weierstrass Theorem, A is dense in the real-valued $C(\mathbb{T})$ for the uniform norm. But $C(\mathbb{T})$ is dense in the real-valued $L^2(\mathbb{T},\mathcal{L},\eta)$. So, A is dense in the real-valued $L^2(\mathbb{T},\mathcal{L},\eta)$. Finally, it follows that the linear span of $(e_n)_{n\in\mathbb{Z}}$ is dense in the complex-valued $L^2(\mathbb{T},\mathcal{L},\eta)$. Since we already know that $(e_n)_{n\in\mathbb{Z}}$ is an orthonormal set in $L^2(\mathbb{T},\mathcal{L},\eta)$, we have the desired conclusion. ∎

We define

$$\hat{f}(n) = \langle e_n, f\rangle = \int \overline{e_n} f d\eta$$

the n^{th} Fourier coefficient of f. It is actually well-defined if $f \in L^1(\mathbb{T},\mathcal{L},\eta)$. Usually with Fourier series we work with symmetric partial sums $S_N f$ defined by

$$S_N f(t) = \sum_{n=-N}^{N} \hat{f}(n) e_n(t).$$

COROLLARY 92 (PLANCHEREL'S THEOREM) *Let* $f \in L^2(\mathbb{T}, \mathcal{L}, \eta)$. *Then we have*

- $f = \sum_{n\in\mathbb{Z}} \langle e_n, f\rangle e_n$. *as an unconditional sum in* $L^2(\mathbb{T}, \mathcal{L}, \eta)$.
- $S_N f \longrightarrow f$ *in* L^2 *norm as* $N \longrightarrow \infty$.
- $\|f\|_2^2 = \sum_{n\in\mathbb{Z}} |\langle e_n, f\rangle e_n|^2 = \sum_{n\in\mathbb{Z}} \left|\hat{f}(n)\right|^2$.

7.1 Dirichlet and Féjer Kernels

We define

$$D_N(t) = \sum_{n=-N}^{N} e_n(t) = \frac{\sin(N+\frac{1}{2})t}{\sin\frac{1}{2}t}$$

obtained from summing a geometric series. This is the Dirichlet kernel. For the Féjer kernel, we have

$$K_N(t) = \sum_{n=-N}^{N} \left(1 - \frac{|n|}{N}\right) e_n(t) = \frac{\sin^2(\frac{Nt}{2})}{N\sin^2\frac{t}{2}}.$$

Note that the terms $n = \pm N$ are not needed in the sum. The key to establishing this awkward summation is

$$\sum_{n=0}^{N-1} e_n(t) = e^{i\frac{(N-1)t}{2}} \frac{\sin(\frac{Nt}{2})}{\sin\frac{t}{2}}$$

obtained again by summing a geometric series. Taking the absolute value squared gives the desired result. One needs to verify that

$$\sum_{p=0}^{N-1}\sum_{q=0}^{N-1} \overline{e_p(t)} e_q(t) = \sum_{p=0}^{N-1}\sum_{q=0}^{N-1} e_{q-p}(t) = \sum_{n=-N}^{N} (N - |n|)\, e_n(t)$$

which is proved by counting terms.

The Dirichlet kernel is badly behaved and so, correspondingly are the partial sums of Fourier series. The Féjer kernel, on the other hand is an example of what we call a ***summability kernel*** and it has nice properties. These are

- $K_N(t) \geq 0$.
- $\|K_N\|_1 = 1$.
- For $\delta > 0$ fixed, we have $\int_{|t|\geq\delta} K_N(t)d\eta(t) \longrightarrow 0$ as $N \longrightarrow \infty$.

The first of these conditions is obvious. For the second we have

$$\begin{aligned}
\|K_N\|_1 &= \int K_N(t)d\eta(t) \\
&= \int \sum_{n=-N}^{N} \left(1 - \frac{|n|}{N}\right) e_n(t)d\eta(t) \\
&= \sum_{n=-N}^{N} \left(1 - \frac{|n|}{N}\right) \int e_n(t)d\eta(t) = 1,
\end{aligned}$$

because only the term $n = 0$ in the last line survives. For the final condition, we need to get our hands dirty. We work on $[-\pi, \pi]$. We have

$$K_N(t) = \frac{\sin^2(\frac{Nt}{2})}{N\sin^2\frac{t}{2}} \leq \frac{1}{N\sin^2\frac{t}{2}} \leq \frac{\pi^2}{Nt^2}$$

using the inequality $\sin(u) \geq \dfrac{2}{\pi}u$ for $0 < u \leq \pi/2$. Thus,

$$\begin{aligned}
\int_{|t|\geq\delta} K_N(t)d\eta(t) &= \frac{1}{2\pi}\int_{|t|\geq\delta} K_N(t)dt \\
&= \frac{1}{\pi}\int_{\delta}^{\pi} K_N(t)dt \\
&\leq \frac{1}{\pi}\int_{\delta}^{\infty} \frac{\pi^2}{Nt^2}dt \\
&= \pi N^{-1}\delta^{-1} \longrightarrow 0 \qquad \text{as } n \longrightarrow \infty.
\end{aligned}$$

Next, we'll introduce the ***convolution product*** . Eventually, this will be defined for $f, g \in L^1(\mathbb{T}, \mathcal{L}, \eta)$. We set

$$f \star g(t) = \int f(t-s)g(s)d\eta(s).$$

The difference $t - s$ is taken modulo 2π. The key fact about convolutions that we need at the moment is

$$e_n \star f = \hat{f}(n) e_n$$

and this is straightforward to verify.

$$\begin{aligned} e_n \star f(t) &= \int e_n(t-s) f(s) d\eta(s) = \int e_n(t) \overline{e_n(s)} f(s) d\eta(s) \\ &= \int \overline{e_n(s)} f(s) d\eta(s) e_n(t) = \hat{f}(n) e_n(t). \end{aligned}$$

We'll start by showing the following theorem.

THEOREM 93 *Let $f \in C(\mathbb{T})$. Then $K_N \star f \longrightarrow f$ in $C(\mathbb{T})$ as $N \longrightarrow \infty$.*

Proof. The proof of this result should be compared with the proof of the Bernstein Approximation Theorem. The idea that is used is exactly the same. The first step is to capture the cancellation. We have

$$(f - K_n \star f)(t) = f(t) - \int K_n(t-s) f(s) d\eta(s) = \int K_n(t-s) \Big(f(t) - f(s) \Big) d\eta(s)$$

because $\int K_n(t-s) d\eta(s) = 1$. Next, we put in the absolute values.

$$\begin{aligned} &|f - K_n \star f|(t) \\ &\leq \int K_n(t-s) \Big| f(t) - f(s) \Big| d\eta(s) \\ &= \int_{|t-s|<\delta} K_n(t-s) \Big| f(t) - f(s) \Big| d\eta(s) + \int_{|t-s|\geq\delta} K_n(t-s) \Big| f(t) - f(s) \Big| d\eta(s) \\ &\leq \omega_f(\delta) \int_{|t-s|<\delta} K_n(t-s) d\eta(s) + 2\|f\|_\infty \int_{|t-s|\geq\delta} K_n(t-s) d\eta(s) \end{aligned}$$

using $|f(t) - f(s)| \leq \omega_f(\delta)$ for $|t - s| < \delta$ and $|f(t) - f(s)| \leq 2\|f\|_\infty$ always.

$$\leq \omega_f(\delta) + 2\pi \|f\|_\infty n^{-1} \delta^{-1}$$

Let $\epsilon > 0$. Then first choose $\delta > 0$ so small that $\omega_f(\delta) < \epsilon/2$ and then, choose N so large that $2\pi\|f\|_\infty N^{-1}\delta^{-1} < \epsilon/2$. We find that

$$|f - K_n \star f|(t) < \epsilon$$

for all $n \geq N$. Note that N does not depend on t so that convergence is uniform.

■

7.2 The Uniform Boundedness Principle

In this section we will develop the uniform boundedness principle which is an abstract principle which applies to Banach spaces (complete normed vector spaces). We will then see how that principal can be applied to show that there exist functions $f \in C(\mathbb{T})$ such that $S_n f$ fails to converge to f even in the pointwise sense.

We will need some general ideas from Banach space theory. First of all, recall that if u is a continuous linear form on a Banach space B, (i.e. a continuous linear mapping from B to the base field, $\mathbb{R}$ or $\mathbb{C}$) then there is a constant C such that $|u(x)| \leq C\|x\|_B$. The smallest constant C that works is taken to be the norm of u. Precisely

$$\|u\|_{B'} = \sup_{\substack{x \in B \\ \|x\|_B \leq 1}} |u(x)|$$

It can actually be shown that the linear space of all bounded linear forms u on B can be made into a Banach space with this norm, but we do not need this fact here.

THEOREM 94 (UNIFORM BOUNDEDNESS PRINCIPLE) *Suppose that $(u_n)_{n=1}^{\infty}$ is a sequence of linear forms on B such that for every fixed $x \in B$, the sequence $(u_n(x))_{n=1}^{\infty}$ is a bounded sequence of scalars. Then there is a constant C such that $\|u_n\|_{B'} \leq C$ for all $n \in \mathbb{N}$.*

Proof. The proof is non obvious and depends on the Baire Category Theorem. It should be compared with the Open Mapping Theorem. For $k \in \mathbb{N}$ we define

$$A_k = \{x; x \in B, |u_n(x)| \leq k \quad \forall n \in \mathbb{N}\}.$$

Then A_k is a closed subset of the complete normed space B for every $k \in \mathbb{N}$ because each u_n is continuous and an arbitrary intersection of closed sets is closed. By hypothesis we have

$$B = \bigcup_{k=1}^{\infty} A_k.$$

So, according to the Baire Category Theorem, there exists $k \in \mathbb{N}$ such that A_k has nonempty interior. From now on in this proof, we denote by k that specific k. So, there exists $x \in B$ and $t > 0$ such that $U_B(x,t) \subseteq A_k$. But A_k is a symmetric (i.e. $x \in A_k \implies -x \in A_k$) convex set. So, using first the symmetry, we have

$U_B(-x,t) \subseteq A_k$. Now we use the convexity. Let $y \in U_B(0_B,t)$. Then we can write

$$y = \frac{1}{2}\Big((x+y) + (-x+y)\Big)$$

and $\pm x + y \in U_B(\pm x, t) \subseteq A_k$. So, by convexity of A_k, $y \in A_k$. So, certainly, for all $n \in \mathbb{N}$

$$\|y\| \le \frac{t}{2} \Longrightarrow |u_n(y)| \le k$$

and indeed, by scaling

$$\|y\| \le 1 \Longrightarrow |u_n(y)| \le \frac{2k}{t}.$$

So, the conclusion holds with $C = \dfrac{2k}{t}$.

The next step in this saga is to compute a lower bound for the L^1 norm of the Dirichlet kernel.

LEMMA 95 *We have the lower bound*

$$\|D_N\|_1 \ge \frac{4}{\pi^2} \sum_{k=1}^{N-1} \frac{1}{k+1}$$

for the L^1 norm of the Dirichlet kernel. In particular, these norms are not bounded in N.

Proof. Note that the zeros of the Dirichlet kernel $D_N(t)$ occur at $t = \dfrac{2k\pi}{2N+1}$ as k runs from 1 to $2N$. The first N of these zeros are in the interval $[0,\pi]$ where the denominator in the Dirichlet kernel is nonnegative. We can therefore write

$$\begin{aligned}
\|D_N\|_1 = \int |D_N(t)| d\eta(t) &= \frac{1}{\pi} \int_0^\pi \frac{|\sin(N+\frac{1}{2})t|}{\sin\frac{1}{2}t} dt \\
&\ge \frac{1}{\pi} \sum_{k=1}^{N-1} \int_{\frac{2k\pi}{2N+1}}^{\frac{2(k+1)\pi}{2N+1}} \frac{|\sin(N+\frac{1}{2})t|}{\sin\frac{1}{2}t} dt
\end{aligned}$$

We change variable using $(N+\frac{1}{2})t = k\pi + s$, which now gives

$$= \frac{1}{\pi} \cdot \frac{2}{2N+1} \sum_{k=1}^{N-1} \int_0^\pi \frac{\sin s}{\sin\left(\frac{k\pi+s}{2N+1}\right)} ds$$

and indeed, $\sin\left(\frac{k\pi+s}{2N+1}\right) \leq \frac{(k+1)\pi}{2N+1}$ for $0 \leq s \leq \pi$

$$\geq \frac{1}{\pi} \cdot \frac{2}{2N+1} \sum_{k=1}^{N-1} \frac{2N+1}{(k+1)\pi} \int_0^\pi \sin s\, ds$$

$$= \frac{4}{\pi^2} \sum_{k=1}^{N-1} \frac{1}{k+1}.$$

PROPOSITION 96 *There exists a function $f \in C(\mathbb{T})$ such that $S_N f(0)$ is unbounded in N. In particular, the Fourier series of f does not converge pointwise.*

Proof. The idea of the proof is to apply the Uniform Boundedness Principle to $B = C(\mathbb{T})$. We define continuous linear forms u_n on B by

$$u_n(f) = S_n f(0) = \int D_n f d\eta.$$

If we can show that the norms of the u_n are unbounded, then it will follow that there exists at least one element $f \in B = C(\mathbb{T})$ such that $u_n(f)$ is an unbounded sequence of scalars. In particular then $u_n(f) = S_n f(0)$ cannot converge as $n \longrightarrow \infty$. But, by proposition 65, $\|D_n\|_{L^1} = \|u_n\|_{B'}$ and the proof is complete. ■

7.3 More about Convolution

If $f, g \in L^1(\mathbb{T}, \mathcal{L}, \eta)$, then we have

$$\int |f| \star |g| d\eta = \int \int |f(t-s)||g(s)| d\eta(s) d\eta(t)$$

$$= \int \int |f(t-s)||g(s)| d\eta(t) d\eta(s)$$

by applying Tonelli's Theorem

$$= \int |g(s)| \int |f(t-s)| d\eta(t) d\eta(s)$$

$$= \int |g(s)| \int |f(t)| d\eta(t) d\eta(s)$$

because η is translation invariant on $\mathbb{T}$

$$= \|f\|_1 \|g\|_1.$$

It follows that for almost all t, the integral $\int |f(t-s)||g(s)|d\eta(s)$ is finite. Hence, $f \star g(t)$ is well-defined for almost all t and indeed, $\|f \star g\|_1 \le \|f\|_1 \|g\|_1$.

To understand what the convolution actually is, we should verify the following identity. Let now $h \in L^\infty(\mathbb{T}, \mathcal{L}, \eta)$. Then

$$\int (f \star g)(t) h(t) d\eta(t) = \int f(t) g(s) h(t+s) d(\eta \otimes \eta)(t,s). \tag{7.1}$$

This is a consequence of Fubini's Theorem

$$\begin{aligned}\int (f \star g)(u) h(u) d\eta(u) &= \int \int f(u-s) g(s) d\eta(s) h(u) d\eta(u) \\ &= \int \int f(u-s) g(s) h(u) d\eta(u) d\eta(s) \\ &= \int \int f(t) g(s) h(t+s) d\eta(t) d\eta(s)\end{aligned}$$

after substituting $u = t+s$ in the inner integral and using the translation invariance of η

$$= \int f(t) g(s) h(t+s) d(\eta \otimes \eta)(t,s).$$

This can even be expressed in the form

$$\int (f \star g)(t) h(t) d\eta(t) = \int h(t+s) d(f \cdot \eta \otimes g \cdot \eta)(t,s).$$

where we have used the notation $f \cdot \eta$ for the measure $f \cdot \eta(A) = \int \mathbb{1}_A f d\eta$ for $A \in \mathcal{L}$. We leave it as an exercise for the reader to show that $f \cdot \eta$ actually is a measure.

This last formula suggests what we should now check, namely that $f \star g = g \star f$. This uses both the translation invariance and the reflection invariance of η. We have

$$\begin{aligned}(f \star g)(t) &= \int f(t-s) g(s) d\eta(s) \\ &= \int f(-v) g(t+v) d\eta(v)\end{aligned}$$

by using $v = s - t$ and translating the measure η by t

$$= \int f(u) g(t-u) d\eta(u)$$

by using $u = -v$ and reflecting the measure η

$$= g \star f(t).$$

What is the connection between convolution and the Fourier coefficients. We put $h(t) = \overline{e_n(t)}$ in (7.1).

$$\begin{aligned}
\widehat{(f \star g)}(n) &= \int (f \star g)(t)\overline{e_n(t)} d\eta(t) \\
&= \int f(t)g(s)\overline{e_n(t+s)} d(\eta \otimes \eta)(t,s) \\
&= \int f(t)g(s)\overline{e_n(t)e_n(s)} d(\eta \otimes \eta)(t,s) \\
&= \int f(t)\overline{e_n(t)} d\eta(t) \cdot \int g(s)\overline{e_n(s)} d\eta(s) \\
&= \hat{f}(n) \cdot \hat{g}(n).
\end{aligned}$$

So, the Fourier coefficients of the convolution product are just the pointwise products of the Fourier coefficients.

Finally, in this section we will prove the uniqueness theorem.

THEOREM 97 *Let $f \in L^1(\mathbb{T}, \mathcal{L}, \eta)$ and suppose that $\hat{f}(n) = 0$ for all $n \in \mathbb{Z}$. Then $f = 0$ in L^1, (i.e. f(t) = 0 for almost all t).*

Proof. Let $p(t) = \sum_{n=-N}^{N} p_n e_n(t)$ be a trigonometric polynomial. Then, we have

$$\int f(t)p(t)d\eta(t) = \sum_{n=-N}^{N} p_n \int f(t)e_n(t) = \sum_{n=-N}^{N} p_n \hat{f}(-n) = 0.$$

Now, let $g \in C(\mathbb{T})$ and consider $g_n = K_n \star g$. Now it is easy to see

- g_n is a trigonometric polynomial.
- $g_n \longrightarrow g$ in the uniform norm by Theorem 93.

It follows that $\int f(t)g(t)d\eta(t) = 0$ for all $g \in C(\mathbb{T})$. Hence f is the zero element of L^1 by Proposition 65, completing the proof. ■

COROLLARY 98 *If* $f \in L^1(\mathbb{T}, \mathcal{L}, \eta)$ *and* $\displaystyle\sum_{n \in \mathbb{Z}} \left|\hat{f}(n)\right|^2 < \infty$, *then* $f \in L^2(\mathbb{T}, \mathcal{L}, \eta)$.

Proof. Let $g = \sum_{n \in \mathbb{Z}} \hat{f}(n) e_n$ as an unconditional sum. Then $g \in L^2 \subset L^1$. Also $\hat{f}(n) = \hat{g}(n)$ for all $n \in \mathbb{Z}$. We now apply the Uniqueness Theorem to $f - g$ and deduce that $f = g \in L^2$. ∎

COROLLARY 99 (RIEMANN–LEBESGUE LEMMA) *If* $f \in L^1(\mathbb{T}, \mathcal{L}, \eta)$, *then*

$$\lim_{|n| \to \infty} \hat{f}(n) = 0.$$

Proof. We clearly have $\|\hat{f}\|_\infty \leq \|f\|_1$. This tells only that $\hat{f}$ is bounded. However, $L^2(\mathbb{T}, \mathcal{L}, \eta)$ is dense in $L^1(\mathbb{T}, \mathcal{L}, \eta)$ and we have $\hat{f} \in C_0(\mathbb{Z})$ for $f \in L^2$. For general $f \in L^1$, find $f_n \in L^2$ with $f_n \underset{n \to \infty}{\longrightarrow} f$ in L^1 norm. Then $\widehat{f_n} \underset{n \to \infty}{\longrightarrow} \hat{f}$ in the uniform norm. But since $\widehat{f_n} \in C_0(\mathbb{Z})$ and $C_0(\mathbb{Z})$ is closed for the uniform norm (in all bounded functions on $\mathbb{Z}$), we see that $\hat{f} \in C_0(\mathbb{Z})$ as required. ∎

8

Differentiation

We start with the remarkable Vitali Covering Lemma. The action takes place on $\mathbb{R}$, but equally well there are corresponding statement in $\mathbb{R}^d$. We denote by μ the Lebesgue measure.

LEMMA 100 (VITALI COVERING LEMMA) *Let $\mathcal{S}$ be a family of bounded open intervals in $\mathbb{R}$ and let S be a Lebesgue subset of $\mathbb{R}$ with $\mu(S) < \infty$ and such that*

$$S \subseteq \bigcup_{I \in \mathcal{S}} I.$$

Then, there exists $N \in \mathbb{N}$ and pairwise disjoint intervals $I_1, I_2, \ldots I_N$ of $\mathcal{S}$ such that

$$\mu(S) \leq 4 \sum_{n=1}^{N} \mu(I_n). \tag{8.1}$$

Proof. Since $\mu(S) < \infty$, there exists K compact, $K \subseteq S$ and $\mu(K) > \frac{3}{4}\mu(S)$. Now, since

$$K \subseteq \bigcup_{I \in \mathcal{S}} I$$

there are just finitely many intervals $J_1, J_2, \ldots, J_M$ with $K \subseteq \bigcup_{m=1}^{M} J_m$. Let these intervals be arranged in order of decreasing length. Thus $1 \leq m_1 < m_2 \leq M$ implies that $\mu(J_{m_1}) \geq \mu(J_{m_2})$. We will call this the J list. We proceed algorithmically. If the J list is empty, then $N = 0$ and we stop. Otherwise, let I_1 be the first element of the J list (in this case J_1). Now, remove from the J list, all intervals that meet I_1 (including I_1 itself). If the J list is empty, then $N = 1$ and we stop. Otherwise, let I_2 be the first remaining element of the J list. Now, remove from the J list, all intervals that meet I_2 (including I_2 itself). If the J list is empty, then $N = 2$ and we stop. Otherwise, let I_3 be the first remaining element of the J list. Now, remove from the J list, all intervals that meet $I_3 \ldots$

Eventually, the process must stop, because there are only finitely many elements in the J list to start with. Clearly, the I_n are pairwise disjoint, because if I_{n_1} meets I_{n_2} and $1 \le n_1 < n_2 \le N$, then, immediately after I_{n_1} was chosen, all those intervals of the J list which meet I_{n_1} were removed. Since I_{n_2} was eventually chosen from this list, it must be that $I_{n_1} \cap I_{n_2} = \emptyset$. Now, we claim that for every J_m is contained in an interval $I_n^\star$ which is our notation for the interval with the same centre as I_n but three times the length. To see this, suppose that J_m was removed from the J list immediately after the choice of I_n. Then J_m was in the J list immediately prior to the choice of I_n and we must have that $\mathrm{length(J_m)} \le \mathrm{length(I_n)}$ for otherwise J_m would be strictly longer than I_n and I_n would not have been chosen as a longest interval at that stage. Also J_m must meet I_n because it was removed immediately after the choice of I_n. It therefore follows that $J_m \subseteq I_n^\star$.

So, $K \subseteq \bigcup_{m=1}^M J_m \subseteq \bigcup_{n=1}^N I_n^\star$ and $\mu(K) \le \sum_{n=1}^N \mu(I_n^\star) = 3\sum_{n=1}^N \mu(I_n)$. It follows that (8.1) holds. ■

8.1 The hardy–Littlewood Maximal Function

We now get an estimate for the Hardy-Littlewood maximal function. Let us define for $f \in L^1(\mathbb{R}, \mathcal{L}, \mu)$

$$Mf(x) = \sup_{h>0} \frac{1}{2h} \left| \int_{x-h}^{x+h} f(t)dt \right|.$$

It is not completely obvious that Mf is measurable, so let us address that issue. We will show that for each fixed x, the map $h \mapsto \dfrac{1}{2h}\displaystyle\int_{x-h}^{x+h} f(t)dt$ is continuous on $h > 0$. A consequence is that it suffices to take the supremum over the set of all positive rational h and the supremum of a countable family of measurable functions *is* measurable. It will be enough to consider $x = 0$.

LEMMA 101 *Let $f \in L^1(\mathbb{R}, \mathcal{L}, \mu)$. Then $h \mapsto \dfrac{1}{2h}\displaystyle\int_{-h}^{h} f(t)dt$ is continuous on $h > 0$.*

Proof. Let $k_h = \frac{1}{2h}\mathbb{1}_{]-h,h[}$. Let $h > 0$ be fixed and let (h_j) be a sequence with $\frac{1}{2}h < h_j < 2h$ and $h_j \to h$ as $j \to \infty$. Then, it is easy to check that

- $k_{h_j}(t) \longrightarrow k_h(t)$ except possibly if $t = \pm h$ and in particular, almost everywhere. It then follows that $k_{h_j}(t)f(t) \longrightarrow k_h(t)f(t)$ for almost all t.
- $\sup_j |k_{h_j} - k_h| \le h^{-1}$ pointwise. It follows that $\sup_j |k_{h_j} - k_h||f| \le h^{-1}|f|$ pointwise and we observe that $h^{-1}|f|$ is integrable.

Thus, by applying the Dominated Convergence Theorem, we find that

$$\frac{1}{2h_j}\int_{-h_j}^{h_j} f(t)dt \longrightarrow \frac{1}{2h}\int_{-h}^{h} f(t)dt$$

as required. ■

THEOREM 102 *We have* $\mu(\{x; |Mf(x)| > s\}) \le 4s^{-1}\|f\|_1$.

The result says that Mf satisfies a Tchebychev type inequality for L^1. It is easy to see that we do not necessarily have $Mf \in L^1$. For example, if $f = \mathbb{1}_{[-1,1]}$, then we have

$$Mf(x) = \begin{cases} 1 & \text{if } |x| < 1, \\ \dfrac{1}{|x|+1} & \text{if } |x| \ge 1. \end{cases}$$

and Mf is not integrable even though f is.

Proof. First of all, there is no loss of generality in assuming that $f \ge 0$. Let $a \in \mathbb{N}$ and let $S = [-a,a] \cap \{x; |Mf(x)| > s\}$. Let $\mathcal{S}$ be the set of intervals $]x-h, x+h[$ such that

$$\frac{1}{2h}\int_{x-h}^{x+h} f(t)dt > s. \tag{8.2}$$

Then if $x \in S$ there is some $h > 0$ such that (8.2) holds and so $]x-h, x+h[$ is in $\mathcal{S}$. Hence the hypotheses of the Vitali Covering Lemma are satisfied. We can then find N disjoint intervals $I_n =]x_n - h_n, x_n + h_n[$ such that $\mu(S) \le 8\sum_{n=1}^{N} h_n$. But

$$2s\sum_{n=1}^{N} h_n \le \sum_{n=1}^{N}\int_{x_n-h_n}^{x_n+h_n} f(t)dt = \int \left(\sum_{n=1}^{N} \mathbb{1}_{]x_n-h_n, x_n+h_n[}\right) f d\mu \le \|f\|_1$$

Note that the disjointness of the intervals is key here. It is used to show that $\sum_{n=1}^{N} \mathbb{1}_{]x_n-h_n,x_n+h_n[} \leq \mathbb{1}$. It follows that $\mu(S) \leq 4s^{-1}\|f\|_1$. Now it suffices to let $a \longrightarrow \infty$ to obtain the desired conclusion. ■

8.2 The Martingale Maximal Function on $\mathbb{R}^*$

This section develops a similar theorem with a different and instructive proof. We work on $[0,1[$ with the Lebesgue field, we'll call it $\mathcal{F}$ and linear measure μ. A dyadic interval of length 2^{-n} is an interval $[(k-1)2^{-n}, k2^{-n}[$ for $k = 1, 2, \ldots, 2^n$ and $n = 0, 1, 2, \ldots$ The maximal function we deal with here for $f \in L^1([0,1[, \mathcal{F}, \mu)$ is

$$Mf(x) = \sup \frac{1}{\mu(I)} \int \mathbb{1}_I f d\mu$$

where the sup is taken over all dyadic intervals that contain x. There is a more succinct way of writing this maximal function

$$Mf(x) = \sup_{n=0}^{\infty} |\mathbf{E}_{\mathcal{F}_n} f|$$

where $\mathcal{F}_n$ is the σ-field (it's actually a field) generated by the dyadic intervals of length 2^{-n}. Note that for given $x \in [0,1[$ and $n \in \mathbb{Z}^+$ there is a unique dyadic interval of length 2^{-n} to which x belongs.

To get further, we need to develop the probabilistic setting. A sequence of σ-fields $\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \mathcal{F}_2 \subseteq \cdots$ is called a ***stochastic base***. The σ-field $\mathcal{F}_n$ contains those events that can be formulated at time n. The σ-field $\mathcal{F}$ is the σ-field generated by the union of all the $\mathcal{F}_n$, and contains all possible events. In our case, times n are nonnegative integers and we can imagine tossing a fair coin. So, at time 1, we toss the coin and if it ends up heads, we are in $[\frac{1}{2}, 1[$ and if it comes up tails, we are in $[0, \frac{1}{2}[$. At time 2, we toss again and we place in the upper half of the interval if we have a head and the lower half if we have a tail. The tossing is repeated indefinitely. Thus, if the result of the first 5 tosses is HTTHT, we are in the interval $[\frac{1}{2} + \frac{1}{16}, \frac{1}{2} + \frac{1}{16} + \frac{1}{32}[$.

Probabilists need to consider random times. In our case, these are mappings from the sample space $[0,1[$ to the time space $\mathbb{Z}^+ \cup \{\infty\}$ which are $\mathcal{F}$ measurable. However, there is a very special class of random times called stopping times. We can think of a gambler who is following a fixed strategy. The quintessence of being a good gambler is knowing when to quit. But if the gambler is to quit at time n, then his decision has to be based on the information that is available to him at time n. If he were able to base his decision of whether to quit or not at time n on the information available at time $n+1$, then he would be clairvoyant. So, a ***stopping time*** is a random time $\tau : [0,1[\longrightarrow \mathbb{Z}^+ \cup \{\infty\}$ with the additional property

$$\{x; \tau(x) = n\} \in \mathcal{F}_n, \quad n = 0, 1, 2, \ldots$$

This also implies

$$\{x; \tau(x) = \infty\} \in \mathcal{F}_\infty = \mathcal{F},$$

and in fact, you can make this part of the definition if you wish.

Let τ be a stopping time. We ask, what information is available at time τ. Well, an event $A \in \mathcal{F}$ can be formulated at time τ if and only if

$$A \cap \{x; \tau(x) = n\} \in \mathcal{F}_n, \quad n = 0, 1, 2, \ldots$$

The collection of all such events A defines a σ-field $\mathcal{F}_\tau$ (prove this). This idea does not make a whole lot of sense for random times (the whole space need not be in $\mathcal{F}_\tau$), but it does make sense for stopping times. Now, since $\mathcal{F}_\tau$ is a σ-field , it has an associated conditional expectation operator $\mathbf{E}_{\mathcal{F}_\tau}$. The next thing to show is that

$$\mathbf{E}_{\mathcal{F}_\tau} f = \left(\sum_{n=0}^{\infty} \mathbb{1}_{\{x;\tau(x)=n\}} \mathbf{E}_{\mathcal{F}_n} f \right) + \mathbb{1}_{\{x;\tau(x)=\infty\}} f$$

We now have enough information to start to tackle the maximal function. Let $f \in L^1([0,1[, \mathcal{F}, \mu)$ and $f \geq 0$. Fix $s > 0$. We define $\tau(x)$ to be the first time n that $\mathbf{E}_{\mathcal{F}_n} f(x) > s$. If it happens that $\mathbf{E}_{\mathcal{F}_n} f(x) \leq s$ for all $n = 0, 1, 2, \ldots$ then we have $\tau(x) = \infty$. This is a stopping time because $\tau(x) = n$ if and only if

$$\mathbf{E}_{\mathcal{F}_k} f(x) \leq s, \text{ for } k = 0, 1, \ldots, n-1$$

and

$$\mathbf{E}_{\mathcal{F}_n} f(x) > s$$

These conditions define an event in $\mathcal{F}_n$.

The key observation is that $\mathbf{E}_{\mathcal{F}_\tau} f(x) > s$ on the set $\{x; \tau(x) < \infty\} = \{x; Mf(x) > s\}$. So,

$$\mu(\{x; Mf(x) > s\}) = \mu(\{x; \mathbf{E}_{\mathcal{F}_\tau} f(x) > s\}) \leq s^{-1} \|\mathbf{E}_{\mathcal{F}_\tau} f\|_1 \leq s^{-1} \|f\|_1.$$

This is the analogue of Theorem 102. There are interesting parallels between, for example the use of *longest* intervals in the Vitali Covering lemma and the stopping time being the *first* time that $\mathbf{E}_{\mathcal{F}_n} f(x) > s$. One final caveat. This argument does not show that $Mf \in L^1$ (a false statement in general) because the stopping time τ depends on s.

8.3 Fundamental Theorem of Calculus

THEOREM 103 *Let $f \in L^1(\mathbb{R})$ and define $F(x) = \int_0^x f(t)dt$. Then $F'(x)$ exists and equals $f(x)$ for almost all $x \in \mathbb{R}$.*

Proof. Let $\epsilon > 0$ and write $f = g + h$ where $g \in C_c(\mathbb{R})$ and $h \in L^1$ with $\|h\|_1 < \epsilon$. Let us also define $G(x) = \int_0^x g(t)dt$ and $H(x) = \int_0^x h(t)dt$. Then $F(x) = G(x) + H(x)$. Now consider

$$\begin{aligned}
&\limsup_{t\to 0}\left|\frac{F(x+t)-F(x)}{t} - f(x)\right| \\
&\le \limsup_{t\to 0}\left|\frac{G(x+t)-G(x)}{t} - g(x)\right| + \limsup_{t\to 0}\left|\frac{H(x+t)-H(x)}{t} - h(x)\right|
\end{aligned}$$

Now, the first $\limsup$ on the right is zero, by the Fundamental Theorem of Calculus, because g is continuous. So,

$$\begin{aligned}
\limsup_{t\to 0}\left|\frac{F(x+t)-F(x)}{t} - f(x)\right| &\le \limsup_{t\to 0}\left|\frac{H(x+t)-H(x)}{t} - h(x)\right| \\
&\le |h(x)| + \sup_{t\neq 0}\left|\frac{H(x+t)-H(x)}{t}\right| \\
&\le |h(x)| + \sup_{t\neq 0} t^{-1}\int_x^{x+t}|h(s)|ds \\
&\le |h(x)| + \sup_{t>0} t^{-1}\int_{x-t}^{x+t}|h(s)|ds \\
&= |h(x)| + 2M|h|(x)
\end{aligned}$$

Let $\delta > 0$ and consider the set

$$A_\delta = \{x; \limsup_{t\to 0+}\left|\frac{F(x+t)-F(x)}{t} - f(x)\right| > \delta\}.$$

Now, if $x \in A_\delta$, then either $|h(x)| > \frac{1}{3}\delta$ or $M|h|(x) > \frac{1}{3}\delta$. The first possibility occurs on a set of measure at most $3\epsilon\delta^{-1}$ (by the Tchebychev Inequality) and the second on a set of measure at most $12\epsilon\delta^{-1}$ by Theorem 102. So, the measure of A_δ is bounded by $15\epsilon\delta^{-1}$. But A_δ does not depend on ϵ, so letting $\epsilon \longrightarrow 0+$, we find that A_δ is a null set. Finally, taking a sequence of positive δs converging to zero, we find that $\{x; \limsup_{t\to 0+} \left|\frac{F(x+t)-F(x)}{t} - f(x)\right| > 0\}$ is also a null set. The result follows. ∎

8.4 Jacobian Determinants and Change of Variables*

We start with the following simple Lemma.

LEMMA 104 *Let P be a parallelepiped contained in a ball of radius R in $\mathbb{R}^d$. Let $P_r = P + B(\mathbf{0}, r)$, that is the set of points that lie within distance r of P. Then, for Lebesgue measure μ,*

$$\mu(P_r) \le \mu(P) + C_d r(R+r)^{d-1},$$

where C_d is a constant that depends only on the ambient dimension d.

Proof. We'll give the proof in the case $d = 2$ and the reader will find that it generalizes easily to higher dimensions. Let $x \in P_r \setminus \mathrm{cl}(P)$. Then the distance from x to $\mathrm{cl}(P)$ is strictly positive and there is a nearest point y of $\mathrm{cl}(P)$ to x. Clearly, points on the line segment from x to y with the exception of y itself are not in $\mathrm{cl}(P)$, So y is a boundary point of $\mathrm{cl}(P)$ and hence also a boundary point of P. Thus, x lies within distance r of ∂P. Now, δP has 4 faces (2^d faces in d dimensions). Each face is contained in an interval of length at most $2R$ (in general, a ball of radius R, cut by a hyperplane). So, the set of points within distance r of ∂P has measure bounded by $16Rr + 4\pi r^2$ (in general, $C_d r(R+r)^{d-1}$ for some suitable C_d). ∎

Another interesting lemma is the following.

LEMMA 105 *Let U be an open subset of $\mathbb{R}^{d-1}$ and $f : U \longrightarrow \mathbb{R}^d$ be a Lipschitz mapping. Then $f(U)$ is a Lebesgue null set.*

Proof. We write U as a countable union of $(d-1)$-dimensional cubes in $\mathbb{R}^{d-1}$. Fix one of these $(d-1)$-dimensional cubes Q of side s. It is enough to show that $f(Q)$ is Lebesgue null. Now, let $N \in \mathbb{N}$ and split up this cube into N^{d-1} equal $(d-1)$-dimensional cubes of side sN^{-1}. Since f is a Lipschitz mapping, each of the smaller cubes maps into a d-dimensional ball of diameter MsN^{-1} where M is the Lipschitz constant of f. So, $f(Q)$ is contained in a set of measure $C_d N^{d-1}(MsN^{-1})^d$. Now let $N \longrightarrow \infty$. ∎

COROLLARY 106 *It is impossible to find a Lipschitz mapping $f : [0,1] \longrightarrow [0,1] \times [0,1]$ which is onto.*

PROPOSITION 107 *Let U be a bounded open subset of $\mathbb{R}^d$. Let $f : U \longrightarrow \mathbb{R}^d$ be a C^1 injection with the property that df_x is nonsingular for every $x \in U$. Let K be a compact subset of U (typically a large chunk of U). Let $\epsilon > 0$. Then there exists $\delta > 0$ such that*

$$\mu(f(Q)) \le (|\det(df_\xi)| + \epsilon)\mu(Q) \tag{8.3}$$

for any cube $Q \subseteq K$, centred at ξ of side less than δ.

Proof. First we assume without loss of generality that $\epsilon \le 1$. Next we use the compactness of K and the continuity of $x \mapsto df_x$ to establish the boundedness and uniform continuity of df.

- There exists a constant M such that $\|df_x\|_{\text{op}} \le M$ for all $x \in K$.
- For every $\epsilon > 0$ there is a constant $\delta > 0$ such that $\|df_x - df_\xi\|_{\text{op}} < \epsilon$ whenever $x, \xi \in K$ and $\|x - \xi\| \le d^{\frac{1}{2}}\delta$.

Now let Q be a cube contained in K, let ξ be the centre of Q and $x \in Q$. Then

$$\|f(x) - f(\xi)\| \le M\|x - \xi\|$$

since the line segment joining x and ξ lies in Q and hence in K. This uses a standard estimate, see the MATH 354 notes for example.

We use the symbol C_d to denote a constant depending only on the ambient dimension d, but the constant may change with different occurrences of the symbol. With this convention, we can write that $f(Q)$ is contained in a ball of radius $C_d M \operatorname{side}(Q)$.

The second statement above is used to obtain a uniform differentiability condition. Let $x(t) = (1-t)\xi + tx$ and $\varphi(t) = f(x(t))$. Then, $\varphi'(t) = df_{x(t)}(x - \xi)$, and it follows that $\|\varphi'(t) - df_\xi(x - \xi)\| \le \epsilon\|x - \xi\|$ for $\|x - \xi\| \le d^{\frac{1}{2}}\delta$. We now have

$$f(x) - f(\xi) = \varphi(1) - \varphi(0)$$

$$
\begin{aligned}
&= \int_0^1 \varphi'(t)dt \\
&= \int_0^1 df_{x(t)}(x-\xi)dt
\end{aligned}
$$

so that

$$\|f(x) - f(\xi) - df_\xi(x-\xi)\| \le \epsilon\|x-\xi\|$$

for $\|x-\xi\| \le d^{\frac{1}{2}}\delta$. Now, let $g(x) = f(\xi) + df_\xi(x-\xi)$ an affine (constant plus linear) mapping. Then for x in the cube Q with centre ξ and side $\le \delta$ we have

$$f(Q) \subseteq g(Q) + B(\mathbf{0}, d^{\frac{1}{2}}\epsilon\,\mathrm{side}(Q)).$$

But $g(Q)$ is a parallelepiped and it follows from Lemma 104 that

$$
\begin{aligned}
\mu(f(Q)) &\le \mu(g(Q)) + C_d(C_d M\mathrm{side}(Q) + d^{\frac{1}{2}}\epsilon\,\mathrm{side}(Q))^{d-1} d^{\frac{1}{2}}\epsilon\,\mathrm{side}(Q) \\
&\le \mu(g(Q)) + C_{d,M}\epsilon\,\mathrm{side}(Q)^d = \Big(|\det(df_\xi)| + C_{d,M}\epsilon\Big)\mu(Q)
\end{aligned}
$$

So, after rescaling ϵ, the result is proved. ■

PROPOSITION 108 *Let U be a bounded open subset of $\mathbb{R}^d$. Let $f : U \longrightarrow \mathbb{R}^d$ be a C^1 injection with the property that df_x is nonsingular for every $x \in U$. Let g be a nonnegative continuous function on $f(U)$. Then*

$$\int_{f(U)} g(x)d\mu(x) \le \int_U g\circ f(x)|\det(df_x)|d\mu(x)$$

Proof. Let W be an open subset of U with $\mathrm{cl}(W) \subseteq U$. Then, it will be enough to show that

$$\int_{f(W)} g(x)d\mu(x) \le \int_W g\circ f(x)|\det(df_x)|d\mu(x).$$

Then, taking a sequence of suitable W increasing to U the desired result will follow from the Monotone Convergence Theorem. (There's an exercise here. If

U is a bounded open subset of $\mathbb{R}^d$, show that there exist a sequence $(W_k)_{k=1}^{\infty}$ of open subsets of $\mathbb{R}^d$ such that $\mathrm{cl}(W_k) \subseteq U$ and $\bigcup_{k=1}^{\infty} W_k = U$.)

So, let the K of the previous proposition be $\mathrm{cl}(W)$. Given $\epsilon > 0$, choose $\delta > 0$ such that (8.3) holds and the oscillation of both $g \circ f$ and $x \mapsto g(f(x))|\det(df_x)|$ over a cube Q of side $\leq \delta$ is $< \epsilon$. We split up W into countably many disjoint cubes Q_j with side less than δ[1]. The boundaries of these cubes are irrelevant, (i.e. $\mu(f(\partial Q_j)) = 0$), this is a consequence of the fact that f is a Lipschitz mapping and Lemma 105. Now we can approximate each of the integrals with sums as follows.

$$\left| \int_W g \circ f(x) |\det(df_x)| d\mu(x) - \sum_{j=1}^{\infty} g \circ f(\xi_j) |\det(df_{\xi_j})| \mu(Q_j) \right| \leq \epsilon \mu(W)$$

and

$$\left| \int_{f(W)} g(x) d\mu(x) - \sum_{j=1}^{\infty} g(f(\xi_j)) \mu(f(Q_j)) \right| \leq \epsilon \mu(f(W))$$

Together with the estimate from (8.3), we get

$$\int_{f(W)} g(x) d\mu(x) \leq \int_W g \circ f(x) |\det(df_x)| d\mu(x) + E$$

where the error term E satisfies

$$E \leq 2\epsilon\mu(W) + \epsilon\mu(f(W))$$

Letting $\epsilon \longrightarrow 0$, we have the desired result because both W and $f(W)$ are bounded sets and have finite measure. ∎

THEOREM 109 *Let U be a bounded open subset of $\mathbb{R}^d$. Let $f : U \longrightarrow \mathbb{R}^d$ be a bounded C^1 injection with the property that df_x is nonsingular for every $x \in U$. Let g be a continuous function on $f(U)$. Then*

$$\int_{f(U)} g(x) d\mu(x) = \int_U g \circ f(x) |\det(df_x)| d\mu(x)$$

[1] This is another exercise. Show that every bounded open subset of Euclidean space is a union of closed dyadic cubes with pairwise disjoint interiors. There's a canonical way of doing this — no cube is included if its dyadic double (i.e. the unique dyadic cube with twice the side containing the given cube) is contained in the given open set. For the purposes of this result, the larger dyadic cubes (i.e. those with side $> \delta$ need to be further subdivided, but this is easy.!

Proof. We start by making the additional assumption that g is nonnegative. We remark that $f(U)$ is a bounded open set and that $f^{-1} : f(U) \longrightarrow U$ satisfies the same conditions as f. Therefore, applying Proposition 108, to f^{-1} and $x \mapsto g(f(x))|\det(df_x)|$, we get

$$\begin{aligned}\int_U g \circ f(x)|\det(df_x)|d\mu(x) &\leq \int_{f(U)} g(x)|\det(df_{f^{-1}(x)}||\det(df_x^{-1})|d\mu(x) \\ &= \int_{f(U)} g(x)d\mu(x).\end{aligned}$$

Because $df_{f^{-1}(x)} \circ df_x^{-1}$ is the identity mapping by the Chain Rule. Combining this with Proposition 108, we have the desired result in case g nonnegative. To get the general result, we write $g = g_+ - g_-$ where $g_\pm = \max(0, \pm g)$, apply the preceding result to $g_\pm$ and use the linearity of the integral. ■

9

Fourier Transforms*

In this chapter, we develop just the basics of Fourier transforms on the line, sometimes called Fourier integrals.

DEFINITION *For a function $f \in L^1(\mathbb{R})$ we define the* ***Fourier transform*** $\hat{f}$ *to be the function on another copy of the line which will be referred to as $\widehat{\mathbb{R}}$ given by*

$$\hat{f}(u) = \int_{\mathbb{R}} f(x)e^{-iux}dx \text{ for } u \in \widehat{\mathbb{R}}.$$

The integral is well defined since it converges absolutely.

Here we have denoted dx the Lebesgue measure on the line and the L^1 space is taken with respect to the Lebesgue or the Borel σ-field. Similarly we have L^p spaces defined on $\widehat{\mathbb{R}}$, but for these spaces we take the measure $\frac{1}{2\pi}du$. This is the dual measure. The definition presented here is just one of several possible normalizations — the one favoured by the French and the Russians. The Americans define the Fourier transform according to a different normalization.

Everything we do in this chapter depends upon some standard integrals typically

$$\frac{1}{\sqrt{2\pi}} \int e^{-\frac{1}{2}x^2} e^{-iux} dx = e^{-\frac{1}{2}u^2} \tag{9.1}$$

which can be established by means of contour integration in complex function theory. If you know how to do this, then all well and good. If not, you will have to take it on trust. More generally, we define for $t > 0$

$$\gamma_t(x) = \frac{1}{t\sqrt{2\pi}} e^{-\frac{1}{2}t^{-2}x^2},$$

in fact γ_t is the ***Gauss kernel*** . Two further identities are obtained from (9.1) by change of variables, namely

$$\widehat{\gamma}_t(u) = \frac{1}{t\sqrt{2\pi}} \int e^{-\frac{1}{2}t^{-2}x^2} e^{-iux} dx = e^{-\frac{1}{2}t^2u^2}$$

and

$$\int_{\widehat{\mathbb{R}}} \widehat{\gamma}_t(u) e^{iux} \frac{1}{2\pi} du = \int_{\widehat{\mathbb{R}}} e^{-\frac{1}{2}t^2u^2} e^{iux} \frac{1}{2\pi} du = \frac{1}{t\sqrt{2\pi}} e^{-\frac{1}{2}t^{-2}x^2} = \gamma_t(x).$$

These formulæ represent the Fourier transform and the inverse Fourier transform for the Gauss kernel.

LEMMA 110 *The Gauss kernel γ_t is a summability kernel on $\mathbb{R}$ as $t \to 0+$.*

DEFINITION *For f and g suitable measurable functions on $\mathbb{R}$, then the convolution product $f \star g$ is defined by*

$$f \star g(x) = \int_{\mathbb{R}} f(x-y) g(y) dy.$$

PROPOSITION 111 *The following scenarios are the most common.*

(i) $f, g \in L^1(\mathbb{R})$*. Then* $f \star g \in L^1(\mathbb{R})$ *and* $\|f \star g\|_1 \leq \|f\|_1 \|g\|_1$.

(ii) $f \in L^p(\mathbb{R})$, $g \in L^1(\mathbb{R})$ *with* $1 \leq p < \infty$*. Then* $f \star g \in L^p(\mathbb{R})$ *and* $\|f \star g\|_p \leq \|f\|_p \|g\|_1$.

(iii) $f \in L^p(\mathbb{R})$, $g \in L^{p'}(\mathbb{R})$ *with* $1 \leq p \leq \infty$*. Then* $f \star g \in L^\infty(\mathbb{R})$ *and* $\|f \star g\|_\infty \leq \|f\|_p \|g\|_{p'}$.

Sketch Proof.

In (i) we verify that $\iint |f(x-y)||g(y)| dy dx < \infty$, which shows that the integral defining $f \star g(x)$ is absolutely convergent for almost all x and that the resulting function is in L^1. This is exactly as in the Fourier series section.

For (ii) we let X be a measurable set of finite measure in $\mathbb{R}$, then

$$\int_X \left\{ \int |f(x-y)||g(y)|dy \right\} dx = \int \left\{ \int_X |f(x-y)|dx \right\} |g(y)|dy$$
$$\leq \text{meas}(X)^{\frac{1}{p'}} \|f\|_p \|g\|_1 < \infty$$

and again we see that the integral converges absolutely for almost all x. Now let h be a nonnegative function in $L^{p'}$, then again by Tonelli's Theorem

$$\iint h(x)|f(x-y)||g(y)|dydx \leq \|h\|_{p'} \|f\|_p \|g\|_1$$

and Theorem 57 now gives that $x \mapsto \iint |f(x-y)||g(y)|dy$ is an L^p function with norm bounded by $\|f\|_p \|g\|_1$. The same is therefore also true of $f \star g$.

There is another completely different approach to (ii). Let us consider the map $F : \mathbb{R} \longrightarrow L^p(\mathbb{R})$ defined by $(F(y))(x) = f(x-y)$, i.e. $F(y) = T_y(f)$, then by Corollary 61, F is continuous. If g is a continuous function of compact support, then we can consider the Riemann integral

$$\int F(y)g(y)dy$$

the integrand being a continuous function on $\mathbb{R}$ taking values in the complete space $L^p(\mathbb{R})$ and vanishing outside a compact set. This leads to

$$\left\| \int F(y)g(y)dy \right\|_p \leq \sup_y \|F(y)\|_p \|g\|_1 = \|f\|_p \|g\|_1.$$

Extending this function to $g \in L^1(\mathbb{R})$ using uniform continuity and the density of $C_c(\mathbb{R})$ in $L^1(\mathbb{R})$ we get a new definition of $f \star g$ satisfying the desired inequality. As an exercise, the reader may show that this definition agrees with the previous one.

Finally, for (iii), its easy to see that $f \star g$ is bounded in absolute value by $\|f\|_p \|g\|_{p'}$. In fact, $f \star g$ is continuous. To see this, we observe first that $T_t(f \star g) = T_t(f) \star g = f \star T_t(g)$ and the result follows since if $1 \leq p < \infty$, then $t \mapsto T_t(f)$ is continuous with values in L^p and if $1 < p \leq \infty$, $t \mapsto T_t(g)$ is continuous with values in $L^{p'}$. ∎

We remark also that a simple change of variables shows that $f \star g = g \star f$ when both sides are defined.

9.1 Fourier Transforms of L^1 functions

It should be clear from the definition of the Fourier transform that $\hat{f}$ is a bounded function for $f \in L^1$. If in addition f has compact support, then we can differentiate under the integral sign as many times as we wish, so that $\hat{f}$ is infinitely differentiable. Since L^1 functions of compact support are dense in the space of all L^1 functions and since the Fourier transform is bounded from $L^1(\mathbb{R})$ to $\ell^\infty(\widehat{\mathbb{R}}_d)$ (i.e. the space of *all* bounded functions on $\widehat{\mathbb{R}}$), we see that the Fourier transforms of L^1 functions are necessarily continuous. In fact, slightly more is true, namely the Riemann–Lebesgue Lemma in the context of the line. For this we will first need the following lemma.

LEMMA 112 *Let $f, g \in L^1(\mathbb{R})$, then $\widehat{f \star g}(u) = \hat{f}(u)\hat{g}(u)$ for all $u \in \widehat{\mathbb{R}}$.*

Proof. We have

$$\begin{aligned}\widehat{f \star g}(u) &= \int\int f(x-y)g(y)e^{-iux}dydx \\ &= \int\int f(x-y)g(y)e^{-iux}dxdy\end{aligned}$$

by Fubini's Theorem and since $f, g \in L^1$

$$= \int\int f(z)g(y)e^{-iu(y+z)}dzdy$$

by making a change of variables in the inner integral

$$\begin{aligned}&= \int\int f(z)g(y)e^{-iuz}e^{-iuy}dzdy \\ &= \hat{f}(u)\hat{g}(u)\end{aligned}$$

as required. ∎

LEMMA 113 (RIEMANN–LEBESGUE LEMMA) *If $f \in L^1(\mathbb{R})$, then $\hat{f} \in C_0(\widehat{\mathbb{R}})$.*

Proof. Since γ_t is a summability kernel, $\gamma_t \star f \longrightarrow f$ in L^1 norm as $t \to 0+$. But

$$\widehat{\gamma_t \star f}(u) = \widehat{\gamma_t}(u)\hat{f}(u) = e^{-\frac{1}{2}t^2u^2}\hat{f}(u).$$

So $\widehat{\gamma_t \star f} \in C_0(\widehat{\mathbb{R}})$ since it is the product of a bounded continuous function with a continuous function tending to zero at infinity. But $\widehat{\gamma_t \star f} \longrightarrow \hat{f}$ uniformly, and it follows that $\hat{f} \in C_0(\widehat{\mathbb{R}})$. ■

9.2 Fourier Transforms of L^2 functions

We start with the following key result.

PROPOSITION 114 *Let $f \in L^1(\mathbb{R}) \cap L^2(\mathbb{R})$, then $\int |f(x)|^2 dx = \int |\hat{f}(u)|^2 \frac{du}{2\pi}$.*

Proof. Since $f \in L^2$ and since γ_t is a summability kernel, $\gamma_t \star f \longrightarrow f$ in L^2 norm as $t \to 0+$. Therefore

$$\int\int \overline{f(x)} f(x-y)\gamma_t(y)dydx \longrightarrow \int \overline{f(x)} f(x)dx = \|f\|_2^2.$$

On the other hand since $f \in L^1$,

$$\begin{aligned}
\int\int \overline{f(x)} f(x-y)\gamma_t(y)dydx &= \int\int \overline{f(x)} f(y)\gamma_t(x-y)dxdy \\
&= \int\int\int \overline{f(x)} f(y) e^{-\frac{1}{2}t^2u^2} e^{iu(x-y)} \frac{du}{2\pi} dxdy \\
&= \int\int\int \overline{f(x)e^{-iux}} f(y) e^{-iuy} e^{-\frac{1}{2}t^2u^2} dxdy \frac{du}{2\pi} \\
&= \int \overline{\hat{f}(u)} \hat{f}(u) e^{-\frac{1}{2}t^2u^2} \frac{du}{2\pi} \\
&\longrightarrow \int |\hat{f}(u)|^2 \frac{du}{2\pi}
\end{aligned}$$

by monotone convergence. ■

COROLLARY 115 *Let $f, g \in L^1(\mathbb{R}) \cap L^2(\mathbb{R})$, then*

$$\int \overline{f(x)} g(x)dx = \int \overline{\hat{f}(u)} \hat{g}(u) \frac{du}{2\pi}.$$

Proof. The proof follows the same line as in the previous proposition. The only difference is that the final step is justified using dominated convergence, since now it is known that $\overline{\hat{f}}\hat{g}$ is in $L^1(\widehat{\mathbb{R}})$ since it is the product of two L^2 functions. ■

COROLLARY 116 (FOURIER INTEGRAL UNIQUENESS THEOREM) *Let f be in $L^1(\mathbb{R})$ and suppose that $\hat{f}(u) = 0$ for all $u \in \widehat{\mathbb{R}}$. Then f is the zero element of L^1.*

Proof. We have $\widehat{f \star \gamma_t}(u) = \hat{f}(u)\widehat{\gamma_t}(u) = 0$ for all $u \in \widehat{\mathbb{R}}$. So, since $f \star \gamma_t \in L^1 \cap L^2$ and by Proposition 114 applied to $f \star \gamma_t$, we have that $\|f \star \gamma_t\|_2 = 0$. So, $f \star \gamma_t$ is zero almost everywhere. But $f \star \gamma_t \longrightarrow f$ in L^1 norm as $t \to 0+$. Therefore, f is zero almost everywhere. ■

At this point, we need to define the inverse Fourier transform. For a function $h \in L^1(\widehat{\mathbb{R}}, \frac{1}{2\pi}du)$, we define for $x \in \mathbb{R}$

$$\check{h}(x) = \int h(u)e^{iux}\frac{du}{2\pi}.$$

Obviously, this definition is very similar to the definition of the Fourier transform (in that $\check{h}(x) = \frac{1}{2\pi}\hat{h}(-x)$ and so, the uniqueness result above will also apply to the inverse Fourier transform, namely

$$h \in L^1(\widehat{\mathbb{R}}),\ \check{h} \equiv 0 \implies h \equiv 0.$$

With this in mind, we can now prove the following famous theorem.

THEOREM 117 (PLANCHEREL'S THEOREM) *The Fourier transform extends by continuity to an isometry of $L^2(\mathbb{R})$ onto $L^2(\widehat{\mathbb{R}})$.*

Proof. Most of the work is done. We extend the Fourier transform from $L^1 \cap L^2$ to L^2 by continuity. This defines a linear isometry from $L^2(\mathbb{R})$ to $L^2(\widehat{\mathbb{R}})$. It remains only to show that the isometry is surjective.

The direct image of $L^2(\mathbb{R})$ in $L^2(\widehat{\mathbb{R}})$ is clearly complete (it is the isometric image of a complete space). It follows that the image is closed in $L^2(\widehat{\mathbb{R}})$. If it is not the whole of $L^2(\widehat{\mathbb{R}})$, then we may find a nonzero element g of $L^2(\widehat{\mathbb{R}})$ orthogonal to the image. Let $f \in L^1(\mathbb{R})$ be arbitrary. We have

$$0 = \int \overline{g(u)}\widehat{f \star \gamma_t}(u)\frac{du}{2\pi}$$

since $f \star \gamma_t \in L^1 \cap L^2$,

$$= \int \overline{g(u)}\widehat{\gamma_t}(u)\hat{f}(u)\frac{du}{2\pi}$$

$$= \int \int \overline{g(u)} \widehat{\gamma}_t(u) e^{-iux} f(x) dx \frac{du}{2\pi}$$
$$= \int \int \overline{g(u)} \widehat{\gamma}_t(u) e^{-iux} f(x) \frac{du}{2\pi} dx$$

by Fubini's Theorem and since $\overline{g}\widehat{\gamma}_t, f \in L^1$,

$$= \int (\overline{g}\widehat{\gamma}_t)\check{\ }(x) f(x) dx$$

Since this is true for all $f \in L^1(\mathbb{R})$, we have $(\overline{g}\widehat{\gamma}_t)\check{\ }(x) = 0$ for almost all x. Then by the uniqueness of the inverse transform $\overline{g}\widehat{\gamma}_t \equiv 0$ and consequently $g \equiv 0$. This contradiction completes the proof. ■

It is important to note that the Plancherel Theorem allows us to make an interpretation of the integral

$$\int f(x) e^{-iux} dx$$

for $f \in L^2$ even when the integral does not converge absolutely. Taken at face value, this is something very mysterious.

9.3 Fourier Inversion

In this section we aim to prove two results about inversion.

THEOREM 118 (THE FOURIER INVERSION THEOREM) *Let $f \in L^1(\mathbb{R})$ and suppose that $\hat{f} \in L^1(\widehat{\mathbb{R}})$. Then $(\hat{f})\check{\ } = f$ almost everywhere.*

Proof. Suppose that $g \in L^1(\mathbb{R}) \cap L^\infty(\mathbb{R})$. Then certainly $g \in L^2(\mathbb{R})$. Also $f \star \gamma_t \in L^2$. Therefore

$$\int \overline{g(x)} (f \star \gamma_t)(x) dx = \int \overline{\hat{g}(u)} \widehat{\gamma}_t(u) \hat{f}(u) \frac{du}{2\pi}$$
$$= \int \overline{g(x)} e^{iux} \widehat{\gamma}_t(u) \hat{f}(u) dx \frac{du}{2\pi}$$

and letting $t \to 0+$ using dominated convergence since $g \in L^1(\mathbb{R})$ and $\hat{f} \in L^1(\widehat{\mathbb{R}})$

$$\int \overline{g(x)} f(x) dx = \int \overline{g(x)} e^{iux} \hat{f}(u) dx \frac{du}{2\pi}$$

$$= \int \overline{g(x)} e^{iux} \hat{f}(u) \frac{du}{2\pi} dx$$

applying Fubini, again since $g \in L^1(\mathbb{R})$ and $\hat{f} \in L^1(\widehat{\mathbb{R}})$

$$= \int \overline{g(x)} (\hat{f})\check{}\,(x) dx$$

Thus

$$\int \Big(f(x) - (\hat{f})\check{}\,(x) \Big) \overline{g(x)} dx = 0$$

for all $g \in L^1(\mathbb{R}) \cap L^\infty(\mathbb{R})$. It's now easy to see that $f(x) = (\hat{f})\check{}\,(x)$ for almost all x. ∎

The final result on inversion relates to the fact that we may extend the Fourier inversion operator to an isometry from $L^2(\widehat{\mathbb{R}})$ to $L^2(\mathbb{R})$ as in the Plancherel Theorem.

PROPOSITION 119 *The resulting operator is the inverse of the extension of the Fourier transform to $L^2(\mathbb{R})$.*

Proof. Let $A : L^2(\mathbb{R}) \longrightarrow L^2(\widehat{\mathbb{R}})$ be the extension of the Fourier transform of $L^1 \cap L^2$ to L^2 and let $B : L^2(\widehat{\mathbb{R}}) \longrightarrow L^2(\mathbb{R})$ be the similarly defined extension of the inverse Fourier transform. Both of these operators are surjective isometries and it will suffice to show that $B \circ A(f) = f$ on a dense subset of $L^2(\mathbb{R})$. Let $f \in L^1(\mathbb{R}) \cap L^2(\mathbb{R})$. Then $f \star \gamma_t \in L^1(\mathbb{R}) \cap L^2(\mathbb{R})$. Also $\widehat{f \star \gamma_t} = \hat{f}\widehat{\gamma_t} \in L^1(\widehat{\mathbb{R}})$ since both $\hat{f}$ and $\widehat{\gamma_t}$ are in $L^2(\widehat{\mathbb{R}})$. By the Fourier Inversion Theorem

$$(\hat{f}\widehat{\gamma_t})\check{}\; = f \star \gamma_t.$$

As $t \to 0+$, $\hat{f}\widehat{\gamma_t} \longrightarrow \hat{f}$ in $L^2(\widehat{\mathbb{R}})$ since $\hat{f} \in L^2(\widehat{\mathbb{R}})$ and by dominated convergence. Since γ_t is a summability kernel, $f \star \gamma_t \to f$ in $L^2(\mathbb{R})$. Therefore

$$B(\hat{f}) = f$$

for all $f \in L^1(\mathbb{R}) \cap L^2(\mathbb{R})$. But $L^1(\mathbb{R}) \cap L^2(\mathbb{R})$ is dense in $L^2(\mathbb{R})$ and the result is proved. ∎

9.4 Defining the Fourier Transform on $L^1 + L^2$

Since we now know how to define the Fourier transform on $L^1(\mathbb{R})$ and $L^2(\mathbb{R})$ separately, we can define it almost everywhere for the sum of an L^1 function and an L^2 function in the obvious way.

DEFINITION *If $f = f_1 + f_2$ with $f_1 \in L^1(\mathbb{R})$ and $f_2 \in L^2(\mathbb{R})$, then we define*

$$\hat{f}(u) = \widehat{f_1}(u) + A(f_2)(u)$$

and we see that $\hat{f} \in L^\infty + L^2$. The definition is good since if $f = f_1 + f_2 = g_1 + g_2$ with $f_1, g_1 \in L^1(\mathbb{R})$ and $f_2, g_2 \in L^2(\mathbb{R})$ then $f_1 - g_1 = g_2 - f_2 \in L^1 \cap L^2$. But $A(h) = \hat{h}$ for functions in $L^1 \cap L^2$ and therefore

$$(f_1 - g_1)\hat{} = A(g_2 - f_2)$$

and by linearity, this now gives $\widehat{f_1} + A(f_2) = \widehat{g_1} + A(g_2)$.

EXAMPLE Let $f(x) = |x|^{-1+z}$ with $0 < \Re z < \frac{1}{2}$. We have

$$\int_{|x|\le 1} |f(x)|dx \le \int_{|x|\le 1} |x|^{-1+\Re z} dx < \infty$$

since $\Re z > 0$ and

$$\int_{|x|>1} |f(x)|^2 dx \le \int_{|x|>1} |x|^{-2+2\Re z} dx < \infty$$

since $\Re z < \frac{1}{2}$. So, $f \in L^1 + L^2$ and the Fourier transform $\hat{f}$ is defined almost everywhere. A change of variable shows that $\hat{f}(u) = c_z|u|^{-z}$ where c_z is some constant depending only on z. It is not difficult to show that

$$\int_{-\infty}^{\infty} \gamma_1(x) f(x) dx = \int_{-\infty}^{\infty} e^{-\frac{1}{2}u^2} \hat{f}(u) \frac{du}{2\pi}$$

both integrals leading to Gamma functions and $c_z = \sqrt{\pi}\dfrac{\Gamma(\frac{1}{2}z)}{\Gamma(\frac{1}{2}(1-z))}2^z$. □

Index