PMATH 450 Course Notes

Lebesgue Integration and Fourier Analysis

Michael Brannan • Spring 2022 • University of Waterloo

Table of Contents

1 Motivation

This course is a continuation of PMATH 351. It can be thought of as a gateway course to many areas of modern analysis, and has many applications such as partial differential equations or even representation theory of groups.

Even though this course is called "Lebesgue Integration and Fourier Analysis", we will focus more on the latter, since there is a lot of overlap with PMATH 451 in terms of measure theory. First, we will begin by giving a hand-wavy derivation of the heat equation. We then try to solve the corresponding PDE, which will give us some motivation for studying Fourier analysis.

1.1 Deriving the Heat Equation

Take a "nice" region $D \subseteq \mathbb{R}^3$ with volume, such as a sphere, cylinder, or cube. Consider a solid body with shape D. At time $t = 0$, the body is heated to an initial temperature

$$u(x, y, z, t)|_{t=0} = u(x, y, z, 0).$$

This is our initial condition. Moreover, for all $t > 0$, the temperature on the boundary ∂D is specified; that is, we know the values of $u(x, y, z, t)$ for all $(x, y, z) \in \partial D$. These are the boundary conditions. Our goal is to find $u(x, y, z, t)$ for all $t > 0$ and $(x, y, z) \in D$.

To begin, we will derive (using physics) the PDE governing u. The solid body with shape given by D is assumed to have constant density $\rho > 0$, and there is a specific heat constant $c > 0$. Then the heat content of D is given by

$$H(t) = \iiint_D c\rho u(x, y, z, t)\,\mathrm{d}V.$$

Behaving as a physicist would, we toss the derivative into the integral without question to obtain

$$H'(t) = \iiint_D c\rho u_t(x, y, z, t)\,\mathrm{d}V. \tag{1.1}$$

Now, Fourier's Law states that heat flows from hotter to colder regions at a rate proportional to the temperature gradient

$$\nabla u(x, y, z) = (u_x, u_y, u_z).$$

With our nice region D, heat only flows in and out through the surface ∂D. Then Fourier states that there exists $\kappa > 0$ such that $H'(t)$ is equal to κ multiplied by the flux of ∇u through ∂D. That is, we have

$$H'(t) = \iint_{\partial D} \kappa(\nabla u) \cdot \mathrm{d}\vec{S},$$

where $\mathrm{d}\vec{S}$ is the surface differential $\vec{n} \cdot \mathrm{d}S$. Recall that the divergence of a vector field $\vec{F} = (F_1, F_2, F_3)$ is defined by $\operatorname{div}(\vec{F}) := (F_1)_x + (F_2)_y + (F_3)_z$, and Gauss' Divergence Theorem states that

$$\iint_{\partial D} \vec{F} \cdot \mathrm{d}\vec{S} = \iiint_D \operatorname{div}(\vec{F})\,\mathrm{d}V.$$

In our case, we have $\vec{F} = \kappa \nabla u = (\kappa u_x, \kappa u_y, \kappa u_z)$, and hence

$$\operatorname{div}(\kappa \nabla u) = \kappa(u_{xx} + u_{yy} + u_{zz}) = \kappa \Delta u,$$

where $\Delta u = u_{xx} + u_{yy} + u_{zz}$ is the Laplacian of u. From our above equation, this yields

$$H'(t) = \iiint_D \kappa \Delta u\,\mathrm{d}V. \tag{1.2}$$

Combining (1.1) and (1.2) and doing some rearranging, we end up with

$$\iiint_D (c\rho u_t - \kappa \Delta u)\, \mathrm{d}V = 0.$$

This holds for all "nice" regions, and implies that

$$c\rho u_t - \kappa \Delta u = 0.$$

Setting $K = \kappa/(c\rho) > 0$, we obtain the heat equation

$$u_t = K\Delta u.$$

We now want to solve this with our given initial and boundary conditions. As we would expect, this is very difficult! This is a PDE, and solving an ODE is already a tall order.

1.2 Solving the Heat Equation

For simplicity, we will instead consider the 1-dimensional heat equation. Let us take a thin rod over the interval $[-\pi, \pi]$. Suppose that this rod is laterally insulated so that heat only flows in the x-direction. In this case, the heat equation is given by

$$u_t = K u_{xx}.$$

Our initial condition is $u(x, 0) = f(x)$ where f is piecewise continuous or even just Riemann integrable, if we want to be more fancy. As for the boundary conditions, this really depends on the physical scenario. We give some examples of them below.

- We may assert that the temperature is 0 on the endpoints, so $u(-\pi, t) = u(\pi, t) = 0$ for all $t \geq 0$. These are called Dirichlet boundary conditions.
- We can also assume that the endpoints are insulated, giving us $u_x(-\pi, t) = u_x(\pi, t) = 0$ for all $t \geq 0$. These are called Neumann boundary conditions.

For our purposes, we will consider a mixture of these and say that we have periodic boundary conditions. To be specific, we want it so that for all $t \geq 0$, we have

$$\begin{aligned} u(-\pi, t) &= u(\pi, t), \\ u_x(-\pi, t) &= u_x(\pi, t). \end{aligned}$$

Now, we employ separation of variables, which allows us to find candidates for PDEs. We look for non-zero solutions of the form

$$u(x, t) = T(t)X(x),$$

where T and X are differentiable and not equal to 0 everywhere. Notice that if u solves the PDE $u_t = K u_{xx}$, then for all $t \geq 0$ and $x \in [-\pi, \pi]$, we have

$$T'(t)X(x) = KT(t)X''(x).$$

This implies that

$$\frac{T'(t)}{KT(t)} = \frac{X''(x)}{X(x)}$$

for all $t \geq 0$ such that $T(t) \neq 0$ and $x \in [-\pi, \pi]$ such that $X(x) \neq 0$. Now notice that if we keep t fixed and vary x, the value of $X''(x)/X(x)$ remains unchanged. Similarly, if we keep x fixed and vary t, the value of $T'(t)/[KT(t)]$ is also unchanged. Therefore, there exists some constant $\lambda \in \mathbb{R}$ such that

$$\lambda = \frac{T'(t)}{KT(t)} = \frac{X''(x)}{X(x)}.$$

This yields the equations

$$T'(t) = -\lambda K T(t), \tag{1.3}$$
$$X''(x) + \lambda X(x) = 0. \tag{1.4}$$

Now, we put the periodic boundary conditions into play. This gives us

$$\begin{aligned} T(t)X(\pi) &= T(t)X(-\pi), \\ T(t)X'(\pi) &= T(t)X'(-\pi). \end{aligned}$$

We will consider equation (1.4) first. Since we assumed that T is not identically 0, we obtain the eigenvalue problem for X given by the following three equations

$$\begin{aligned} X''(x) + \lambda X(x) &= 0, \\ X(\pi) &= X(-\pi), \\ X'(\pi) &= X'(-\pi). \end{aligned}$$

Let us now determine what values of λ will work.

Case 1. Suppose that $\lambda > 0$. Then we can write $\lambda = \omega^2$ for some $\omega > 0$. We obtain the equation

$$X''(x) + \omega^2 X(x) = 0,$$

whose only solutions are of the form

$$X(x) = C\cos(\omega x) + D\sin(\omega x)$$

for some constants C and D. Using the first boundary condition $X(\pi) = X(-\pi)$ gives us

$$2D\sin(\omega\pi) = 0,$$

so either $D = 0$ or $\omega \in \mathbb{N}$. Similarly, the second boundary condition $X'(\pi) = X'(-\pi)$ implies that either $C = 0$ or $2C\omega\sin(\omega\pi) = 0$, and the latter scenario means $\omega \in \mathbb{N}$. Therefore, we have established that for $n \in \mathbb{N}$, the functions

$$X_n(x) = C_n\cos(nx) + D_n\sin(nx)$$

with constants C_n and D_n are solutions to the eigenvalue problem.

Case 2. Suppose that $\lambda = 0$. Then $X''(x) = 0$, which means that

$$X(x) = C + Dx$$

for some constants C and D. It is easily verified that $X(\pi) = X(-\pi)$ gives $D = 0$, and that $X'(\pi) = X'(-\pi)$ gives nothing new. So $X_0(x) = C_0$ is a solution to the eigenvalue problem.

Case 3. Suppose that $\lambda < 0$. Then we can write $\lambda = -\omega^2$ for some $\omega > 0$. It follows that all solutions to $X''(x) - \omega^2 X = 0$ are of the form

$$X(x) = C\cosh(\omega x) + D\sinh(\omega x).$$

Now $X(\pi) = X(-\pi)$ implies that $2D\sinh(\omega\pi) = 0$ and $X'(\pi) = X'(-\pi)$ gives us $2C\omega\sinh(\omega\pi) = 0$. These together have no nonzero solutions.

Therefore, we have found that X is either of the form $X_0(x) = C_0$ for some constant C_0, or

$$X_n(x) = C_n\cos(nx) + D_n\sin(nx)$$

for some $n \in \mathbb{N}$ and constants C_n and D_n. Using equation (1.3), we see that $\lambda = 0$ implies that $T(t)$ is constant, and $\lambda = n^2 > 0$ implies that

$$T(t) = \exp(-Kn^2t).$$

Then, the solutions for u are $u_0(x,t) = T_0X_0 = C_0$, and

$$u_n(x,t) = \exp(-Kn^2t)(C_n\cos(nx) + D_n\sin(nx))$$

for all $n \in \mathbb{N}$. Using the Fourier method for PDEs, we notice that $u_t = Ku_{xx}$ is linear, so we can take linear combinations such as

$$u(x,t) = \sum_{n=0}^{N} u_n(x,t)$$

and still obtain a solution. Moreover, by formally interchanging sums and derivatives, the boundary conditions are also linear. However, finite sums can be insufficient for the initial conditions to be satisfied too. Thus, we instead consider formal infinite sums to get

$$u(x,t) = \sum_{n=0}^{\infty} u_n(x,t) = C_0 + \sum_{n=1}^{\infty} \exp(-Kn^2t)(C_n\cos(nx) + D_n\sin(nx)).$$

Assuming that the initial condition holds, this means we can write

$$f(x) = u(x,0) = C_0 + \sum_{n=1}^{\infty}(C_n\cos(nx) + D_n\sin(nx)).$$

The above form is known as a **Fourier series**. Now, we recall that we can write $\cos(nx) = (e^{inx} + e^{-inx})/2$ and $\sin(nx) = (e^{inx} - e^{-inx})/2$, so if we let

$$A_n = \begin{cases} C_0, & \text{if } n = 0, \\ (C_n - iD_n)/2, & \text{if } n > 0, \\ (C_{-n} + iD_{-n})/2, & \text{if } n < 0, \end{cases}$$

then we obtain the nice formula

$$f(x) = \sum_{n=-\infty}^{\infty} A_n e^{inx}.$$

This leads us to a few questions.

1. Are we justified in interchanging summation and differentiation?
2. Given some nice function f from $[-\pi,\pi]$ to $\mathbb{R}$ or $\mathbb{C}$, is it possible to express f as the infinite sum $f(x) = \sum_{n=-\infty}^{\infty} A_n e^{inx}$?
   (a) If so, in what sense does the sum converge?
   (b) How are f and the coefficients A_n related?

1.3 Basic Notation

We will get into answering the above questions later. First, we will make some definitions.

Definition 1.1

- We define $\mathbb{T} = \{z \in \mathbb{C} : |z| = 1\} = \{e^{i\theta} : \theta \in [-\pi,\pi]\}$ to be the unit circle in $\mathbb{C}$.
- We define $C(\mathbb{T})$ to be the continuous $\mathbb{C}$-valued functions on $\mathbb{T}$. Notice that we can view $C(\mathbb{T})$ as the space of 2π-periodic functions $\{f \in C[-\pi,\pi] : f(\pi) = f(-\pi)\}$.
- We define $R(\mathbb{T})$ to be the Riemann integrable functions over $\mathbb{T}$. Note that $R(\mathbb{T}) \supseteq C(\mathbb{T})$.

The space $C(\mathbb{T})$ has many nice norms.

- One such norm is
$$\|f\|_\infty = \sup_{\theta\in[-\pi,\pi]} |f(\theta)|.$$
In fact, $(C(\mathbb{T}), \|\cdot\|_\infty)$ is complete, so every Cauchy sequence in $C(\mathbb{T})$ converges to a limit in $C(\mathbb{T})$ with respect to $\|\cdot\|_\infty$.

- Another norm is given by
$$\|f\|_1 = \frac{1}{2\pi}\int_{-\pi}^{\pi} |f(\theta)|\,\mathrm{d}\theta.$$
Note that $C(\mathbb{T})$ is not complete with respect to $\|\cdot\|_1$; in fact, it is not even complete for $R(\mathbb{T})$, which hints to us that Riemann integrability may not be enough.

- For functions $f, g \in C(\mathbb{T})$, one can define an inner product by
$$\langle f, g\rangle = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(\theta)\overline{g(\theta)}\,\mathrm{d}\theta.$$
This gives us a norm
$$\|f\|_2 = \langle f, f\rangle^{1/2} = \left(\frac{1}{2\pi}\int_{-\pi}^{\pi} |f(\theta)|^2\,\mathrm{d}\theta\right)^{1/2}.$$

Now, let $f \in C(\mathbb{T})$, and assume that it makes sense to write it as
$$f(\theta) = \sum_{n=-\infty}^{\infty} A_n e^{in\theta}.$$
For example, the series could be uniformly convergent. What are the coefficients A_n?

Lemma 1.2

The set $\{e^{in\theta} : n \in \mathbb{Z}\}$ is an orthonormal system in $C(\mathbb{T})$, with
$$\langle e^{in\theta}, e^{im\theta}\rangle = \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{i(n-m)\theta}\,\mathrm{d}\theta = \begin{cases} 1, & \text{if } n = m, \\ 0, & \text{if } n \neq m. \end{cases}$$

Therefore, we would expect that
$$\begin{aligned}
\langle f, e^{in\theta}\rangle &= \frac{1}{2\pi}\int_{-\pi}^{\pi} f(\theta)e^{-in\theta}\,\mathrm{d}\theta \\
&= \frac{1}{2\pi}\lim_{N\to\infty}\int_{-\pi}^{\pi}\sum_{k=-N}^{N} A_k e^{ik\theta - in\theta}\,\mathrm{d}\theta \\
&= \lim_{N\to\infty}\sum_{k=-N}^{N} A_k\delta_{kn} = A_n.
\end{aligned}$$

Putting the main ideas into one line, we expect that
$$A_n = \langle f, e^{in\theta}\rangle = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(\theta)e^{-in\theta}\,\mathrm{d}\theta.$$

Definition 1.3

Let $f \in C(\mathbb{T})$ (or $R(\mathbb{T})$). The **n-th Fourier coefficient** of f is defined to be

$$\hat{f}(n) := \frac{1}{2\pi}\int_{-\pi}^{\pi} f(\theta)e^{-in\theta}\,\mathrm{d}\theta = \langle f, e^{in\theta}\rangle.$$

The (complex) Fourier series of f is then

$$f \sim \sum_{n\in\mathbb{Z}} \hat{f}(n)e^{in\theta}.$$

We now revisit the questions we asked earlier. In what sense does the Fourier series for f converge? For $f \in C(\mathbb{T})$, do we have

$$f(\theta) = \lim_{N\to\infty} \sum_{n=-N}^{N} \hat{f}(n)e^{in\theta} \tag{1.5}$$

for all $\theta \in [-\pi, \pi]$, meaning that we have pointwise convergence? Denoting $S_N(f)$ to be the N-th partial sum of the Fourier series, do we have

$$\lim_{N\to\infty} \|f - S_N(f)\|_2 = 0, \tag{1.6}$$

$$\lim_{N\to\infty} \|f - S_N(f)\|_1 = 0, \tag{1.7}$$

$$\lim_{N\to\infty} \|f - S_N(f)\|_\infty = 0, \tag{1.8}$$

for all $f \in C(\mathbb{T})$? We will show later that (1.8) implies both (1.5) and (1.6), and that (1.6) implies (1.7). So it would be great for (1.8) to hold (uniform convergence). Unfortunately, we have the following fact, which we will prove later.

Fact 1.4

There exists $f \in C(\mathbb{T})$ and $\theta_0 \in \mathbb{T}$ such that

$$\left|\lim_{N\to\infty} S_N(f)(\theta_0)\right| = \infty.$$

That is, the Fourier series diverges.

So that isn't ideal, but the good news is that we have the following result due to Carleson.

Fact 1.5: Carleson

For all $f \in C(\mathbb{T})$, we have

$$\lim_{N\to\infty} S_N(f)(\theta) = f(\theta)$$

for "almost all" $\theta \in [-\pi, \pi]$.

Note that "almost all" is a measure theoretic notion which we will define more rigorously later. We list one more useful fact.

Fact 1.6

The sequence $\{S_N(f)\}_{N=1}^{\infty}$ is Cauchy with respect to $\|\cdot\|_2$ on $C(\mathbb{T})$, so

$$\lim_{N,M\to\infty} \|S_N(f) - S_M(f)\|_2 = 0.$$

In fact, we have $\|f - S_N(f)\|_2 \to 0$ (mean square convergence).

But wait! We know that $(C(\mathbb{T}), \|\cdot\|_2)$ is a normed vector space, but it isn't complete! Using metric space theory from PMATH 351, there exists a completion $L^2(\mathbb{T}) = \overline{C(\mathbb{T})}$ of $C(\mathbb{T})$ with respect to $\|\cdot\|_2$, so $S_N(f)$ is actually converging in this larger space.

What exactly is $L^2(\mathbb{T})$? It's the space of "Lebesgue measurable functions" $f : [-\pi, \pi] \to \mathbb{C}$ that are "square integrable" with respect to Lebesgue measure! Lebesgue measure is a generalization of Riemann's integration theory. It is very useful in modern mathematics, particularly in studying Fourier series and their convergence.

2 Lebesgue Measure and Integration

2.1 Riemann Integration

Recall that in Riemann's theory of integration, we start with a bounded function $f : [a, b] \to \mathbb{R}$. We could then obtain $\int_a^b f(x)\,\mathrm{d}x$ via approximations of Riemann sums. More specifically, we take a partition

$$P = \{a = t_0 < t_1 < \cdots < t_n = b\}$$

of the interval $[a, b]$. For each $1 \le i \le n$, we set $m_i = \inf_{x\in[t_{i-1},t_i)} f(x)$ and $M_i = \sup_{x\in[t_{i-1},t_i)} f(x)$. We define the **lower Riemann sum** by

$$L(f, P) = \sum_{i=1}^{n} m_i(t_i - t_{i-1}),$$

and similarly, the **upper Riemann sum** by

$$U(f, P) = \sum_{i=1}^{n} M_i(t_i - t_{i-1}).$$

It is clear for all partitions P of $[a, b]$ that $L(f, P) \le U(f, P)$. Moreover, suppose P and Q are both partitions of $[a, b]$, and set $P \vee Q$ to be the partition consisting of all points in P and Q. Then recall that $P \vee Q$ refines both P and Q, and we have

$$L(f, P) \le L(f, P \vee Q) \le U(f, P \vee Q) \le U(f, Q).$$

Interchanging P and Q above gives us $L(f, Q) \le U(f, P)$, so we can deduce that

$$\sup_P L(f, P) \le \inf_P U(f, P).$$

That is, any lower Riemann sum of a given partition will always be at most the upper Riemann sum of any other partition.

Definition 2.1

We say that $f : [a, b] \to \mathbb{R}$ is **Riemann integrable** if

$$\sup_P L(f, P) = \inf_P U(f, P).$$

In this case, we write

$$\int_a^b f(x)\,\mathrm{d}x = \sup_P L(f, P) = \inf_P U(f, P).$$

We write $R[a, b]$ to denote the vector space of Riemann integrable functions $f : [a, b] \to \mathbb{R}$.

Riemann's theory is good for many purposes, such as for the Fundamental Theorem of Calculus or analysis over smooth manifolds. But there are also many deficiencies.

For one, it forces f to be bounded and "almost continuous". It also doesn't generalize to integration over sets that are not "like" $\mathbb{R}$ or $\mathbb{R}^N$. Sometimes, one wants to integrate functions over irregular sets, such as fractals.

Worst of all, there are no good limit theorems! Ideally, we want some kind of "monotone convergence theorem" which says that if we have a sequence of Riemann integrable functions $(f_n)_{n=1}^{\infty} \subseteq R[a, b]$ satisfying $f_1 \le f_2 \le \cdots$ and $f(x) = \lim_{n\to\infty} f_n(x)$ exists, then f is also Riemann integrable with

$$\int_a^b f(x)\,\mathrm{d}x = \lim_{n\to\infty} \int_a^b f_n(x)\,\mathrm{d}x.$$

Unfortunately, this result is false! Note that the pointwise limit of Riemann integrable functions might not even be Riemann integrable. Let $\{r_n\}_{n=1}^{\infty}$ be an enumeration of $\mathbb{Q} \cap [0,1]$, and for each $n \in \mathbb{N}$, define the function $f_n : [0,1] \to \mathbb{R}$ by

$$f_n(x) = \begin{cases} 1, & \text{if } x \in \{r_1, \ldots, r_n\}, \\ 0, & \text{otherwise.} \end{cases}$$

Then for all $n \in \mathbb{N}$, we have $f_n \in R[0,1]$ with

$$\int_0^1 f_n(x)\,\mathrm{d}x = 0.$$

Moreover, we see that $(f_n)_{n=1}^{\infty}$ converges pointwise to

$$f(x) = \begin{cases} 1, & \text{if } x \in \mathbb{Q} \cap [0,1], \\ 0, & \text{otherwise,} \end{cases}$$

the indicator function of $\mathbb{Q}$ over $[0,1]$. Notice that this is a nice monotone limit since $f_1 \leq f_2 \leq \cdots$. Even still, $f \notin R[0,1]$ since the rationals and irrationals are dense in $\mathbb{R}$, so given any partition P of $[0,1]$, the upper Riemann sum is $U(f,P) = 1$ and the lower Riemann sum is $L(f,P) = 0$.

2.2 Lebesgue Outer Measure

In the previous section, we saw that Riemann's theory of integration had some flaws. Lebesgue had the idea that we could do Riemann sums over partitions of the y-axis, instead of partitions over the x-axis. It turns out that this idea allows many more functions to be integrable.

Take a function $f : [a,b] \to \mathbb{R}$, and suppose that we have a partition $P = \{y_0 < y_1 < \cdots < y_n\}$ of the y-axis. Then we will take sums of terms of the form

$$y_i \cdot \ell(\{x \in [a,b] : f(x) \in (y_{i-1}, y_i]\}),$$

where $\ell(E)$ denotes the "length" of E. Notice that the set above is the preimage of the half-open interval $(y_{i-1}, y_i]$, and so we have an approximation

$$\int_a^b f(x)\,\mathrm{d}x \approx \sum_{i=1}^{n} y_i \cdot \ell(f^{-1}(y_{i-1}, y_i]).$$

Now, consider the case where f is the indicator function of $\mathbb{Q}$ over $[0,1]$ as we discussed above. Then we would have

$$\int_0^1 f(x)\,\mathrm{d}x \approx 1 \cdot \ell(\mathbb{Q} \cap [0,1]) + 0 \cdot \ell(\mathbb{Q}^c \cap [0,1]).$$

We would expect this to equal 0 if we want our "monotone convergence theorem" to hold, and also because $\mathbb{Q}$ is a countable set. As such, a desired property of our "length" function would be to have $\ell(\mathbb{Q} \cap [0,1]) = 0$.

We have now packaged the problem into generalizing the notion of length from nice sets (such as unions of intervals) to more complicated sets of the form $f^{-1}(y_{i-1}, y_i]$. This turns out to be a difficult task even for a continuous function f. We wish to have a function

$$m : \mathcal{P}(\mathbb{R}) \to [0, \infty]$$

satisfying the following properties:

(1) For any interval I from a to b (which could be open, closed, or half-open), we have

$$m(I) = b - a.$$

(2) **Translation invariance.** For all $x \in \mathbb{R}$, we have
$$m(E+x) = m(E),$$
where $E + x = \{y + x : y \in E\}$.

(3) **Countable additivity.** If $E_n \subseteq \mathbb{R}$ are disjoint for all $n \in \mathbb{N}$ and $E = \bigsqcup_{n=1}^{\infty} E_n$, then
$$m(E) = \sum_{n=1}^{\infty} m(E_n).$$

Bad news: there is no function $m : \mathcal{P}(\mathbb{R}) \to [0,\infty]$ satisfying all of (1) to (3)!

To see this, we will assume there is such a function m satisfying all 3 properties, and find a subset $E \subseteq [0,1)$ such that $m(E)$ is not well-defined. Define an equivalence relation on $[0,1)$ by $x \sim y$ if and only if $x - y \in \mathbb{Q}$. We leave it as an exercise to verify that this is indeed an equivalence relation. As usual, let $[x]$ be the equivalence class of each $x \in [0,1)$.

Choose a set E of representatives of all the equivalence classes $[x]$ for $x \in [0,1)$. Note that this is possible by the Axiom of Choice, and we have $E \subseteq [0,1)$ with $|E \cap [x]| = 1$ for all $x \in [0,1)$.

We claim that $m(E)$ is not well-defined. Let $\{r_n\}_{n=1}^{\infty}$ be an enumeration of $\mathbb{Q} \cap [0,1)$, and for each $n \in \mathbb{N}$, set
$$E_n = E + r_n \pmod 1 = ((E + r_n) \cap [0,1)) \sqcup ((E + r_n - 1) \cap [0,1)).$$

Since m satisfies properties (1) to (3), we have that
$$\begin{aligned} m(E_n) &= m((E + r_n) \cap [0,1)) + m((E + r_n - 1) \cap [0,1)) && \text{by (3)} \\ &= m((E + r_n) \cap [0,1)) + m((E + r_n) \cap [1,2)) && \text{by (2)} \\ &= m(E + r_n) && \text{by (3)} \\ &= m(E) && \text{by (2)} \end{aligned}$$
for all $n \in \mathbb{N}$. We leave it as an exercise to check that $[0,1) = \bigsqcup_{n=1}^{\infty} E_n$. It follows that
$$1 = m([0,1)) = \sum_{m=1}^{\infty} m(E_n) = \sum_{n=1}^{\infty} m(E)$$
where the second equality comes from (3), so there is no admissible value for $m(E)$.

In light of this result, we might be asking too much for m to satisfy all 3 properties. What can we do instead?

(a) We can restrict the domain of the function m to a more "tractable" family of subsets of $\mathbb{R}$. Naturally, we would want to allow all intervals, as well as open and closed sets, to be in this family.

(b) One other approach is to take a function m that works for all subsets $E \subseteq \mathbb{R}$ and agrees with our intuitive notion of length for intervals, but in doing so, sacrificing some of the desirable properties.

The standard approach is the first one, and we shall adopt it. We now introduce a candidate function that could be used for m.

Definition 2.2

The **Lebesgue outer measure** of a subset $E \subseteq \mathbb{R}$ is defined to be
$$m^*(E) = \inf\left\{ \sum_{n=1}^{\infty} \ell(I_n) : \{I_n\}_{n=1}^{\infty} \text{ a cover of } E \text{ by intervals} \right\}.$$

We look at some properties of the Lebesgue outer measure.

Proposition 2.3

(a) We have $m^*(\varnothing) = 0$ and $m^*(E) \geq 0$ for all $E \subseteq \mathbb{R}$.

(b) **Translation invariance.** For all $E \subseteq \mathbb{R}$ and $x \in \mathbb{R}$, we have $m^*(E + x) = m^*(E)$.

(c) **Monotonicity.** If $E \subseteq F \subseteq \mathbb{R}$, then $m^*(E) \leq m^*(F)$.

(d) **Countable subadditivity.** Suppose that $E = \bigcup_{n=1}^\infty E_n$ where the subsets $E_n \subseteq \mathbb{R}$ are not necessarily disjoint. Then

$$m^*(E) \leq \sum_{n=1}^\infty m^*(E_n).$$

Proof.

(a) This is clear from definition.

(b) Coverings of E by countable families of intervals are in bijection with those of $E + x$. In particular, we have $E \subseteq \bigcup_{n=1}^\infty I_n$ if and only if $E + x \subseteq \bigcup_{n=1}^\infty (I_n + x)$, with

$$\sum_{n=1}^\infty \ell(I_n) = \sum_{n=1}^\infty \ell(I_n + x).$$

Taking infima, we obtain $m^*(E) = m^*(E + x)$.

(c) Any covering $\bigcup_{n=1}^\infty I_n$ of F also gives a covering of E. On the other hand, coverings of E are not necessarily coverings of F. Then the infimum in $m^*(E)$ is taken over a larger collection than with $m^*(F)$, so we get $m^*(E) \leq m^*(F)$.

(d) Without loss of generality, suppose that $\sum_{n=1}^\infty m^*(E_n) < \infty$. Let $\varepsilon > 0$. Since $m^*(E_n) < \infty$ for all $n \in \mathbb{N}$, there exists a covering $\bigcup_{k=1}^\infty I_{k,n}$ of E_n such that

$$\sum_{k=1}^\infty \ell(I_{k,n}) < m^*(E_n) + \frac{\varepsilon}{2^n}.$$

Then we have $E = \bigcup_{n=1}^\infty E_n \subseteq \bigcup_{n,k=1}^\infty I_{k,n}$, so it follows that

$$\begin{aligned} m^*(E) &\leq \sum_{n=1}^\infty \sum_{k=1}^\infty \ell(I_{k,n}) \\ &\leq \sum_{n=1}^\infty \left(m^*(E_n) + \frac{\varepsilon}{2^n} \right) \\ &= \sum_{n=1}^\infty m^*(E_n) + \varepsilon. \end{aligned}$$

Since $\varepsilon > 0$ was arbitrary, the result follows. □

The following result tells us that we can compute $m^*(E)$ using "small" open intervals I_n in our covers.

Proposition 2.4

Let $E \subseteq \mathbb{R}$ and $\delta > 0$. Then

$$m^*(E) = \inf \left\{ \sum_{n=1}^\infty \ell(I_n) : \{I_n\}_{n=1}^\infty \text{ a cover of } E \text{ by open intervals such that } \ell(I_n) < \delta \right\}.$$

PROOF. It is clear that $m^*(E)$ is at most the right hand side, because the collection we are taking the infimum over in the definition of m^* is more general than the one in this proposition, which forces the intervals to be open and have length less than δ.

So, we turn to proving the other direction. Without loss of generality, we may assume that $m^*(E) < \infty$. Let $\varepsilon > 0$. We can find intervals $\{J_n\}_{n=1}^{\infty}$ such that $E \subseteq \bigcup_{n=1}^{\infty} J_n$ and

$$\sum_{n=1}^{\infty} \ell(J_n) \leq m^*(E) + \varepsilon.$$

Without loss of generality, we can also partition these intervals $\{J_n\}_{n=1}^{\infty}$ into subintervals so that $\ell(J_n) < \delta$ for all $n \in \mathbb{N}$. Now, choose open intervals $I_n \supseteq J_n$ such that $\ell(I_n) \leq \max\{\delta, \ell(J_n) + \varepsilon/2^n\}$. We see that $E \subseteq \bigcup_{n=1}^{\infty} I_n$ with

$$\sum_{n=1}^{\infty} \ell(I_n) \leq \sum_{n=1}^{\infty} \left(\ell(J_n) + \frac{\varepsilon}{2^n}\right) \leq (m^*(E) + \varepsilon) + \varepsilon = m^*(E) + 2\varepsilon.$$

Taking infima, we see that $m^*(E)$ is at least the right hand side and we are done. □

Next, we show that Lebesgue outer measure really generalizes length.

THEOREM 2.5

Let I be an interval with left endpoint a and right endpoint b, where $a < b \in \mathbb{R}$. Then

$$m^*(I) = \ell(I) = b - a.$$

PROOF. We first prove this for the case where $I = [a, b]$ is a compact interval. By taking $I_1 = I$ and $I_2 = \varnothing$ for $n \geq 2$, the collection $\{I_n\}_{n=1}^{\infty}$ is a cover of $[a, b]$ by intervals, which implies that

$$m^*(I) \leq \sum_{n=1}^{\infty} \ell(I_n) = \ell(I_1) = b - a.$$

Thus, we have $m^*(I) \leq b - a$. We now turn to showing that $m^*(I) \geq b - a$. Let $\varepsilon > 0$. By Proposition 2.4, we can find an open cover $\bigcup_{n=1}^{\infty}(a_n, b_n)$ of I such that

$$\sum_{n=1}^{\infty}(b_n - a_n) \leq m^*(I) + \varepsilon.$$

But I is compact, so we can find a finite subcover; that is, there exists $N \in \mathbb{N}$ such that $I \subseteq \bigcup_{n=1}^{N}(a_n, b_n)$. Without loss of generality, we may toss away any intervals I_n such that $I_n \cap I = \varnothing$, and reorder the I_n's if necessary to get $a_1 < a_2 < \cdots < a_N$. Since I is an interval, it is connected. Thus, the intervals (a_n, b_n) must overlap, and we obtain

$$\ell(I) = b - a \leq \sum_{n=1}^{N}(b_n - a_n) \leq \sum_{n=1}^{\infty}(b_n - a_n) \leq m^*(I) + \varepsilon.$$

But $\varepsilon > 0$ was arbitrary, so we have $m^*(I) \geq \ell(I) = b - a$, as desired.

We now prove the result in the case that $I = (a, b]$. For all $0 < \varepsilon < b - a$, we have $[a + \varepsilon, b] \subseteq (a, b] \subseteq [a, b]$. By the monotonicity of Lebesgue outer measure (Proposition 2.3), we see that

$$(b - a) - \varepsilon = m^*[a + \varepsilon, b] \leq m^*(a, b] \leq m^*[a, b] = b - a.$$

Since ε is arbitrary (subject to $0 < \varepsilon < b - a$), we deduce that $m^*(a, b] = b - a$. The cases where $I = [a, b)$ and $I = (a, b)$ are proved similarly. □

We have now shown that the Lebesgue outer measure is translation invariant and is a "good" generalization of length. Does countable additivity hold for m^*? The answer is of course no, because we showed that there is no function $m : \mathcal{P}(\mathbb{R}) \to [0, \infty]$ which simultaneously extends length, is translation invariant, and countably additive.

So when can countable or even just finite additivity hold? We first consider some special cases.

Definition 2.6

A set $E \subseteq \mathbb{R}$ is said to have **Lebesgue (outer) measure zero** if $m^*(E) = 0$.

The following lemma says that when the sets have Lebesgue measure zero, then m^* is countably additive.

Lemma 2.7

If $(E_n)_{n=1}^{\infty}$ are not necessarily disjoint sets with $m^*(E_n) = 0$ for all $n \in \mathbb{N}$, then

$$m^*\left(\bigcup_{n=1}^{\infty} E_n\right) = \sum_{n=1}^{\infty} m^*(E_n) = 0.$$

Proof. By subadditivity of m^*, we obtain

$$0 \leq m^*\left(\bigcup_{n=1}^{\infty} E_n\right) \leq \sum_{n=1}^{\infty} m^*(E_n) = 0. \qquad \square$$

Next, we consider the notion of distance between two non-empty subsets of $\mathbb{R}$.

Definition 2.8

Let $E, F \subseteq \mathbb{R}$ be non-empty. We define

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

to be the **distance** between E and F.

Notice that if $E, F \subseteq \mathbb{R}$ satisfy $d(E, F) > 0$, then they are certainly disjoint. In this case, it turns out that finite additivity holds for m^*.

Proposition 2.9

If $E, F \subseteq \mathbb{R}$ are such that $d(E, F) > 0$, then

$$m^*(E \sqcup F) = m^*(E) + m^*(F).$$

Proof. We will assume that $m^*(E), m^*(F) < \infty$. Otherwise, we get equality for free by observing that $E \sqcup F \supseteq E$ and $E \sqcup F \supseteq F$ and using the monotonicity of m^*.

By the countable subadditivity of m^*, we have $m^*(E \sqcup F) \leq m^*(E) + m^*(F)$, so we only need to prove the other direction. Let $\delta = d(E, F) > 0$, and let $\varepsilon > 0$. Then there is a covering of $E \sqcup F$ by intervals $\bigcup_{n=1}^{\infty} I_n$ such that $\ell(I_n) < \delta$ for all $n \in \mathbb{N}$ and

$$\sum_{n=1}^{\infty} \ell(I_n) < m^*(E \sqcup F) + \varepsilon.$$

Without loss of generality, we can toss away any intervals I_n such that $I_n \cap (E \sqcup F) = \varnothing$. Moreover, the restriction that $\ell(I_n) < \delta$ for all $n \in \mathbb{N}$ means that each I_n touches one of E or F, but not both. So we can partition $\{I_n\}_{n=1}^{\infty}$ into $\{I'_n\}_{n=1}^{\infty} \cup \{I''_n\}_{n=1}^{\infty}$, where the intervals I'_n only touch E and the intervals I''_n only touch F. Observe now that $\{I'_n\}_{n=1}^{\infty}$ is a covering of E and $\{I''_n\}_{n=1}^{\infty}$ is a covering of F, so we obtain

$$m^*(E) + m^*(F) \leq \sum_{n=1}^{\infty} \ell(I'_n) + \sum_{n=1}^{\infty} \ell(I''_n) \leq \sum_{n=1}^{\infty} \ell(I_n) < m^*(E \sqcup F) + \varepsilon.$$

Since $\varepsilon > 0$ was arbitrary, this completes the proof. □

Corollary 2.10

If $K_1, \dots, K_n \subseteq \mathbb{R}$ are pairwise disjoint compact sets, then

$$m^*\left(\bigsqcup_{i=1}^{n} K_i\right) = \sum_{i=1}^{n} m^*(K_i).$$

Proof. Observe that $d(E, F) > 0$ when E and F are compact with $E \cap F = \varnothing$. So this result follows by induction and applying Proposition 2.9. □

2.3 Lebesgue Measure and Lebesgue Measurable Sets

Our goal is to find a **large** class of subsets $\mathcal{L} \subseteq \mathcal{P}(\mathbb{R})$ so that countable additivity of m^* holds for $\mathcal{L}$. We want $\mathcal{L}$ to contain all intervals, closed sets, open sets, and anything else that can be built from them by countable unions and intersections. In other words, we ideally want $\mathcal{L}$ to be a so-called σ-algebra, which we define below.

Definition 2.11

Let X be a non-empty set. A family $\mathcal{M}$ of subsets of X is called a **σ-algebra** if the following three properties hold:

(1) $\varnothing \in \mathcal{M}$ and $X \in \mathcal{M}$;

(2) $E \in \mathcal{M}$ if and only if $E^c \in \mathcal{M}$; and

(3) if $\{E_n\}_{n=1}^{\infty}$ is a countable sequence in $\mathcal{M}$, then $\bigcup_{n=1}^{\infty} E_n \in \mathcal{M}$.

Remark 2.12

It follows immediately from the definition that σ-algebras are closed under countable intersections as well. Indeed, for a σ-algebra $\mathcal{M}$ and a countable sequence $\{E_n\}_{n=1}^{\infty}$ in $\mathcal{M}$, we have

$$\bigcap_{n=1}^{\infty} E_n = \left(\bigcup_{n=1}^{\infty} E_n^c\right)^c \in \mathcal{M}$$

by De Morgan's law. In fact, we could use countable intersection in the definition of a σ-algebra and derive countable unions from it.

Example 2.13

(a) Some examples of σ-algebras are $\mathcal{P}(X)$ and $\mathcal{M}_E = \{\varnothing, X, E, E^c\}$ for a subset $E \subseteq X$.

(b) The σ-algebras from part (a) are not very interesting; $\mathcal{P}(X)$ is too big and $\mathcal{M}_E$ is too small to work with. We give a slightly more interesting example. Let $\mathcal{F} \subseteq \mathcal{P}(X)$, and define

$$\mathcal{M}_{\mathcal{F}} = \bigcap_{\substack{\sigma\text{-algebras } \mathcal{M} \\ \text{on } X \text{ such} \\ \text{that } \mathcal{F} \subseteq \mathcal{M}}} \mathcal{M}.$$

Then $\mathcal{M}_{\mathcal{F}}$ is also a σ-algebra on X. In fact, it is the smallest σ-algebra containing $\mathcal{F}$. We call $\mathcal{M}_{\mathcal{F}}$ the **σ-algebra generated by** $\mathcal{F}$.

(c) Let (X, τ) be a topological space. Then

$$\mathcal{B}_X = \mathcal{M}_\tau \subseteq \mathcal{P}(X)$$

is called the **Borel σ-algebra**, and is the σ-algebra generated by open sets in X.

With the definition of a σ-algebra out of the way, we can now discuss what a measure on one should look like.

Definition 2.14

Let X be a non-empty set, and let $\mathcal{M}$ be a σ-algebra on X. We say that a function $\mu : \mathcal{M} \to [0, \infty]$ is a **measure on** $\mathcal{M}$ if we have

(1) $\mu(\varnothing) = 0$, and

(2) if $(E_n)_{n=1}^\infty$ is a sequence of pairwise disjoint sets in $\mathcal{M}$, then

$$\mu\left(\bigsqcup_{n=1}^\infty E_n\right) = \sum_{n=1}^\infty \mu(E_n).$$

In particular, the second property means that the measure μ is countably additive on the σ-algebra $\mathcal{M}$. Recall that we wanted our large set $\mathcal{L}$ above to be a σ-algebra on $\mathbb{R}$. Therefore, our hope is that $m = m^*|_{\mathcal{L}}$ is a measure on $\mathcal{L}$.

Let us now actually construct this large set $\mathcal{L}$. We first begin with a definition.

Definition 2.15

A set $E \subseteq \mathbb{R}$ is said to satisfy **Carathéodory's condition** if for any $A \subseteq \mathbb{R}$, we have

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

That is, m^* is additive when partitioning any set A with $\{E, E^c\}$.

Then, we set $\mathcal{L} := \{E \subseteq \mathbb{R} : E \text{ satisfies Carathéodory's condition}\}$. It turns out that this choice of $\mathcal{L}$ gives us exactly what we want. Of course, we will need to check that $\mathcal{L}$ is actually a rich collection of sets, but we will leave that for later.

Theorem 2.16: Carathéodory

The set $\mathcal{L}$ defined as above is a σ-algebra over $\mathbb{R}$, and $m = m^*|_{\mathcal{L}}$ is a measure on $\mathcal{L}$. We call m the **Lebesgue measure** on $\mathbb{R}$.

Proof. We will prove both of these claims simultaneously by beginning with the finite case, then lifting it to the infinite case.

Step 0. Recall that an **algebra** of sets $\mathcal{A} \subseteq \mathcal{P}(X)$ has the following properties:

(1) $\varnothing \in \mathcal{A}$ and $X \in \mathcal{A}$;

(2) $E \in \mathcal{A}$ if and only if $E^c \in \mathcal{A}$; and

(3) if $E, F \in \mathcal{A}$, then $E \cup F \in \mathcal{A}$.

In particular, this is weaker than being a σ-algebra, only being closed under finite unions (and by extension, finite intersections). We show that $\mathcal{L}$ is an algebra of sets in $\mathbb{R}$.

Proof of Step 0.

(1) Observe that for any $A \subseteq \mathbb{R}$, we have

$$m^*(A) = m^*(A \cap \mathbb{R}) + m^*(A \cap \varnothing).$$

Then $\mathbb{R} \in \mathcal{L}$ and by symmetry, $\varnothing \in \mathcal{L}$ as well.

(2) If $E \in \mathcal{L}$, then for any $A \subseteq \mathbb{R}$, we have

$$m^*(A) = m^*(A \cap E) + m^*(A \cap E^c),$$

so we see that $E^c \in \mathcal{L}$.

(3) Let $E, F \in \mathcal{L}$ and $A \subseteq \mathbb{R}$. We will show that $E \cap F \in \mathcal{L}$, and so $E \cup F = (E^c \cap F^c)^c \in \mathcal{L}$ by De Morgan's laws. First, observe that

$$\begin{aligned} m^*(A) &= m^*(A \cap E) + m^*(A \cap E^c) \\ &= m^*(A \cap E \cap F) + m^*(A \cap E \cap F^c) + m^*(A \cap E^c), \end{aligned}$$

where the first equality uses the fact that $E \in \mathcal{L}$, and the second equality comes from the fact that $F \in \mathcal{L}$ and applying it to $m^*(A \cap E)$. The idea now is to combine some terms to obtain a corresponding term $m^*(A \cap (E \cap F)^c)$. Indeed, we have

$$\begin{aligned} m^*(A \cap (E \cap F)^c) &= m^*(A \cap (E^c \cup F^c)) \\ &= m^*(A \cap (E^c \cup F^c) \cap E) + m^*(A \cap (E^c \cup F^c) \cap E^c) \\ &= m^*(A \cap F^c \cap E) + m^*(A \cap E^c). \end{aligned}$$

These last two terms are exactly what we had before, so combining everything gives us

$$m^*(A) = m^*(A \cap (E \cap F)) + m^*(A \cap (E \cap F)^c).$$

Thus, $E \cap F \in \mathcal{L}$, and we conclude that $\mathcal{L}$ is an algebra over $\mathbb{R}$. ■

Step 1. The Lebesgue outer measure m^* is finitely additive on $\mathcal{L}$. That is, if $E_1, \ldots, E_n \in \mathcal{L}$ are pairwise disjoint, then

$$m^*\left(\bigsqcup_{i=1}^{n} E_i\right) = \sum_{i=1}^{n} m^*(E_i).$$

Proof of Step 1. Let $E_1, \ldots, E_n \in \mathcal{L}$ be pairwise disjoint, and let $A \subseteq \mathbb{R}$. We know that $\bigsqcup_{i=1}^n E_i \in \mathcal{L}$ since we have already shown that $\mathcal{L}$ is an algebra. Thus, noting that $E_n \in \mathcal{L}$, we get

$$m^*\left(A \cap \bigsqcup_{i=1}^n E_i\right) = m^*\left(A \cap \bigsqcup_{i=1}^n E_i \cap E_n\right) + m^*\left(A \cap \bigsqcup_{i=1}^n E_i \cap E_n^c\right) = m^*(A \cap E_n) + m^*\left(A \cap \bigsqcup_{i=1}^{n-1} E_i\right),$$

where the final equality is because the sets are pairwise disjoint. Repeating this argument inductively yields

$$m^*\left(A \cap \bigsqcup_{i=1}^n E_i\right) = \sum_{i=1}^n m^*(A \cap E_i).$$

Taking $A = \mathbb{R}$ completes the proof of this step. ■

Step 2. If $\{E_i\}_{i=1}^\infty$ is a countable family of disjoint sets in $\mathcal{L}$, then $E = \bigsqcup_{i=1}^\infty E_i \in \mathcal{L}$ and

$$m^*(E) = \sum_{i=1}^\infty m^*(E_i).$$

Proof of Step 2. It suffices to show that

$$m^*(A) \geq m^*(A \cap E) + m^*(A \cap E^c)$$

for any $A \subseteq \mathbb{R}$ with $m^*(A) < \infty$. Indeed, by the subadditivity of m^*, we always have $m^*(A) \leq m^*(A \cap E) + m^*(A \cap E^c)$. Moreover, if $m^*(A) = \infty$, then there is certainly equality.

Let $A \subseteq \mathbb{R}$ be such that $m^*(A) < \infty$. Let $F_n = \bigsqcup_{i=1}^n E_i$, which is in $\mathcal{L}$ since it is an algebra. Arguing as in Step 1, we have

$$m^*(A) = m^*(A \cap F_n) + m^*(A \cap F_n^c) = \sum_{i=1}^n m^*(A \cap E_i) + m^*(A \cap F_n^c).$$

Notice that $F_n \subseteq E$, so $E^c \subseteq F_n^c$, which in turn implies $A \cap E^c \subseteq A \cap F_n^c$. By the monotonicity of m^*, we get

$$m^*(A \cap E^c) \leq m^*(A \cap F_n^c).$$

Applying this inequality to our above equation gives

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

Since this holds for all $n \in \mathbb{N}$, we can take $n \to \infty$ and the inequality will still hold. This means that

$$\begin{aligned} m^*(A) &\geq \sum_{i=1}^\infty m^*(A \cap E_i) + m^*(A \cap E^c) \\ &\geq m^*\left(\bigsqcup_{i=1}^\infty (A \cap E_i)\right) + m^*(A \cap E^c) \\ &= m^*(A \cap E) + m^*(A \cap E^c), \end{aligned}$$

where the second inequality follows from the countable subadditivity of m^*. Letting $A = E$, we deduce that

$$m^*(E) \geq \sum_{i=1}^\infty m^*(E_i) + 0 \geq m^*(E),$$

so we have the desired equality. ■

Step 3. We claim that $\mathcal{L}$ is closed under countable unions.

Proof of Step 3. Let $\{E_i\}_{i=1}^{\infty}$ be a countable family of not necessarily pairwise disjoint sets in $\mathcal{L}$. Let $E_1' = E_1$, and for each $i \geq 2$, set $E_i' = E_i \setminus (E_1 \cup \cdots \cup E_{i-1})$. Then $\{E_i'\}_{i=1}^{\infty}$ is a countable family of disjoint sets in $\mathcal{L}$ with

$$\bigcup_{i=1}^{\infty} E_i = \bigsqcup_{i=1}^{\infty} E_i',$$

and this is in $\mathcal{L}$ since we showed that countable unions of pairwise disjoint sets in $\mathcal{L}$ are also in $\mathcal{L}$. ■

This last step shows that $\mathcal{L}$ is a σ-algebra over $\mathbb{R}$ and Step 2 establishes that $m = m^*|_{\mathcal{L}}$ is a measure on $\mathcal{L}$, completing the proof. □

In light of this result, we say that a set satisfying Carathéodory's condition is a **Lebesgue measurable set**. Now, let's determine which sets belong to $\mathcal{L}$. Below, we see that sets of Lebesgue measure zero are an easy example of Lebesgue measurable sets.

Proposition 2.17

If $E \subseteq \mathbb{R}$ with $m^*(E) = 0$, then $E \in \mathcal{L}$.

Proof. Let $A \subseteq \mathbb{R}$. Notice that $m^*(A \cap E) \leq m^*(E) = 0$ by the monotonicity of m^*, so $m^*(A \cap E) = 0$. Similarly, we have $m^*(A \cap E^c) \leq m^*(A)$. Therefore, we obtain

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

and so we have the equality $m^*(A) = m^*(A \cap E) + m^*(A \cap E^c)$. □

In fact, it turns out that $\mathcal{B}_{\mathbb{R}}$, the Borel σ-algebra over $\mathbb{R}$, is contained in $\mathcal{L}$.

Theorem 2.18

We have $\mathcal{B}_{\mathbb{R}} \subseteq \mathcal{L}$.

Proof. Note that every open set can be written as the countable union of open intervals. Since $\mathcal{L}$ is a σ-algebra, it suffices to show that any open interval $I = (a, b)$ is in $\mathcal{L}$, where $a < b \in \mathbb{R}$ are finite.

Let $I = (a, b)$, and fix $A \subseteq \mathbb{R}$ such that $m^*(A) < \infty$. It is enough to show that

$$m^*(A) \geq m^*(A \cap I) + m^*(A \cap I^c).$$

Pick $n \in \mathbb{N}$ large enough so that $I_n = [a + \frac{1}{n}, b - \frac{1}{n}] \subseteq I$. Notice that

$$A = (A \cap I) \sqcup (A \cap I^c) \supseteq (A \cap I_n) \sqcup (A \cap I^c),$$

with $d(A \cap I_n, A \cap I^c) \geq \frac{1}{n} > 0$. By Proposition 2.9, we know that m^* is additive on sets with positive separation, so we have

$$m^*(A) \geq m^*((A \cap I_n) \sqcup (A \cap I^c)) = m^*(A \cap I_n) + m^*(A \cap I^c).$$

Now, we are close to the desired inequality. Note that $A \cap I = (A \cap I_n) \sqcup (A \cap I \setminus I_n)$, so by the subadditivity and monotonicity of m^*, we obtain

$$\begin{aligned} m^*(A) &\leq m^*(A \cap I_n) + m^*(A \cap I \setminus I_n) \\ &\leq m^*(A \cap I_n) + m^*(I \setminus I_n) \\ &= m^*(A \cap I_n) + \tfrac{2}{n}, \end{aligned}$$

where $m^*(I \setminus I_n) = m^*((a, a + \frac{1}{n})) + m^*((b - \frac{1}{n}, b)) = \frac{2}{n}$ since the intervals are disjoint. It follows that

$$\lim_{n\to\infty} m^*(A \cap I_n) = m^*(A \cap I),$$

and we deduce that $m^*(A) \geq m^*(A \cap I) + m^*(A \cap I^c)$ as desired. $\square$

Is it the case that $\mathcal{L}$ is actually equal to $\mathcal{B}_{\mathbb{R}}$? The answer is no, but it is difficult to explicitly write down a Lebesgue measurable set that is not in $\mathcal{B}_{\mathbb{R}}$. Instead, we can argue using the cardinalities of the sets.

We first consider the cardinality of $\mathcal{B}_{\mathbb{R}}$. Recall the following definitions from PMATH 351.

Definition 2.19

Let $E \subseteq \mathbb{R}$ be a set.

- We say that E is a G_δ**-set** if E is a countable intersection of open sets.
- We say that E is an F_σ**-set** if E is a countable union of closed sets.

Notice that every G_δ-set and F_σ-set is also a Borel set since $\mathcal{B}_{\mathbb{R}}$ is a σ-algebra, which is generated by the open sets in $\mathbb{R}$ and is closed under complements, countable unions and intersections. Moreover, we can also iterate these operations, so for instance, $G_{\delta\sigma}$ contains all sets that are countable unions of G_δ-sets.

Every open set $U \subseteq \mathbb{R}$ is a countable union of open intervals with rational endpoints. Let $\mathcal{F}$ be the set of all open intervals with rational endpoints, and observe that $\mathcal{F}$ is countable. Then any Borel set $E \subseteq \mathbb{R}$ can be generated from $\mathcal{F}$ by iterating the operations

$$G_\delta \to G_{\delta\sigma} \to G_{\delta\sigma\delta} \to \cdots$$

countably often. This means that $|\mathcal{B}_{\mathbb{R}}| = |\mathbb{N}|^{|\mathbb{N}|} = |\mathbb{R}|$.

We now consider the cardinality of $\mathcal{L}$. On Assignment 2, we will show that there is a set $\Delta \subseteq [0,1] \subseteq \mathbb{R}$ called the **Cantor middle thirds set** such that Δ is uncountable and $m^*(\Delta) = 0$. By Proposition 2.17, this means that $\Delta \in \mathcal{L}$. Then monotonicity tells us that any subset $E \subseteq \Delta$ is also in $\mathcal{L}$. In particular, we have $\mathcal{L} \supseteq \mathcal{P}(\Delta)$ and so

$$|\mathcal{L}| \geq |\mathcal{P}(\Delta)| = |\mathcal{P}(\mathbb{R})| = |\mathbb{R}|^{|\mathbb{R}|} > |\mathbb{R}| = |\mathcal{B}_{\mathbb{R}}|.$$

Thus, we get the following result.

Theorem 2.20

We have $\mathcal{B}_{\mathbb{R}} \subsetneq \mathcal{L}$.

Next, we show that $m = m^*|_{\mathcal{L}}$ is the unique measure on $\mathcal{L}$ extending length. Before we do that, we note that every measure μ is subadditive as a consequence of additivity. Indeed, let $E, F \subseteq \mathbb{R}$. We first show that

$$\mu(E \cup F) + \mu(E \cap F) = \mu(E) + \mu(F).$$

To begin, observe that we have $\mu(E) = \mu(E \cap F) + \mu(E \cap F^c)$ and $\mu(F) = \mu(F \cap E) + \mu(F \cap E^c)$ with $(E \cap F^c) \cap (F \cap E^c) = \varnothing$. Then additivity of μ gives us

$$\begin{aligned}\mu(E) + \mu(F) &= \mu(E \cap F^c) + 2\mu(E \cap F) + \mu(F \cap E^c) \\ &= \mu((E \cap F^c) \sqcup (E \cap F) \sqcup (F \cap E^c)) + \mu(E \cap F) \\ &= \mu(E \cup F) + \mu(E \cap F).\end{aligned}$$

It easily follows from this that

$$\mu(E \cup F) = \mu(E) + \mu(F) - \mu(E \cap F) \leq \mu(E) + \mu(F).$$

Next, we show that μ is monotone. Suppose that $E \subseteq F \subseteq \mathbb{R}$. Then we can write $F = (F \cap E) \sqcup (F \cap E^c) = E \sqcup (F \cap E^c)$, and since $\mu(E \cap F^c) \geq 0$, it follows from additivity that

$$\mu(E) \leq \mu(F \cap E^c) + \mu(E) = \mu(F).$$

Theorem 2.21: Uniqueness of Lebesgue measure

If $\mu : \mathcal{L} \to [0, \infty]$ is a measure on $\mathcal{L}$ such that $\mu(I) = m(I) = \ell(I)$ for all open intervals $I \subseteq \mathbb{R}$, then $\mu(E) = m(E)$ for all $E \in \mathcal{L}$.

Proof. First, we deal with the case where $E \in \mathcal{L}$ is bounded. Then there exists some $n \in \mathbb{N}$ such that $E \subseteq (-n, n)$. By the subadditivity of μ and m, we have $\mu(E) \leq \mu((-n, n)) = 2n$ and similarly, $m(E) \leq 2n$.

Let $\varepsilon > 0$. There exist open intervals $(I_k)_{k=1}^\infty$ such that $E \subseteq \bigcup_{k=1}^\infty I_k$ and

$$\sum_{k=1}^\infty \ell(I_k) < m(E) + \varepsilon.$$

By the monotonicity and subadditivity of μ, we see that

$$\mu(E) \leq \mu\left(\bigcup_{k=1}^\infty I_k\right) \leq \sum_{k=1}^\infty \mu(I_k) = \sum_{k=1}^\infty \ell(I_k) < m(E) + \varepsilon.$$

Since $\varepsilon > 0$ was arbitrary, we have $\mu(E) \leq m(E)$ for all bounded sets $E \in \mathcal{L}$. By replacing E above with $(-n, n) \setminus E$ and repeating the argument, we can deduce that $\mu((-n, n) \setminus E) \leq m((-n, n) \setminus E)$, so

$$2n = \mu((-n, n)) = \mu((-n, n) \setminus E) + \mu(E) \leq m((-n, n) \setminus E) + m(E) = m((-n, n)) = 2n$$

and we have equality throughout. Hence, we have $\mu(E) = m(E)$ for all bounded $E \in \mathcal{L}$.

Suppose now that $E \in \mathcal{L}$ is arbitrary. We can write $E = \bigsqcup_{n=1}^\infty E_n$ where each $E_n \in \mathcal{L}$ is bounded via the construction $E_n = E \cap (-n-1, n+1) \setminus (-n, n)$ for all $n \in \mathbb{N}$. Then countable additivity of the measures assures that

$$\mu(E) = \sum_{n=1}^\infty \mu(E_n) = \sum_{n=1}^\infty m(E_n) = m(E).$$ □

We have seen that not all Borel measurable sets are Lebesgue measurable. But the following structure theorem for $\mathcal{L}$ tells us that every $E \in \mathcal{L}$ is "almost" Borel measurable, up to throwing away sets of measure zero.

Theorem 2.22: Structure of Lebesgue measurable sets

Let $E \subseteq \mathbb{R}$. The following are equivalent:

(1) E is Lebesgue measurable.

(2) For all $\varepsilon > 0$, there exists an open set V and closed set F such that $F \subseteq E \subseteq V$ and $m(V \setminus F) < \varepsilon$.

(3) There exists an F_σ-set A and a G_δ-set B such that $A \subseteq E \subseteq B$ and $m(B \setminus A) = 0$.

Proof. (1) $\Rightarrow$ (2). Let $K_n = [-n, n]$ for each $n \in \mathbb{N}$, and note that $\mathbb{R} = \bigcup_{n=1}^{\infty} K_n$. Let E be a Lebesgue measurable set and let $\varepsilon > 0$. Notice that $m(E \cap K_n) \leq m(K_n) = 2n < \infty$ for all $n \in \mathbb{N}$. In particular, there exists an open set $V_n \subseteq \mathbb{R}$ such that

$$m(V_n \setminus (E \cap K_n)) < \frac{\varepsilon}{2^{n+1}}.$$

Indeed, just take $V_n = \bigcup_{k=1}^{\infty}(a_k, b_k) \supseteq E \cap K_n$ such that

$$m(V_n) < m(E \cap K_n) + \frac{\varepsilon}{2^{n+1}},$$

then use monotonicity to get $m(V_n \setminus (E \cap K_n)) \leq m(V_n)$.

Now, put $V = \bigcup_{n=1}^{\infty} V_n$. Observe that V is open with $E \subseteq V$, and

$$m(V \setminus E) = m\left(\bigcup_{n=1}^{\infty} V_n \setminus E\right) \leq m\left(\bigcup_{n=1}^{\infty} V_n \setminus (E \cap K_n)\right) \leq \sum_{n=1}^{\infty} m(V_n \setminus (E \cap K_n)) < \sum_{n=1}^{\infty} \frac{\varepsilon}{2^{n+1}} = \frac{\varepsilon}{2}.$$

We can repeat the same argument with E^c replacing E. This gives us an open set $W \supseteq E^c$ such that $m(W \setminus E^c) < \varepsilon/2$. Letting $F = W^c$, we see that F is closed and $F \subseteq E \subseteq V$. Using the additivity of Lebesgue measure gives

$$m(V \setminus F) = m(V \setminus E) + m(E \setminus F) = m(V \setminus E) + m(W \setminus E^c) < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

(2) $\Rightarrow$ (3). Applying (2) with $\varepsilon_k = 1/k$ for all $k \in \mathbb{N}$, we obtain closed sets $\{F_k\}_{k=1}^{\infty}$ and open sets $\{V_k\}_{k=1}^{\infty}$ such that $F_k \subseteq E \subseteq V_k$ and $m(V_k \setminus F_k) < 1/k$ for all $k \in \mathbb{N}$. Then $A = \bigcup_{k=1}^{\infty} F_k \subseteq E$ is an F_σ-set and $B = \bigcap_{k=1}^{\infty} V_k \supseteq E$ is a G_σ-set. We see that

$$B \setminus A \subseteq V_k \setminus F_k$$

for all $k \in \mathbb{N}$. Note that $m(B \setminus A)$ is a well-defined quantity since A and B are both Borel and hence Lebesgue measurable. Then monotonicity of m implies that

$$m(B \setminus A) \leq m(V_k \setminus F_k) < \frac{1}{k}.$$

Since this holds for all $k \in \mathbb{N}$, we in fact have $m(B \setminus A) = 0$.

(3) $\Rightarrow$ (1). Recall from Proposition 2.17 that if $W \subseteq \mathbb{R}$ with $m^*(W) = 0$, then $W \in \mathcal{L}$, and monotonicity shows that every subset $W_0 \subseteq W$ is also in $\mathcal{L}$. From (3), there exists an F_σ-set A and a G_δ-set B such that $A \subseteq E \subseteq B$. Writing $E = A \cup (E \setminus A)$, notice that $A \in \mathcal{L}$ since it is Borel. On the other hand, we have $E \setminus A \subseteq B \setminus A$ with $m(B \setminus A) = 0$, so $E \setminus A \in \mathcal{L}$. Thus, E is Lebesgue measurable as $\mathcal{L}$ is a σ-algebra and so closed under unions. $\square$

2.4 Measurable Functions

Recall that when integrating a function $f : \mathbb{R} \to \mathbb{R}$ using Lebesgue's approach of partitioning the y-axis, we needed to measure the length of sets of the form

$$f^{-1}(a, b) = \{x \in \mathbb{R} : f(x) \in (a, b)\}.$$

Therefore, we would like $m(f^{-1}(a, b))$ to be a well-defined quantity, or equivalently, we want $f^{-1}(a, b)$ to be a Lebesgue measurable set for any open interval (a, b).

More generally, let $f : X \to Y$ be an arbitrary function. Let B and $\{B_i\}_{i=1}^{\infty}$ be subsets of Y. Then we have $f^{-1}(B^c) = (f^{-1}(B))^c$ and

$$f^{-1}\left(\bigcup_{i=1}^{\infty} B_i\right) = \bigcup_{i=1}^{\infty} f^{-1}(B_i),$$

so the inverse image operation f^{-1} respects all the σ-algebra operations.

In particular, every open set $U \subseteq \mathbb{R}$ is of the form

$$U = \bigcup_{n=1}^{\infty} (a_n, b_n).$$

So our above condition where we want $f^{-1}(a, b) \in \mathcal{L}$ for any open interval (a, b) is equivalent to having

$$f^{-1}(U) = \bigcup_{i=1}^{\infty} f^{-1}(a_n, b_n) \in \mathcal{L}$$

for any open set $U \subseteq \mathbb{R}$. This leads us to the definition of a Lebesgue measurable function.

Definition 2.23

Let D be a Lebesgue measurable set. A function $f : D \to \mathbb{R}$ is called a **Lebesgue measurable function** if for any open set $U \subseteq \mathbb{R}$, we have

$$f^{-1}(U) \in \mathcal{L}.$$

Remark 2.24

This definition is reminiscent of the characterization of continuous functions via open sets; in fact, it is a generalization of it. Suppose that $f : D \to \mathbb{R}$ is continuous where $D \in \mathcal{L}$. Then for every open subset $U \subseteq \mathbb{R}$, we know that $f^{-1}(U)$ is relatively open in D; that is, it is of the form $D \cap V$ where $V \subseteq \mathbb{R}$ is open, so $f^{-1}(U) \in \mathcal{L}$.

Thinking more abstractly, we can generalize the above definition where the domain has a σ-algebra associated with it, and the codomain is a topological space.

Definition 2.25

Let X be a set, and let $\mathcal{M}$ be a σ-algebra on X. Let Y be a topological space with the topology τ_Y of open sets. Then $f : X \to Y$ is **measurable** with respect to $\mathcal{M}$ if for all open sets $U \in \tau_Y$, we have

$$f^{-1}(U) \in \mathcal{M}.$$

Proposition 2.26

Let $f : X \to Y$ be as in the above definition. The following are equivalent:

(1) f is measurable with respect to $\mathcal{M}$.

(2) For all Borel sets $E \in \mathcal{B}_Y$, we have $f^{-1}(E) \in \mathcal{M}$.

(3) If $\mathcal{F}$ is any collection of Borel sets in Y that generates $\mathcal{B}_Y$ (that is, $\mathcal{B}_Y = \mathcal{M}_{\mathcal{F}}$), then $f^{-1}(E) \in \mathcal{M}$ for all $E \in \mathcal{F}$.

Proof. We leave the proof as an exercise. Hint: Use the observation above where the preimage is closed under all the operations of a σ-algebra. $\square$

Proposition 2.27

Let $D \in \mathcal{L}$, and let $f : D \to \mathbb{R}$ be a Lebesgue measurable function. The following are equivalent:

(1) f is Lebesgue measurable.

(2) $f^{-1}((\alpha, \infty)) \in \mathcal{L}$ for all $\alpha \in \mathbb{R}$.

(3) $f^{-1}([\alpha, \infty)) \in \mathcal{L}$ for all $\alpha \in \mathbb{R}$.

(4) $f^{-1}((-\infty, \alpha)) \in \mathcal{L}$ for all $\alpha \in \mathbb{R}$.

(5) $f^{-1}((-\infty, \alpha]) \in \mathcal{L}$ for all $\alpha \in \mathbb{R}$.

Proof. It is clear that (1) implies (2) since (α, ∞) is open for all $\alpha \in \mathbb{R}$. For (2) implies (3), observe that

$$[\alpha, \infty) = \bigcap_{n=1}^{\infty} (\alpha - 1/n, \infty),$$

and since the inverse image preserves σ-algebra operations, we obtain

$$f^{-1}([\alpha, \infty)) = f^{-1}\left(\bigcap_{n=1}^{\infty} (\alpha - 1/n, \infty)\right) = \bigcap_{n=1}^{\infty} f^{-1}((\alpha - 1/n, \infty)) \in \mathcal{L}.$$

For (3) implies (4), we have $(-\infty, \alpha) = [\alpha, \infty)^c$ and so

$$f^{-1}((-\infty, \alpha)) = f^{-1}([\alpha, \infty)^c) = (f^{-1}([\alpha, \infty)))^c \in \mathcal{L}.$$

For (4) implies (5), we can use a similar argument to (2) implies (3), noting that

$$(-\infty, \alpha] = \bigcap_{n=1}^{\infty} (-\infty, \alpha + 1/n).$$

Finally, for (5) implies (1), it suffices to show that $f^{-1}((a, b)) \in \mathcal{L}$ for every bounded open interval (a, b). Indeed, we see that

$$(a, b) = (-\infty, a]^c \cap \bigcup_{n=1}^{\infty} (-\infty, b - 1/n],$$

and taking the inverse image completes the proof. $\square$

Using this characterization, we can see that there are many more examples of measurable functions other than continuous functions.

Example 2.28

(1) For a set $E \in \mathcal{L}$, the characteristic function $\chi_E : \mathbb{R} \to \{0, 1\}$ on E given by

$$\chi_E(x) = \begin{cases} 1, & \text{if } x \in E, \\ 0, & \text{if } x \notin E \end{cases}$$

is measurable, because given $\alpha \in \mathbb{R}$, we have

$$f^{-1}((\alpha, \infty)) = \begin{cases} E, & \text{if } \alpha \in (0, 1), \\ \varnothing, & \text{if } \alpha \geq 1, \\ \mathbb{R}, & \text{if } \alpha \leq 1, \end{cases}$$

all of which are Lebesgue measurable sets.

(2) Let $f : \mathbb{R} \to \mathbb{R}$ be a non-decreasing function. Then f is Lebesgue measurable since given $\alpha \in \mathbb{R}$, we have

$$f^{-1}((\alpha, \infty)) = \begin{cases} \varnothing, & \text{if } f(x) \le \alpha \text{ for all } x \in \mathbb{R}, \\ \mathbb{R}, & \text{if } f(x) > \alpha \text{ for all } x \in \mathbb{R}, \\ [a, \infty), & \text{if } f(x) > \alpha \text{ for all } x \ge a, \\ (a, \infty), & \text{if } f(x) > \alpha \text{ for all } x > a. \end{cases}$$

A similar computation shows that non-increasing functions are Lebesgue measurable, and thus all monotone functions are Lebesgue measurable.

(3) Let $D \in \mathcal{L}$. Let Y and Z be topological spaces. Let $g : Y \to Z$ be a continuous function, and let $f : D \to Y$ be Lebesgue measurable. Then $g \circ f : D \to Z$ is Lebesgue measurable. Indeed, if $U \subseteq Z$ is open, then $g^{-1}(U)$ is open in Y. From this, we see that

$$f^{-1}(g^{-1}(U)) = (g \circ f)^{-1}(U) \in \mathcal{L}.$$

More generally, we can allow for $g : Y \to Z$ to be **Borel measurable** where $g^{-1}(U)$ is a Borel set (rather than having to be open) for any open subset $U \subseteq Z$, and the same proof follows.

It would be nice to extend Lebesgue measurability to complex-valued functions $f : D \to \mathbb{C}$ such that by taking the typical decomposition $f = u + iv$ of real-valued functions $u, v : D \to \mathbb{R}$, we have that f is Lebesgue measurable if and only if u and v are. This is true, but first, we will prove a more general result.

Theorem 2.29

Let $D \in \mathcal{L}$, and let $u, v : D \to \mathbb{R}$ be Lebesgue measurable functions. Let Y be a topological space and let $\Phi : \mathbb{R}^2 \to Y$ be a continuous function. Then $h : D \to Y$ given by

$$h(x) = \Phi(u(x), v(x))$$

is also Lebesgue measurable.

Proof. Let $f : D \to \mathbb{R}^2$ be defined by $f(x) = (u(x), v(x))$. Then we have $h = \Phi \circ f$. Since Φ is continuous, it is enough to show that f is measurable due to part (3) of Example 2.28.

Recall that the topology on $\mathbb{R}^2$ is generated by bounded open rectangles

$$U = (a, b) \times (c, d),$$

so it suffices to show that $f^{-1}(U) \in \mathcal{L}$. We see that

$$f^{-1}(U) = u^{-1}((a, b)) \cap v^{-1}((c, d)).$$

Since u and v are Lebesgue measurable, both the above preimages are in $\mathcal{L}$, so we deduce that $f^{-1}(U) \in \mathcal{L}$. $\square$

Corollary 2.30

Let $D \in \mathcal{L}$.

(a) If $u, v : D \to \mathbb{R}$ is measurable, then $f = u + iv : D \to \mathbb{C}$ is also measurable.

(b) If $f : D \to \mathbb{C}$ is measurable, then $u = \mathrm{Re}(f)$, $v = \mathrm{Im}(f)$, and $|f|$ are all measurable.

(c) If $f, g : D \to \mathbb{C}$ are measurable, then αf, $f + g$, and fg are measurable for all $\alpha \in \mathbb{C}$.

PROOF.

(a) We can consider the topological isomorphism for $\mathbb{R}^2 \cong \mathbb{C}$. Taking $\Phi(u,v) = u + iv$ in Theorem 2.29, we obtain the result.

(b) Notice that Re, Im, and $|\cdot|$ are all continuous. Taking the composition of f with these functions and applying part (3) of Example 2.28 gives the result.

(c) For any $\alpha \in \mathbb{C}$, it is clear that αf is measurable. Assume for the moment that f and g are real-valued functions. Applying Theorem 2.29 with $\Phi(s,t) = s + t$ and $\Phi(s,t) = st$ which are both continuous, we find that $f + g$ and fg are measurable in this special case.

Now, suppose that f and g are complex-valued. Suppose that $f = u + iv$ and $g = w + iy$ where u, v, w, y are real-valued. By (1), we see that u, v, w, y are measurable. Then $f + g = (u + w) + i(v + y)$ and $fg = (uw - vy) + i(uy + vw)$ are also measurable by what we have just shown. $\square$

DEFINITION 2.31

Let $f : D \to \mathbb{C}$ and $g : E \to \mathbb{C}$ be functions where $D, E \in \mathcal{L}$. We say that $f = g$ **almost everywhere (a.e.)** if the set

$$\{f \neq g\} := (E \,\Delta\, D) \cap \{x \in E \cap D : f(x) \neq g(x)\}$$

has Lebesgue measure zero, where $E \,\Delta\, D$ denotes the symmetric difference of E and D.

We can think of f and g as essentially the "same" function, even if D and E have trivial intersection.

PROPOSITION 2.32

Let $D \in \mathcal{L}$, and let $f : D \to \mathbb{C}$ be a measurable function. Suppose that $g : E \to \mathbb{C}$. If $f = g$ almost everywhere, then $E \in \mathcal{L}$ and g is a measurable function satisfying

$$m(f^{-1}(U)) = m(g^{-1}(U))$$

for all open sets $U \subseteq \mathbb{C}$.

PROOF. Observe that we can write

$$\begin{aligned} E &= \{x \in E \cap D : f(x) = g(x)\} \cup (\{f \neq g\} \cap E) \\ &= (D \setminus \{f \neq g\}) \cup (\{f \neq g\} \cap E). \end{aligned}$$

The first set is the measurable set D minus the set of measure zero $\{f \neq g\}$ whereas the second has measure zero, so $E \in \mathcal{L}$. Now suppose that $U \subseteq \mathbb{C}$ is open. We have

$$g^{-1}(U) = (f^{-1}(U) \setminus \{f \neq g\}) \cup (g^{-1}(U) \cap \{f \neq g\}),$$

where the first set above is the set of elements in the preimage such that f and g agree, and the second set consists of the elements where either f and g disagree or f is not defined where g is. Noting that $f^{-1}(U) \in \mathcal{L}$, it follows from an analogous argument to above that $g^{-1}(U) \in \mathcal{L}$. Since $\{f \neq g\}$ is null, it is easy to see that $m(g^{-1}(U)) = m(f^{-1}(U))$. $\square$

The previous proposition tells us that when $f = g$ almost everywhere, then $f = g$ in the eyes of Lebesgue measure. When f is measurable, we are free to redefine f on measure zero sets, or extend the domain by f up to sets of measure zero. These operations do not "materially" change f.

To end this section, we say some words on extended real-valued functions. We let $[-\infty, \infty] = \mathbb{R} \cup \{\pm\infty\}$ be the **extended real numbers**, and we want to consider functions $f : D \to [-\infty, \infty]$ for some $D \in \mathcal{L}$. We will call f **Lebesgue measurable** if $f^{-1}((\alpha, \infty]) \in \mathcal{L}$ for all $\alpha \in \mathbb{R}$. In this case, we have $f^{-1}(\{\infty\}) \in \mathcal{L}$ and $f^{-1}(\{-\infty\}) \in \mathcal{L}$. Note that this notion only makes sense for real-valued functions, and not in $\mathbb{C}$.

2.5 Limits of Measurable Functions

Let $D \subseteq \mathbb{R}$, and let $(f_n)_{n=1}^{\infty}$ be a sequence of continuous functions defined on D. If $f_n \to f$ pointwise, recall that there is no guarantee that f is continuous. For example, letting $f_n(x) = x^n$ on $[0,1]$, then $f_n \to f$ where $f(x) = 0$ for $x \in [0,1)$ and $f(1) = 1$. We can only guarantee that $f = \lim_{n\to\infty} f_n$ is continuous when the convergence is uniform (which is equivalent to convergence in the supremum norm).

It turns out that measurability is a little nicer in this regard and does not require uniform convergence. Recall that $\sup_{n\in\mathbb{N}} f_n$ and $\inf_{n\in\mathbb{N}} f_n$ are defined such that $(\sup_{n\in\mathbb{N}} f_n)(x) = \sup_{n\in\mathbb{N}}(f_n(x))$ and $(\inf_{n\in\mathbb{N}} f_n)(x) = \inf_{n\in\mathbb{N}}(f_n(x))$.

Theorem 2.33

Let $(f_n)_{n=1}^{\infty}$ be a sequence of measurable extended real-valued functions defined on some $D \in \mathcal{L}$. Then $\sup_{n\in\mathbb{N}} f_n$ and $\inf_{n\in\mathbb{N}} f_n$ are both measurable functions.

Proof. Let $g = \sup_{n\in\mathbb{N}} f_n$ and take $\alpha \in \mathbb{R}$. Notice that $g(x) > \alpha$ if and only if there exists some $n \in \mathbb{N}$ such that $f_n(x) > \alpha$. Since each f_n is measurable, we obtain

$$g^{-1}((\alpha,\infty]) = \bigcup_{n=1}^{\infty} f_n^{-1}((\alpha,\infty]) \in \mathcal{L},$$

and thus g is measurable. It follows that $\inf_{n\in\mathbb{N}} f_n = -\sup_{n\in\mathbb{N}}(-f_n)$ is also measurable. □

Corollary 2.34

Let $D \in \mathcal{L}$.

(a) Let $f, g : D \to [-\infty,\infty]$ be measurable functions. Then $\max\{f,g\}$ and $\min\{f,g\}$ are both measurable.

(b) Let $f : D \to [-\infty,\infty]$ be a function. Define $f^+ = \max\{f,0\}$ and $f^- = -\min\{f,0\}$, and observe that $f = f^+ - f^-$. Then f is measurable if and only if f^+ and f^- are both measurable.

(c) Let $(f_n)_{n=1}^{\infty}$ be a sequence of real or complex valued functions on D. Suppose that $\lim_{n\to\infty} f_n(x) = f(x)$ exists for all $x \in D$. Then f is measurable.

Proof.

(a) Take the sequence $f_1 = f$ and $f_n = g$ for all $n \geq 2$, then apply Theorem 2.33 by observing that $\max\{f,g\} = \sup_{n\in\mathbb{N}} f_n$ and $\min\{f,g\} = \inf_{n\in\mathbb{N}} f_n$.

(b) The forward direction follows from (a). The backwards direction follows from part (c) of Corollary 2.30.

(c) Let $(f_n)_{n=1}^{\infty}$ be a sequence of real valued functions on D. Observe that

$$\limsup_{n\to\infty} f_n = \inf_{n\in\mathbb{N}} \left(\sup_{k\geq n} f_k \right)$$

is measurable by applying Theorem 2.33, and so is $\liminf_{n\to\infty} f_n = \sup_{n\in\mathbb{N}}(\inf_{k\geq n} f_k)$. Since $f_n \to f$ pointwise, we see that $f = \limsup_{n\to\infty} f_n = \liminf_{n\to\infty} f_n$ is measurable.

For the complex valued case, we can write $f_n = u_n + iv_n$ where u_n and v_n are real valued for each $n \in \mathbb{N}$. Then $u_n \to u$ and $v_n \to v$ with u and v being measurable, and thus $f = u + iv$ is measurable by Corollary 2.30. □

Corollary 2.34 tells us that measurable functions are closed under pointwise limits. We now discuss some other modes of convergence.

Definition 2.35

Let $D \in \mathcal{L}$. A sequence of measurable functions $(f_n)_{n=1}^{\infty}$ defined on D converges **pointwise almost everywhere** to a function f on D if

$$\lim_{n\to\infty} f_n(x) = f(x)$$

for almost every $x \in D$. We write $f_n \to f$ almost everywhere (a.e.).

Note that if we do not have pointwise convergence at a point, it either converges to the wrong value or the sequence doesn't converge at the point. The following lemma tells us that if $f_n \to f$ a.e., then it doesn't really matter what happens for points $x \in D$ where $f(x) \neq \lim_{n\to\infty} f_n(x)$.

Lemma 2.36

Let $D \in \mathcal{L}$. If $(f_n)_{n=1}^{\infty}$ is a sequence of measurable functions on D and $f_n \to f$ a.e., then f is measurable.

Proof. Define the function

$$g(x) = \begin{cases} \lim_{n\to\infty} f_n(x), & \text{if the limit exists,} \\ f(x), & \text{otherwise.} \end{cases}$$

Then by construction, we have $g = f$ almost everywhere. By Proposition 2.32, f is measurable. $\square$

Next, we talk about uniform convergence and almost uniform convergence.

Definition 2.37

Suppose that $(f_n)_{n=1}^{\infty}$ and f are functions from D to $\mathbb{C}$, $\mathbb{R}$, or $[-\infty, \infty]$.

- We say that $f_n \to f$ **uniformly** if for all $\varepsilon > 0$, there exists $N \in \mathbb{N}$ such that $n \geq N$ implies

$$|f_n(x) - f(x)| < \varepsilon.$$

- We say that $f_n \to f$ **almost uniformly** if for all $\varepsilon > 0$, there exists a measurable subset $E \subseteq D$ such that $m(E) < \varepsilon$ and $f_n \to f$ uniformly on $D \setminus E$.

It is clear that uniform convergence implies almost uniform convergence and pointwise convergence, while pointwise convergence implies pointwise almost everywhere convergence. Our goal is to determine the relationship between almost uniform convergence and pointwise almost everywhere convergence.

Lemma 2.38

If $f_n \to f$ almost uniformly, then $f_n \to f$ a.e. pointwise.

Proof. For each $m \in \mathbb{N}$, there exists a measurable subset $E_m \subseteq D$ with $m(E_m) < 1/m$ such that $f_m \to f$ uniformly on $D \setminus E_m$. Let $E = \bigcap_{m=1}^{\infty} E_m$. Using the continuity from above property of the Lebesgue measure from Assignment 2, we obtain

$$m(E) = \lim_{N\to\infty} m\left(\bigcap_{m=1}^{N} E_m\right) \leq \lim_{N\to\infty} m(E_N) = 0$$

since $(\bigcap_{m=1}^{N} E_m)_{N=1}^{\infty}$ is a decreasing sequence of sets. Therefore, E is a null set.

To show that $f_n \to f$ a.e. pointwise, it only remains to show that $f_n \to f$ for all $x \in D \setminus E$. Observe that

$$D \setminus E = D \setminus \bigcap_{m=1}^{\infty} E_m = \bigcup_{m=1}^{\infty} (D \setminus E_m).$$

In particular, if $x \in D \setminus E$, then there exists some $m \in \mathbb{N}$ such that $x \in D \setminus E_m$. Then $f_n \to f$ uniformly there, which implies pointwise convergence. $\square$

What about the converse? Does almost everywhere convergence imply almost uniform convergence? It turns out that this is true under the assumption that $m(D) < \infty$. This is known as Egorov's Theorem.

Theorem 2.39: Egorov's Theorem

Suppose that $D \in \mathcal{L}$ is such that $m(D) < \infty$. Let $(f_n)_{n=1}^{\infty}$ be a sequence of measurable functions $f_n : D \to \mathbb{C}$, and let $f : D \to \mathbb{C}$. If $f_n \to f$ pointwise a.e. on D, then $f_n \to f$ almost uniformly on D.

Proof. We may assume without loss of generality that $f_n \to f$ pointwise by replacing D with the set $\{x \in D : \lim_{n\to\infty} f_n(x) = f(x)\}$; we are only dropping a set of measure zero. Given $n, k \in \mathbb{N}$, we let

$$E(n,k) = \bigcup_{m \geq n} \{x \in D : |f_m(x) - f(x)| \geq 1/k\}.$$

We can think of this as the set of points were f_m is doing a bad job of converging to f.

It is clear that $E(n,k)$ is measurable for all $n, k \in \mathbb{N}$. Now, keep $k \in \mathbb{N}$ fixed. Then $(E(n,k))_{n=1}^{\infty}$ is a decreasing sequence of sets with $\bigcap_{n=1}^{\infty} E(n,k) = \varnothing$. Indeed, if there existed an element $x \in \bigcap_{n=1}^{\infty} E(n,k)$, then for all $n \in \mathbb{N}$, there exists $m \in \mathbb{N}$ such that $|f_m(x) - f(x)| \geq 1/k$. This contradicts our assumption that $f_n \to f$ for all $x \in D$. By the continuity from above of Lebesgue measure from Assignment 2, we have

$$\lim_{n\to\infty} m(E(n,k)) = m\left(\bigcap_{n=1}^{\infty} E(n,k)\right) = 0.$$

Note that we are using the fact that $m(E(1,n)) \leq m(D) < \infty$ here!

Let $\varepsilon > 0$. Choose a subsequence $(n_k)_{k=1}^{\infty}$ such that

$$m(E(n_k,k)) < \frac{\varepsilon}{2^k},$$

and set $E = \bigcup_{k=1}^{\infty} E(n_k,k)$. Then we have

$$m(E) \leq \sum_{k=1}^{\infty} m(E(n_k,k)) < \sum_{k=1}^{\infty} \frac{\varepsilon}{2^k} = \varepsilon.$$

For all $x \in D \setminus E$, we see that $x \in D \setminus E(n_k,k)$ for all $k \in \mathbb{N}$. Then $|f_m(x) - f(x)| < 1/k$ for all $k \in \mathbb{N}$ and $m \geq n_k$, implying that $f_m \to f$ uniformly on $D \setminus E$. $\square$

Example 2.40

For a simple example of Egorov's Theorem, consider the sequence of functions $f_n(x) = x^n$ on $[0,1]$. Recall that $f_n \to f$ pointwise on $[0,1]$ where $f(x) = 0$ on $[0,1)$ and $f(1) = 1$. But for all $\varepsilon > 0$, we see that $f_n \to f$ uniformly on $[0, 1-\varepsilon]$. In particular, we have $f_n \to f$ almost uniformly on $[0,1]$.

Remark 2.41

The assumption that $m(D) < \infty$ is necessary for Egorov's Theorem. Take $D = \mathbb{R}$, and define the sequence of functions $f_n = \chi_{[-n,n]}$. Then $\lim_{n\to\infty} f_n(x) = 1$ for all $x \in \mathbb{R}$.

We claim that $f_n \nrightarrow 1$ almost uniformly. Indeed, if $E \subseteq \mathbb{R}$ satisfies $m(E) < \varepsilon < \infty$, then $m(E^c) = \infty$. That is, E^c is unbounded. So there exists a sequence of points $x_k \in E^c$ such that $x_k > k$ for all $k \in \mathbb{N}$. This gives us

$$\sup_{x\in E^c} |f_k(x) - 1| \geq \sup_{k\in\mathbb{N}} |f_k(x_k) - 1| = 1,$$

and thus

$$\limsup_{k\to\infty} \|f_k - 1\|_\infty = 1.$$

We conclude that $f_n \nrightarrow 1$ uniformly on $E^c = \mathbb{R} \setminus E$.

We now consider simple functions, and see how measurable functions are well approximated by them. This will allow us to jump into Lebesgue integration.

Definition 2.42

Let $D \in \mathcal{L}$. A function $f : D \to \mathbb{C}$ is **simple** if its range is finite. That is, we have

$$\operatorname{range}(f) = \{\alpha_1, \dots, \alpha_n\} \subseteq \mathbb{C}$$

where the α_i are distinct elements.

Notice that if we let $E_i = f^{-1}(\{\alpha_i\})$, then $D = \bigsqcup_{i=1}^n E_i$ and $f = \sum_{i=1}^n \alpha_i \chi_{E_i}$. Moreover, f is measurable if and only if E_i is measurable for all $1 \leq i \leq n$.

In general, there are many ways to write a simple function as a linear combination of characteristic functions χ_{A_i}, where $A_i \in \mathcal{L}$. But the form $f = \sum_{i=1}^n \alpha \chi_{E_i}$ above is really the most natural representation of f as a linear combination of characteristic functions; we call it the **standard representation** of f.

The following theorem tells us that every measurable function is a limit of simple measurable functions.

Theorem 2.43

(a) Let $D \in \mathcal{L}$, and let $f : D \to [0, \infty]$ be a measurable function. Then there exists a sequence $(f_n)_{n=1}^\infty$ of non-negative simple measurable functions on D such that $0 \leq f_1 \leq f_2 \leq \cdots \leq f$ and for all $x \in D$, we have

$$\lim_{n\to\infty} f_n(x) = \sup_{n\in\mathbb{N}} f_n(x) = f(x).$$

Moreover, for any $R > 0$, we have $f_n \to f$ uniformly on the set $E_R = \{x \in D : f(x) \leq R\} \subseteq D$.

(b) Let $D \in \mathcal{L}$ and let $f : D \to \mathbb{C}$ be measurable. Then there exists a sequence of simple measurable functions $f_n : D \to \mathbb{C}$ such that

$$f(x) = \lim_{n\to\infty} f_n(x)$$

for all $x \in D$ and $0 \leq |f_1| \leq |f_2| \leq \cdots \leq |f|$. Moreover, for all $R > 0$, we have $f_n \to f$ uniformly on the set $E_R = \{x \in D : |f_n(x)| \leq R\} \subseteq D$.

Proof. For part (a), we first prove a special case. Let $g : [0, \infty] \to [0, \infty]$ be defined by $g(t) = t$. We want to approximate g by non-decreasing piecewise constant functions g_n.

Fix $n \in \mathbb{N}$. For $0 \leq t < \infty$, there is a unique $k \in \mathbb{N}$ such that $t \in [\frac{k}{2^n}, \frac{k+1}{2^n}]$. We write $k = k(t, n)$. Then, we define $g_n : [0, \infty] \to [0, \infty)$ by

$$g_n(t) = \begin{cases} k(n,t)/2^n, & \text{if } t < n, \\ n, & \text{if } n \leq t \leq \infty. \end{cases}$$

Observe that for all $n \in \mathbb{N}$, g_n is simple and non-decreasing. Moreover, for all $t \in [0, \infty]$, we have

$$g_n(t) \leq g_{n+1}(t) \leq g(t) = t.$$

One more thing to note is that for $t \in [0, n]$, we have

$$g_n(t) = \frac{k}{2^n} \leq t \leq \frac{k}{2^n} + \frac{1}{2^n} = g_n(t) + \frac{1}{2^n}.$$

This implies that $|g_n(t) - g(t)| \leq 2^{-n}$, so by picking n large enough so that $n > R$, we have $g_n \to g$ uniformly on the bounded interval $[0, R]$. Finally, we see that $g_n(t) \to g(t)$ for all $t \in [0, \infty]$.

We now prove the general case of part (a). Let $f : D \to [0, \infty]$ be a measurable function and define $f_n = g_n \circ f$ for all $n \in \mathbb{N}$. Then each f_n is simple as the g_n's are simple. Moreover, for all $x \in D$, we have

$$f_n(x) = g_n(f(x)) \leq g_{n+1}(f(x)) = f_{n+1}(x),$$

so the f_n's are increasing. We see that $f_n(x) \to f(x)$ for all $x \in D$ since $g_n(t) \to t$.

Let $R > 0$ and suppose that $x \in E_R = \{x \in D : f(x) \leq R\}$. Let $t = f(x)$. Then

$$|f(x) - f_n(x)| = |t - g_n(t)| < 2^{-n}$$

provided that $t \leq R < n$. Thus, $f_n \to f$ uniformly on E_R.

Finally, we claim that f_n is measurable for all $n \in \mathbb{N}$. Indeed, we saw earlier that the g_n are non-decreasing functions for all $n \in \mathbb{N}$, so $g_n^{-1}((\alpha, \infty])$ is an interval for all $\alpha \in \mathbb{R}$, say I. Then

$$f_n^{-1}((\alpha, \infty]) = f^{-1}(g_n^{-1}((\alpha, \infty])) = f^{-1}(I) \in \mathcal{L}$$

since f is measurable. This completes the proof of part (a).

To prove part (b), we can break $f : D \to \mathbb{C}$ into real and imaginary parts $u, v : D \to \mathbb{R}$ so that $f = u + iv$. We can then write these as $u = u^+ - u^-$ and $v = v^+ - v^-$, which are non-negative functions on D. We leave it as an exercise to verify the remaining details. $\square$

2.6 Lebesgue Integration

We will iteratively define the Lebesgue integral for measurable functions f on a measurable set $E \in \mathcal{L}$. First, we consider the Lebesgue integral of non-negative simple measurable functions. We then use this to define the integral of f where $f : E \to [0, \infty]$ is measurable, and derive a number of consequences of the definition.

Definition 2.44

Let $\varphi : \mathbb{R} \to [0, \infty)$ be a simple and measurable function with range $\{\alpha_1, \ldots, \alpha_n\}$. Let $\varphi = \sum_{i=1}^n \alpha_i \chi_{E_i}$ be its standard form with $E_i = \varphi^{-1}(\{\alpha_i\}) \in \mathcal{L}$. Then the **Lebesgue integral** of φ is

$$\int_{\mathbb{R}} \varphi \,\mathrm{d}m = \int_{\mathbb{R}} \varphi = \int \varphi := \sum_{i=1}^n \alpha_i m(E_i).$$

By convention, we will define $0 \cdot \infty = \infty \cdot 0 = 0$. For example, if $\alpha_i = 0$ and $m(E_i) = \infty$, then $\alpha_i m(E_i) = 0$.

Let us now extend this definition to arbitrary non-negative measurable extended real-valued functions.

Definition 2.45

Let $f : \mathbb{R} \to [0, \infty]$ be a measurable function. Then we define the **Lebesgue integral** of f to be

$$\int_{\mathbb{R}} f \, \mathrm{d}m = \int_{\mathbb{R}} f = \int f := \sup \left\{ \int_{\mathbb{R}} \varphi \, \mathrm{d}m : 0 \leq \varphi \leq f, \text{ where } \varphi \text{ is simple and measurable} \right\}.$$

For a measurable set $E \in \mathcal{L}$, we define

$$\int_E f \, \mathrm{d}m = \int_{\mathbb{R}} f \chi_E \, \mathrm{d}m.$$

We now list some basic properties of the Lebesgue integral which follow easily from the definition.

Proposition 2.46

(a) If $E \in \mathcal{L}$ and $0 \leq f \leq g$ are measurable functions, then

$$\int_E f \, \mathrm{d}m \leq \int_E g \, \mathrm{d}m.$$

(b) If $A \subseteq B$ are measurable sets and $f \geq 0$ is a measurable function, then

$$\int_A f \, \mathrm{d}m \leq \int_B f \, \mathrm{d}m.$$

(c) Let $E \in \mathcal{L}$. If $f \geq 0$ is a measurable function and $c \in [0, \infty)$, then

$$\int_E cf \, \mathrm{d}m = c \int_E f \, \mathrm{d}m.$$

(d) Let $E \in \mathcal{L}$, and suppose that $f \geq 0$ is a measurable function such that $f(x) = 0$ for all $x \in E$. Then we have

$$\int_E f \, \mathrm{d}m = 0.$$

(e) If $E \in \mathcal{L}$ with $m(E) = 0$, then for any measurable function $f \geq 0$, we have

$$\int_E f \, \mathrm{d}m = 0.$$

We want to have additivity of the Lebesgue integral. This will follow from the Lebesgue Monotone Convergence Theorem. This is a very powerful theorem, and it actually says a lot more: we actually have countable additivity of the Lebesgue integral!

Towards this, we will first prove some useful properties concerning non-negative simple measurable functions. It turns out that we can actually construct a measure by applying the Lebesgue integral to a non-negative simple measurable function, and that we have additivity of the Lebesgue integral when the integrands are simple functions.

Proposition 2.47

Let φ and ψ be non-negative simple measurable functions.

(1) For any $E \in \mathcal{L}$, define

$$\mu(E) = \int_E \varphi \, \mathrm{d}m = \int_{\mathbb{R}} \varphi \chi_E \, \mathrm{d}m.$$

Then μ is a measure.

(2) We have

$$\int_{\mathbb{R}} (\varphi + \psi) \, \mathrm{d}m = \int_{\mathbb{R}} \varphi \, \mathrm{d}m + \int_{\mathbb{R}} \psi \, \mathrm{d}m.$$

Proof. Let $\varphi = \sum_{i=1}^n \alpha_i \chi_{E_i}$ and $\psi = \sum_{j=1}^m \beta_j \chi_{F_j}$ be the standard forms of φ and ψ respectively.

(1) For $E \in \mathcal{L}$, observe that

$$\mu(E) = \int_{\mathbb{R}} \varphi \chi_E \, \mathrm{d}m = \int_{\mathbb{R}} \sum_{i=1}^n \alpha_i \chi_{E_i \cap E} \, \mathrm{d}m = \sum_{i=1}^m \alpha_i m(E_i \cap E).$$

It is clear that $\mu(\varnothing) = 0$ and $\mu(E) \geq 0$ since $\varphi \geq 0$. For countable additivity, suppose that $E = \bigsqcup_{k=1}^\infty A_k$ for some $A_k \in \mathcal{L}$. Then we obtain

$$\begin{aligned} \mu(E) &= \sum_{i=1}^n \alpha_i m\left(E \cap \bigsqcup_{k=1}^\infty A_k\right) = \sum_{i=1}^n \alpha_i \sum_{k=1}^\infty m(E_i \cap A_k) \\ &= \sum_{k=1}^\infty \sum_{i=1}^n \alpha_i m(E_i \cap A_k) = \sum_{k=1}^\infty \mu(A_k). \end{aligned}$$

Thus, we conclude that μ is a measure.

(2) Let $E_{ij} = E_i \cap F_j$ for $1 \leq i \leq n$ and $1 \leq j \leq m$. We see that $\mathbb{R} = \bigsqcup_{i,j} E_{ij}$. Note that $\varphi + \psi$ is a non-negative simple measurable function, so $\mu(E) = \int_E (\varphi + \psi) \, \mathrm{d}m$ is a measure by (1). This gives us

$$\int_{\mathbb{R}} (\varphi + \psi) \, \mathrm{d}m = \mu(\mathbb{R}) = \sum_{i,j} \mu(E_{ij}) = \sum_{i,j} \int_{E_{ij}} (\varphi + \psi) \, \mathrm{d}m.$$

Now, observe that $\varphi + \psi$ takes on value $\alpha_i + \beta_j$ on each $E_{ij} = E_i \cap F_j$. It follows that

$$\begin{aligned} \int_{\mathbb{R}} (\varphi + \psi) \, \mathrm{d}m = \mu(\mathbb{R}) &= \sum_{i,j} \int_{E_{ij}} (\varphi + \psi) \, \mathrm{d}m \\ &= \sum_{i,j} \int_{\mathbb{R}} (\alpha_i + \beta_j) \chi_{E_i \cap F_j} \, \mathrm{d}m \\ &= \sum_{i,j} (\alpha_i + \beta_j) m(E_i \cap F_j) \\ &= \sum_{i,j} \alpha_i m(E_i \cap F_j) + \sum_{i,j} \beta_j m(E_i \cap F_j) \\ &= \sum_{i,j} \int_{E_{ij}} \varphi \, \mathrm{d}m + \sum_{i,j} \int_{E_{ij}} \psi \, \mathrm{d}m \\ &= \int_{\mathbb{R}} \varphi \, \mathrm{d}m + \int_{\mathbb{R}} \psi \, \mathrm{d}m, \end{aligned}$$

where the final equality is obtained by applying (1) to φ and ψ individually. $\square$

We are now ready to state and prove the Lebesgue Monotone Convergence Theorem, which is our first big limit theorem. It tells us that the Lebesgue integral is much better behaved than the Riemann integral.

Theorem 2.48: Lebesgue Monotone Convergence Theorem

Let $(f_n)_{n=1}^{\infty}$ be a sequence of non-negative extended real-valued measurable functions on $E \in \mathcal{L}$. Suppose that $0 \leq f_1 \leq f_2 \leq \cdots$ and for all $x \in E$, let

$$f(x) = \lim_{n\to\infty} f_n(x) = \sup_{n\in\mathbb{N}} f_n(x).$$

Then f is measurable and we have

$$\int_E f \,\mathrm{d}m = \lim_{n\to\infty} \int_E f_n \,\mathrm{d}m.$$

We state an immediate corollary of Lebesgue's Monotone Convergence Theorem, which gives us a concrete way of defining the Lebesgue integral of f.

Corollary 2.49

Let $E \in \mathcal{L}$. Let $f \geq 0$ be measurable. For any sequence $(\varphi_n)_{n=1}^{\infty}$ of non-negative simple measurable functions with $0 \leq \varphi_1 \leq \varphi_2 \leq \cdots \leq f$ and $\lim_{n\to\infty} \varphi_n(x) = f(x)$ for all $x \in E$, we have

$$\int_E f \,\mathrm{d}m = \lim_{n\to\infty} \int_E \varphi_n \,\mathrm{d}m.$$

Let us now prove the theorem.

Proof. We already know that f is measurable by Theorem 2.33. By assumption, we have $0 \leq f_1 \leq f_2 \leq \cdots \leq f = \sup_{n\in\mathbb{N}} f_n$. Then for all $n \in \mathbb{N}$, using part (a) of Proposition 2.46 gives us

$$0 \leq \int_E f_n \,\mathrm{d}m \leq \int_E f_{n+1} \,\mathrm{d}m \leq \int_E f \,\mathrm{d}m.$$

From this, we obtain

$$I(f) = \lim_{n\to\infty} \int_E f_n \,\mathrm{d}m = \sup_{n\in\mathbb{N}} \int_E f_n \,\mathrm{d}m \leq \int_E f \,\mathrm{d}m.$$

This is actually one direction of the desired equality, so it only remains to show that

$$I(f) \geq \int_E f_n \,\mathrm{d}m. \tag{2.1}$$

Fix $\varepsilon \in (0, 1)$ and a simple and measurable function φ such that $0 \leq \varphi \leq f$. We claim it suffices to show that

$$I(f) \geq \varepsilon \int_E \varphi \,\mathrm{d}m. \tag{2.2}$$

Why is this enough? If we take $\varepsilon \to 1$ in (2.2), then this implies that $I(f) \geq \int_E \varphi \,\mathrm{d}m$ for all simple measurable functions $0 \leq \varphi \leq f$. Then taking the supremum over φ on the right hand side implies (2.1).

Set $E_n = \{x \in \mathbb{R} : f_n(x) \geq \varepsilon\varphi(x)\}$. By definition, we have $E_n \in \mathcal{L}$. Moreover, since $(f_n)_{n=1}^{\infty}$ is an increasing sequence, we have $E_1 \subseteq E_2 \subseteq \cdots$ with $E = \bigcup_{n=1}^{\infty} E_n$. Applying Proposition 2.47, we see that $\mu : \mathcal{L} \to [0, \infty]$ defined by

$$\mu(F) = \int_F \varepsilon\varphi \,\mathrm{d}m$$

is a measure. Then by the continuity of measure from below, we get

$$\int_E \varepsilon\varphi \, \mathrm{d}m = \mu(E) = \lim_{n\to\infty} \mu(E_n) = \lim_{n\to\infty} \int_{E_n} \varepsilon\varphi \, \mathrm{d}m.$$

But $\varepsilon\varphi \le f_n$ on E_n for all $n \in \mathbb{N}$ by definition, and thus

$$\int_{E_n} \varepsilon\varphi \, \mathrm{d}m \le \int_{E_n} f_n \, \mathrm{d}m \le \int_E f_n \, \mathrm{d}m \le \lim_{n\to\infty} \int_E f_n \, \mathrm{d}m = I(f).$$

Taking $n \to \infty$ gives us (2.2), which completes the proof. $\square$

We now discuss many consequences of Lebesgue's Monotone Convergence Theorem. The first one we will prove is the countable additivity of the Lebesgue integral for general non-negative measurable functions.

Corollary 2.50

Let $(f_n)_{n=1}^\infty$ be a sequence of non-negative measurable functions on $E \in \mathcal{L}$. Then $\sum_{n=1}^\infty f_n$ is measurable and we have

$$\int_E \left(\sum_{n=1}^\infty f_n \right) \mathrm{d}m = \sum_{n=1}^\infty \int_E f_n \, \mathrm{d}m.$$

Proof. We first prove finite additivity. Suppose that $(\varphi_n)_{n=1}^\infty$ and $(\psi_n)_{n=1}^\infty$ are non-negative simple measurable functions such that $\varphi_n \nearrow f_1$ and $\psi_n \nearrow f_2$ (where $\varphi_n \nearrow f_1$ means that $\varphi_1 \le \varphi_2 \le \cdots$ and $\varphi_n \to f_1$ pointwise). Then we see that $\varphi_n + \psi_n \nearrow f_1 + f_2$. It follows that

$$\begin{aligned}
\int_E (f_1 + f_2) \, \mathrm{d}m &= \lim_{n\to\infty} \int_E (\varphi_n + \psi_n) \, \mathrm{d}m \\
&= \lim_{n\to\infty} \left(\int_E \varphi_n \, \mathrm{d}m + \int_E \psi_n \, \mathrm{d}m \right) \\
&= \lim_{n\to\infty} \int_E \varphi_n \, \mathrm{d}m + \lim_{n\to\infty} \int_E \psi_n \, \mathrm{d}m \\
&= \int_E f_1 \, \mathrm{d}m + \int_E f_2 \, \mathrm{d}m,
\end{aligned}$$

where the first and last equality follow from Lebesgue's Monotone Convergence Theorem, and the second equality follows from Proposition 2.47 since φ_n and ψ_n are simple for all $n \in \mathbb{N}$. By induction, we obtain finite additivity.

Next, observe that $f_1 + \cdots + f_N \nearrow \sum_{n=1}^\infty f_n$. Applying Lebesgue's Monotone Convergence Theorem gives

$$\int_E \left(\sum_{n=1}^\infty f_n \right) \mathrm{d}m = \lim_{N\to\infty} \int_E \left(\sum_{n=1}^N f_n \right) \mathrm{d}m = \lim_{N\to\infty} \sum_{n=1}^N \int_E f_n \, \mathrm{d}m = \sum_{n=1}^\infty \int_E f_n \, \mathrm{d}m. \qquad \square$$

The next corollary tells us we can construct a measure from general non-negative measurable functions as well, not just simple ones.

Corollary 2.51

Suppose that $f : \mathbb{R} \to [0, \infty]$ is measurable. Define $\mu : \mathcal{L} \to [0, \infty]$ by

$$\mu(E) = \int_E f \, \mathrm{d}m.$$

Then μ is a measure.

Proof. It is clear that $\mu(\varnothing) = 0$ and $\mu(E) \geq 0$ for all $E \in \mathcal{L}$. Towards countable additivity, suppose that $E = \bigsqcup_{n=1}^{\infty} E_n$. Let $f_n = f\chi_{E_n}$ for each $n \in \mathbb{N}$. Observe that

$$f\chi_E = \sum_{n=1}^{\infty} f\chi_{E_n} = \sum_{n=1}^{\infty} f_n,$$

so applying Corollary 2.50 gives us

$$\mu(E) = \int_E f \,\mathrm{d}m = \int_{\mathbb{R}} f\chi_E \,\mathrm{d}m = \int_{\mathbb{R}} \left(\sum_{n=1}^{\infty} f_n \right) \mathrm{d}m = \sum_{n=1}^{\infty} \int_{\mathbb{R}} f_n \,\mathrm{d}m = \sum_{n=1}^{\infty} \int_{E_n} f \,\mathrm{d}m = \sum_{n=1}^{\infty} \mu(E_n). \qquad \square$$

We can think of f as the "density of μ relative to m". In the literature, this is often written $f = \frac{\mathrm{d}\mu}{\mathrm{d}m}$, and we call this the Radon-Nikodym derivative. This topic is focused on more heavily in PMATH 451.

The next result makes intuitive sense for how the integral should behave when $f = 0$ almost everywhere.

Corollary 2.52

If $f \geq 0$ is measurable and $E \in \mathcal{L}$, then $f = 0$ almost everywhere on E if and only if

$$\int_E f \,\mathrm{d}m = 0.$$

Proof. Suppose that $f = 0$ almost everywhere on E. Then we have

$$\int_E f \,\mathrm{d}m = \int_{\{x \in E \,:\, f(x) = 0\}} f \,\mathrm{d}m + \int_{\{x \in E \,:\, f(x) > 0\}} f \,\mathrm{d}m = 0,$$

where the first integral is 0 because $f(x) = 0$ everywhere on that set and the second integral is 0 because $m(\{x \in E : f(x) > 0\}) = 0$.

Conversely, suppose that $\int_E f \,\mathrm{d}m = 0$. For all $n \in \mathbb{N}$, define

$$E_n = \{x \in E : f(x) \geq 1/n\} \subseteq E.$$

We see that $E_n \nearrow \{x \in E : f(x) > 0\}$. For all $n \in \mathbb{N}$, the monotonicity of the Lebesgue integral implies that

$$0 = \int_E f \,\mathrm{d}m \geq \int_{E_n} f \,\mathrm{d}m \geq \int_{E_n} \frac{1}{n} \,\mathrm{d}m \geq 0.$$

This means that $\frac{1}{n} m(E_n) = 0$ and hence $m(E_n) = 0$ for all $n \in \mathbb{N}$. By the continuity of measure, we get

$$m(\{x \in E : f(x) > 0\}) = \lim_{n\to\infty} m(E_n) = 0.$$

Thus, $f = 0$ almost everywhere. $\square$

Remark 2.53

The monotonicity of the sequence $(f_n)_{n=1}^{\infty}$ in Lebesgue's Monotone Convergence Theorem is essential. For example, let $f_n = n\chi_{[0,1/n]}$ for all $n \in \mathbb{N}$. Then $f_n \to f$ pointwise where $f(0) = \infty$ and $f(x) = 0$ for all $x \neq 0$. This is certainly not a monotone limit. By Corollary 2.52, we have $\int_{\mathbb{R}} f \,\mathrm{d}m = 0$. But each f_n is simple and we see that

$$\int_{\mathbb{R}} f_n \,\mathrm{d}m = n \cdot \frac{1}{n} = 1$$

for all $n \in \mathbb{N}$. In particular, the conclusion of Lebesgue's Monotone Convergence Theorem fails, since

$$0 = \int_{\mathbb{R}} f \,\mathrm{d}m < \lim_{n\to\infty} \int_{\mathbb{R}} f_n \,\mathrm{d}m = 1.$$

Does the above inequality always hold? That is, do we have

$$\int_{\mathbb{R}} \left(\lim_{n\to\infty} f_n \right) \mathrm{d}m \leq \lim_{n\to\infty} \int_{\mathbb{R}} f_n \, \mathrm{d}m$$

given a sequence of non-negative measurable functions $(f_n)_{n=1}^{\infty}$? The problem is that these limits may not make sense. But by making a slight relaxation, it turns out that this does hold in general. This is known as Fatou's Lemma.

Theorem 2.54: Fatou's Lemma

If $(f_n)_{n=1}^{\infty}$ be a sequence of non-negative measurable functions, then

$$\int_{\mathbb{R}} \left(\liminf_{n\to\infty} f_n \right) \mathrm{d}m \leq \liminf_{n\to\infty} \int_{\mathbb{R}} f_n \, \mathrm{d}m.$$

Proof. By definition, we have

$$\liminf_{n\to\infty} f_n = \sup_{n\in\mathbb{N}} \left(\inf_{k\geq n} f_k \right).$$

Setting $h_n = \inf_{k\geq n} f_k$, we see that $h_n \leq h_{n+1} \leq \cdots$ and $h_n = \inf_{k\geq n} f_k \leq f_k$ for all $k \geq n$. This gives us

$$\int_{\mathbb{R}} h_n \, \mathrm{d}m = \int_{\mathbb{R}} \left(\inf_{k\geq n} f_k \right) \mathrm{d}m \leq \int_{\mathbb{R}} f_k \, \mathrm{d}m$$

for all $k \geq n$, and thus

$$\int_{\mathbb{R}} h_n \, \mathrm{d}m \leq \inf_{k\geq n} \int_{\mathbb{R}} f_k \, \mathrm{d}m.$$

Finally, applying Lebesgue's Monotone Theorem implies that

$$\int_{\mathbb{R}} \left(\liminf_{n\to\infty} f_n \right) \mathrm{d}m = \lim_{n\to\infty} \int_{\mathbb{R}} h_n \, \mathrm{d}m = \sup_{n\in\mathbb{N}} \int_{\mathbb{R}} h_n \, \mathrm{d}m \leq \sup_{n\in\mathbb{N}} \left(\inf_{k\geq n} \int_{\mathbb{R}} f_k \, \mathrm{d}m \right) = \liminf_{n\to\infty} \int_{\mathbb{R}} f_n \, \mathrm{d}m. \qquad \square$$

We now consider arbitrary complex-valued measurable functions. In order to avoid $\infty - \infty$ type situations, we first make the following definition.

Definition 2.55

The set of **Lebesgue integrable functions** over $\mathbb{R}$ is defined to be the collection

$$L^1 = L^1(\mathbb{R}) = \left\{ f : \mathbb{R} \to \mathbb{C} : f \text{ is measurable and } \int_{\mathbb{R}} |f| \, \mathrm{d}m < \infty \right\}.$$

A function $f \in L^1$ is said to be **absolutely integrable**. Given $E \in \mathcal{L}$, we can also consider the analogue

$$L^1(E) = \left\{ f : E \to \mathbb{C} : f \text{ is measurable and } \int_E |f| \, \mathrm{d}m < \infty \right\}.$$

For $E \in \mathcal{L}$ and $f \in L^1(E)$, we can write $f = u + iv$ where $u = \mathrm{Re}(f)$ and $v = \mathrm{Im}(f)$. We can again split these into $u = u^+ - u^-$ and $v = v^+ - v^-$, and we observe that $0 \leq u^{\pm}, v^{\pm} \leq |f|$. This means that $\int_E u^{\pm} \, \mathrm{d}m < \infty$ and $\int_E v^{\pm} \, \mathrm{d}m < \infty$, so $u^{\pm}, v^{\pm} \in L^1(E)$.

Definition 2.56

Let $E \in \mathcal{L}$ and $f \in L^1(E)$. We define the **Lebesgue integral of f over $\mathbb{C}$** to be

$$\int_E f \, \mathrm{d}m = \left(\int_E u^+ \, \mathrm{d}m - \int_E u^- \, \mathrm{d}m \right) + i \left(\int_E v^+ \, \mathrm{d}m - \int_E v^- \, \mathrm{d}m \right).$$

Theorem 2.57

Let E be a measurable set.

(a) $L^1(E)$ is a $\mathbb{C}$-vector space under pointwise addition and scalar multiplication.

(b) The map $f \mapsto \int_E f \, \mathrm{d}m$ is a linear functional over $L^1(E)$.

(c) For any $f \in L^1(E)$, we have

$$\left| \int_E f \, \mathrm{d}m \right| \leq \int_E |f| \, \mathrm{d}m.$$

Proof.

(a) Let $f, g \in L^1(E)$ and $\alpha, \beta \in \mathbb{C}$. We know that $\alpha f + \beta g$ is measurable with

$$|\alpha f + \beta g| \leq |\alpha||f| + |\beta||g|.$$

It follows that

$$\int_E |\alpha f + \beta g| \, \mathrm{d}m \leq \int_E (|\alpha||f| + |\beta||g|) \, \mathrm{d}m \leq |\alpha| \int_E |f| \, \mathrm{d}m + |\beta| \int_E |g| \, \mathrm{d}m < \infty,$$

so $\alpha f + \beta g \in L^1(E)$.

(b) To show that $f \mapsto \int_E f \, \mathrm{d}m$ is a linear functional, it suffices to check that $\int_E (f+g) \, \mathrm{d}m = \int_E f \, \mathrm{d}m + \int_E g \, \mathrm{d}m$ and $\int_E \alpha f \, \mathrm{d}m = \alpha \int_E f \, \mathrm{d}m$ for all $f, g \in L^1(E)$ and $\alpha \in \mathbb{C}$. Moreover, it is enough to consider the case where f and g are real-valued since the result follows from considering Definition 2.56.

Let $h = f + g$. Write $f = f^+ - f^-$, $g = g^+ - g^-$, and $h = h^+ - h^-$. Then we see that

$$h^+ + f^- + g^- = f^+ + g^+ + h^-.$$

Recall from Corollary 2.50 that the Lebesgue integral is additive for non-negative measurable functions and thus

$$\int_E h^+ \, \mathrm{d}m + \int_E f^- \, \mathrm{d}m + \int_E g^- \, \mathrm{d}m = \int_E f^+ \, \mathrm{d}m + \int_E g^+ \, \mathrm{d}m + \int_E h^- \, \mathrm{d}m.$$

Note that all the above values are finite since $f, g \in L^1(E)$, so rearranging yields

$$\begin{aligned} \int_E (f+g) \, \mathrm{d}m &= \int_E h^+ \, \mathrm{d}m - \int_E h^- \, \mathrm{d}m \\ &= \int_E f^+ \, \mathrm{d}m - \int_E f^- \, \mathrm{d}m + \int_E g^+ \, \mathrm{d}m - \int_E g^- \, \mathrm{d}m \\ &= \int_E f \, \mathrm{d}m + \int_E g \, \mathrm{d}m. \end{aligned}$$

Next, we show that $\int_E \alpha f \, \mathrm{d}m = \alpha \int_E f \, \mathrm{d}m$ for all $f \in L^1(E)$ and $\alpha \in \mathbb{C}$. If $\alpha \geq 0$, then there is nothing to prove. For $\alpha = -1$, we see that

$$\alpha f = -u + i(-v) = (u^- - u^+) + i(v^- - v^+).$$

Then we obtain

$$\begin{aligned}\int_E (-f)\,\mathrm{d}m &= \left(\int_E u^- \,\mathrm{d}m - \int_E u^+ \,\mathrm{d}m\right) + i\left(\int_E v^- \,\mathrm{d}m - \int_E v^+ \,\mathrm{d}m\right) \\ &= -\left[\left(\int_E u^+ \,\mathrm{d}m - \int_E u^- \,\mathrm{d}m\right) + i\left(\int_E v^+ \,\mathrm{d}m - \int_E v^- \,\mathrm{d}m\right)\right] \\ &= -\int_E f \,\mathrm{d}m.\end{aligned}$$

The $\alpha = i$ case can be proved similarly. Then the result for all $\alpha \in \mathbb{C}$ follows from the above cases by using the fact that $\int_E (f+g)\,\mathrm{d}m = \int_E f\,\mathrm{d}m + \int_E g\,\mathrm{d}m$.

(c) Note that $\int_E f\,\mathrm{d}m \in \mathbb{C}$, so we can pick $\alpha \in \mathbb{C}$ with $|\alpha| = 1$ so that

$$\left|\int_E f\,\mathrm{d}m\right| = \alpha \int_E f\,\mathrm{d}m.$$

Then we have

$$0 \le \left|\int_E f\,\mathrm{d}m\right| = \alpha\int_E f\,\mathrm{d}m = \int_E \alpha f\,\mathrm{d}m = \int_E \mathrm{Re}(\alpha f)\,\mathrm{d}m + i\int_E \mathrm{Im}(\alpha f)\,\mathrm{d}m.$$

Note that the term $i\int_E \mathrm{Im}(\alpha f)\,\mathrm{d}m$ vanishes since the above quantities are real. Since $\mathrm{Re}(\alpha f) \le |\alpha f| = |f|$, we deduce that

$$\left|\int_E f\,\mathrm{d}m\right| = \int_E \mathrm{Re}(\alpha f)\,\mathrm{d}m \le \int_E |f|\,\mathrm{d}m. \qquad \square$$

Now, we discuss Lebesgue's Dominated Convergence Theorem. This result is very powerful, and we will see many applications of it down the road. One such result is the completeness of the L^p spaces.

Theorem 2.58: Lebesgue's Dominated Convergence Theorem

Let $E \in \mathcal{L}$, and let $(f_n)_{n=1}^\infty$ be a sequence of functions in $L^1(E)$. Assume that there exists a non-negative function $g \in L^1(E)$ such that $0 \le |f_n| \le g$ for all $n \in \mathbb{N}$, and suppose that

$$f(x) = \lim_{n\to\infty} f_n(x)$$

exists for all $x \in E$. Then $f \in L^1(E)$ and

$$\lim_{n\to\infty}\int_E |f - f_n|\,\mathrm{d}m = 0.$$

In particular, by applying part (c) of Theorem 2.57, we see that

$$\left|\int_E f\,\mathrm{d}m - \int_E f_n\,\mathrm{d}m\right| = \left|\int_E (f - f_n)\,\mathrm{d}m\right| \le \int_E |f - f_n|\,\mathrm{d}m \to 0,$$

and so we get a similar result to Lebesgue's Monotone Convergence Theorem, namely

$$\int_E f\,\mathrm{d}m = \lim_{n\to\infty}\int_E f_n\,\mathrm{d}m.$$

Proof. We have $f \in L^1(E)$ since $|f_n| \le g$ for all $n \in \mathbb{N}$ and so $|f| \le g$. Since $g \in L^1(E)$, we obtain

$$\int_E |f|\,\mathrm{d}m \le \int_E g\,\mathrm{d}m < \infty.$$

Next, define $F_n(x) = 2g(x) - |f(x) - f_n(x)| \geq 0$. It is clear that each F_n is measurable. Note that

$$\liminf_{n\to\infty} F_n = 2g$$

since $|f(x) - f_n(x)| \to 0$, so applying Fatou's Lemma (Theorem 2.54) to $(F_n)_{n=1}^{\infty}$ gives us

$$\begin{aligned}\int_E 2g \, \mathrm{d}m = \int_E \liminf_{n\to\infty} F_n \, \mathrm{d}m &\leq \liminf_{n\to\infty} \int_E F_n \, \mathrm{d}m \\ &= \liminf_{n\to\infty} \left(\int_E 2g \, \mathrm{d}m - \int_E |f - f_n| \, \mathrm{d}m \right) \\ &= \int_E 2g \, \mathrm{d}m + \liminf_{n\to\infty} \left(- \int_E |f - f_n| \, \mathrm{d}m \right).\end{aligned}$$

Observe that

$$0 \leq \liminf_{n\to\infty} \left(- \int_E |f - f_n| \, \mathrm{d}m \right) = - \limsup_{n\to\infty} \int_E |f - f_n| \, \mathrm{d}m,$$

and so we have

$$0 \leq \limsup_{n\to\infty} \int_E |f - f_n| \, \mathrm{d}m \leq 0.$$

Since $f_n \to f$ pointwise, we conclude that

$$\lim_{n\to\infty} \int_E |f - f_n| \, \mathrm{d}m = 0. \qquad \square$$

Remark 2.59

We note that there are "almost everywhere" versions of Lebesgue's Monotone Convergence Theorem and Dominated Convergence Theorem. The almost everywhere version of the Monotone Convergence Theorem states that if $(f_n)_{n=1}^{\infty}$ is a sequence of non-negative measurable functions on $E \in \mathcal{L}$ with $f_n \leq f_{n+1}$ almost everywhere for all $n \in \mathbb{N}$, then

$$f(x) = \lim_{n\to\infty} f_n(x) = \sup_{n\in\mathbb{N}} f_n(x)$$

exists for almost every $x \in E$ and

$$\int_E f \, \mathrm{d}m = \lim_{n\to\infty} \int_E f_n \, \mathrm{d}m.$$

We give a brief sketch of the proof. By taking

$$F = \{x \in E : f_n(x) > f_{n+1}(x) \text{ for some } n \in \mathbb{N}\} = \bigcup_{n\in\mathbb{N}} \{x \in E : f_n(x) > f_{n+1}(x)\},$$

we observe that each set $\{x \in E : f_n(x) > f_{n+1}(x)\}$ has measure 0 and so $m(F) = 0$. Then we can restrict our attention to F^c and observe that $f_n \nearrow f = \sup_{n\in\mathbb{N}} f_n$ there, in which case we can apply the normal version of Lebesgue's Monotone Convergence Theorem. The value of the integral is unaffected by the measure zero set F.

The almost everywhere version of the Dominated Convergence Theorem states that if $(f_n)_{n=1}^{\infty}$ is a sequence of functions in $L^1(E)$ such that $|f_n| \leq g$ almost everywhere for all $n \in \mathbb{N}$ and $f_n \to f$ almost everywhere, then $f \in L^1(E)$ and

$$\lim_{n\to\infty} \int_E |f - f_n| \, \mathrm{d}m = 0.$$

We leave its proof as an exercise.

Finally, we show that the Riemann integral is a special case of the Lebesgue integral in some sense.

Theorem 2.60

Let $f : [a, b] \to \mathbb{R}$ be a Riemann integrable function. Then f is measurable and $f \in L^1([a, b])$. Moreover, we have

$$\int_a^b f(x)\,\mathrm{d}x = \int_{[a,b]} f\,\mathrm{d}m.$$

Proof. Let $f : [a, b] \to \mathbb{R}$ be Riemann integrable. Note that by taking a sequence $(P_n)_{n=1}^{\infty}$ of partition refinements of $[a, b]$ with mesh approaching 0, we obtain step functions $(u_n)_{n=1}^{\infty}$ and $(\ell_n)_{n=1}^{\infty}$ associated with the upper and lower Riemann sums. In particular, we have $\ell_n \le f \le u_n$ for all $n \in \mathbb{N}$ and

$$\int_a^b f(x)\,\mathrm{d}x = \sup_{n\in\mathbb{N}} \int_a^b \ell_n(x)\,\mathrm{d}x = \inf_{n\in\mathbb{N}} \int_a^b u_n(x)\,\mathrm{d}x.$$

Moreover, we have $\ell_n \le \ell_{n+1}$ and $u_n \ge u_{n+1}$ for all $n \in \mathbb{N}$. It is easy to see that the Riemann integral coincides with the definition of the Lebesgue integral for simple step functions. Thus, we have

$$\int_a^b u_n(x)\,\mathrm{d}x = \int_{[a,b]} u_n\,\mathrm{d}m,$$
$$\int_a^b \ell_n(x)\,\mathrm{d}x = \int_{[a,b]} \ell_n\,\mathrm{d}m.$$

Define $\ell = \sup_{n\in\mathbb{N}} \ell_n$ and $u = \inf_{n\in\mathbb{N}} u_n$. We know that both ℓ and u are measurable since they are pointwise limits of step functions, which are measurable. We can also see that $\ell \le f \le u$.

Consider the sequence of functions $(2u_1 - u_n + \ell_n)_{n=1}^{\infty}$. Observe that each of these functions is non-negative, and this is a non-decreasing sequence. We have $2u_1 - u_n + \ell_n \nearrow 2u_1 - u + \ell$, and applying Lebesgue's Monotone Convergence Theorem gives us

$$\lim_{n\to\infty} \int_{[a,b]} (2u_1 - u_n + \ell_n)\,\mathrm{d}m = \int_{[a,b]} (2u_1 - u + \ell)\,\mathrm{d}m. \tag{2.3}$$

The left hand side of (2.3) is equal to

$$\lim_{n\to\infty} \int_{[a,b]} (2u_1 - u_n + \ell_n)\,\mathrm{d}m = \int_{[a,b]} 2u_1\,\mathrm{d}m + \lim_{n\to\infty} \int_{[a,b]} (u_n - \ell_n)\,\mathrm{d}m = \int_{[a,b]} 2u_1\,\mathrm{d}m$$

where the limit above went to 0 since the Lebesgue integral coincides with the Riemann integral for the simple step functions and we know that f is Riemann integrable. The right hand side of (2.3) is

$$\int_{[a,b]} (2u_1 - u + \ell)\,\mathrm{d}m = \int_{[a,b]} 2u_1\,\mathrm{d}m + \int_{[a,b]} (u - \ell)\,\mathrm{d}m.$$

Putting everything together, we know that $u - \ell \ge 0$ and

$$\int_{[a,b]} (u - \ell)\,\mathrm{d}m = 0.$$

Then we must have $u = \ell = f$ almost everywhere, and so f is measurable. Since f is Riemann integrable, it is also bounded, so $f \in L^1([a, b])$. We conclude that

$$\int_{[a,b]} f\,\mathrm{d}m = \int_{[a,b]} u\,\mathrm{d}m = \lim_{n\to\infty} u_n\,\mathrm{d}m = \int_a^b f(x)\,\mathrm{d}x. \qquad \square$$

3 Banach and Hilbert Spaces

3.1 Banach and Hilbert Spaces

Functional analysis is the study of normed vector spaces and the continuous linear maps between them. Some of the most important examples of complete metric spaces are Banach spaces. Recall from PMATH 351 that a metric space is complete if every Cauchy sequence is convergent. We first give the definition of a normed vector space and a Banach space, and then illustrate these concepts with some examples.

Definition 3.1

A **normed vector space** is a pair $(V, \|\cdot\|)$ where V is a vector space over $\mathbb{C}$ and $\|\cdot\| : V \to [0, \infty)$ is a **norm** on V. That is, $\|\cdot\|$ satisfies the following properties:

(1) $\|v\| \geq 0$ for all $v \in V$, and $\|v\| = 0$ if and only if $v = 0$;

(2) $\|\alpha v\| = |\alpha|\|v\|$ for all $\alpha \in \mathbb{C}$ and $v \in V$; and

(3) $\|v + w\| \leq \|v\| + \|w\|$ for all $v, w \in V$.

Note that the property that $\|v\| = 0$ if and only if $v = 0$ says that $\|\cdot\|$ is non-degenerate; without this requirement, $\|\cdot\|$ is called a **semi-norm**. Moreover, given $v, w \in V$, the above properties together give rise to a canonical metric defined by

$$d(v, w) = \|v - w\|.$$

Thus, a normed vector space is a special case of a metric space.

Definition 3.2

A **Banach space** is a complete normed vector space.

We begin with the prototypical example of a normed vector space.

Example 3.3

Let $v = (v_1, v_2, \ldots, v_n) \in \mathbb{C}^n$. We define the 1-norm on $\mathbb{C}^n$ by

$$\|v\|_1 = |v_1| + |v_2| + \cdots + |v_n|.$$

The 2-norm on $\mathbb{C}^n$ is given by

$$\|v\|_2 = (|v_1|^2 + |v_2|^2 + \cdots + |v_n|^2)^{1/2}.$$

More generally, for $1 \leq p < \infty$, the p-norm on $\mathbb{C}^n$ is

$$\|v\|_p = (|v_1|^p + |v_2|^p + \cdots + |v_n|^p)^{1/p}.$$

The ∞-norm or supremum norm on $\mathbb{C}^n$ is

$$\|v\|_\infty = \max\{|v_1|, |v_2|, \ldots, |v_n|\}.$$

In fact, $\mathbb{C}^n$ equipped with each of these norms is a finite-dimensional Banach space.

We can also consider the continuous analogues for the space $C[a,b]$ of continuous functions on $[a,b]$.

Example 3.4

Let $f \in C[a,b]$. The 1-norm on $C[a,b]$ is

$$\|f\|_1 = \int_a^b |f(x)| \, \mathrm{d}x.$$

The 2-norm on $C[a,b]$ is given by

$$\|f\|_2 = \left(\int_a^b |f(x)|^2 \, \mathrm{d}x \right)^{1/2}.$$

Finally, the ∞-norm on $C[a,b]$ is

$$\|f\|_\infty = \sup_{x \in [a,b]} |f(x)|.$$

Note that $C[a,b]$ is a Banach space equipped with the supremum norm, but it is not complete with respect to the 1-norm and 2-norm.

Next, we consider inner product spaces and Hilbert spaces. The latter is a special case of a Banach space.

Definition 3.5

An **inner product space** is a pair $(H, \langle \cdot, \cdot \rangle)$ where H is a vector space over $\mathbb{C}$ and $\langle \cdot, \cdot \rangle : H \times H \to \mathbb{C}$ satisfies the following properties.

(1) $\langle x, y \rangle = \overline{\langle y, x \rangle}$ for all $x, y \in H$.

(2) For fixed $y \in H$, the map $x \mapsto \langle x, y \rangle$ is a linear functional.

(3) **Positive semidefiniteness:** $\langle x, x \rangle \geq 0$ for all $x \in H$.

(4) **Non-degeneracy:** $\langle x, x \rangle = 0$ if and only if $x = 0$.

Note that properties (1) and (2) imply that $\langle \cdot, \cdot \rangle$ is a **sesquilinear form** on H. That is, $\langle \cdot, \cdot \rangle$ is linear in the left variable, and conjugately linear in the right variable. Moreover, we can define a norm on H by

$$\|x\| = \langle x, x \rangle^{1/2}.$$

We verify this fact in the following proposition.

Proposition 3.6

Let $(H, \langle \cdot, \cdot \rangle)$ be an inner product space.

(1) **Cauchy-Schwarz inequality:** For all $x, y \in H$, we have $|\langle x, y \rangle| \leq \|x\| \|y\|$.

(2) $(H, \langle \cdot, \cdot \rangle)$ is a normed vector space.

Proof.

(1) Fix $x, y \in H$. The result is clear if $\|x\| = 0$ or $\|y\| = 0$. Assume now that $x \neq 0$ and $y \neq 0$. For $t \in \mathbb{R}$, define $p(t) = \|x - ty\|^2 \geq 0$. Observe that

$$p(t) = \langle x - ty, x - ty \rangle = \|x\|^2 + t^2 \|y\|^2 + 2t \operatorname{Re}\langle x, y \rangle.$$

By setting $a = \|y\|^2$, $b = 2\operatorname{Re}\langle x, y\rangle$, and $c = \|x\|^2$, we have $p(t) = at^2 + bt + c \geq 0$. This means that $p(t)$ has at most one distinct real root, and thus $b^2 - 4ac \leq 0$. In particular, we have

$$4(\operatorname{Re}\langle x, y\rangle)^2 \leq 4\|x\|^2\|y\|^2,$$

or equivalently, $|\operatorname{Re}\langle x, y\rangle| \leq \|x\|\|y\|$. Choose $\alpha \in \mathbb{C}$ such that $|\alpha| = 1$ and $\langle \alpha x, y\rangle = |\langle x, y\rangle|$. Then we deduce that

$$|\langle x, y\rangle| = \langle \alpha x, y\rangle = |\operatorname{Re}\langle \alpha x, y\rangle| \leq \|\alpha x\|\|y\| = \|x\|\|y\|.$$

(2) We will show that $\|\cdot\|$ satisfies the triangle inequality; the other properties are obvious. Given $x, y \in H$, we have

$$\|x + y\|^2 = \|x\|^2 + \|y\|^2 + 2\operatorname{Re}\langle x, y\rangle \leq \|x\|^2 + \|y\|^2 + 2\|x\|\|y\| = (\|x\| + \|y\|)^2. \qquad \square$$

Since we have shown that $\|x\| = \langle x, x\rangle^{1/2}$ is indeed a norm on an inner product space $(H, \langle \cdot, \cdot \rangle)$, it makes sense to make the following definition.

Definition 3.7

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

3.2 Banach Spaces of Measurable Functions

Recall that for $E \in \mathcal{L}$, we defined

$$L^1(E) = \left\{ f : E \to \mathbb{C} : f \text{ is measurable and } \int_E |f| \,\mathrm{d}m < \infty \right\},$$

and we showed in Theorem 2.57 that $L^1(E)$ is a vector space over $\mathbb{C}$. Given $f \in L^1(E)$, we will define the **L^1-norm** of f by the value of the integral above, so

$$\|f\|_1 = \int_E |f| \,\mathrm{d}m.$$

Note that $\|\cdot\|_1$ is close to being a norm, since it is easy to see that $\|\alpha f\|_1 = |\alpha|\|f\|_1$ and

$$\|f + g\|_1 = \int_E |f + g| \,\mathrm{d}m \leq \int_E (|f| + |g|) \,\mathrm{d}m = \|f\|_1 + \|g\|_1$$

for any $f, g \in L^1(E)$ and $\alpha \in \mathbb{C}$. However, we are not quite there. Notice that

$$\|f\|_1 = \int_E |f| \,\mathrm{d}m = 0$$

only implies that $f = 0$ almost everywhere, so we are not guaranteed non-degeneracy.

To repair this, we can define an equivalence relation on $L^1(E)$ by $f \sim g$ if and only if $f - g = 0$ almost everywhere. Then $L^1(E)/\!\sim$ is a quotient space where $f \in L^1(E)$ is associated with the equivalence class $[f] \in L^1(E)/\!\sim$. It can be verified that $L^1(E)/\!\sim$ is a vector space over $\mathbb{C}$ with $\alpha[f] + \beta[g] = [\alpha f + \beta g]$ for all $\alpha, \beta \in \mathbb{C}$ and $f, g \in L^1(E)$. Moreover, we can define a norm on it by $\|[f]\|_1 = \|f\|_1$.

From now on, we will abuse notation and identify $L^1(E)$ with $L^1(E)/\!\sim$. That is, we only identify functions $f \in L^1(E)$ up to almost everywhere equality. Then $(L^1(E), \|\cdot\|_1)$ is a normed vector space.

Definition 3.8

Let $E \in \mathcal{L}$. We define the set of **square-integrable functions** over E by

$$L^2(E) = \left\{ f : E \to \mathbb{C} : f \text{ is measurable and } \int_E |f|^2 \, \mathrm{d}m < \infty \right\}.$$

Equivalently, we have

$$L^2(E) = \{ f : E \to \mathbb{C} \mid f \text{ is measurable and } f^2 \in L^1(E) \}.$$

Proposition 3.9

Let $E \in \mathcal{L}$ and $f, g \in L^2(E)$. Then $L^2(E)$ is an inner product space with inner product given by

$$\langle f, g \rangle = \int_E f \overline{g} \, \mathrm{d}m.$$

In particular, this induces a norm on $L^2(E)$ via

$$\|f\|_2 = \langle f, f \rangle^{1/2} = \left(\int_E |f|^2 \, \mathrm{d}m \right)^{1/2},$$

which we call the L^2**-norm**.

Proof. Let $f, g \in L^2(E)$ and $\alpha \in \mathbb{C}$. It is clear that $\alpha f \in L^2(E)$. To see that $f + g \in L^2(E)$, note that $|f(x) + g(x)| \leq 2 \max\{|f(x)|, |g(x)|\}$, and thus

$$|f(x) + g(x)|^2 \leq 4 \max\{|f(x)|^2, |g(x)|^2\} \leq 4(|f(x)|^2 + |g(x)|^2).$$

It follows that

$$\int_E |f + g|^2 \, \mathrm{d}m \leq \int_E 4(|f|^2 + |g|^2) \, \mathrm{d}m < \infty,$$

so $f + g \in L^2(E)$ and we conclude that $L^2(E)$ is a $\mathbb{C}$-vector space.

Next, we verify that $\langle f, g \rangle$ is well-defined by showing that $f\overline{g} \in L^1(E)$ for any $f, g \in L^2(E)$. Note that if $a, b \geq 0$, then $ab \leq \frac{1}{2}(a^2 + b^2)$. By taking $a = |f(x)|$ and $b = |g(x)|$, we obtain

$$|f(x)\overline{g}(x)| = |f(x)||g(x)| \leq \frac{1}{2}(|f(x)|^2 + |g(x)|^2).$$

Hence, we see that

$$\int_E f\overline{g} \, \mathrm{d}m \leq \frac{1}{2} \int_E |f|^2 \, \mathrm{d}m + \frac{1}{2} \int_E |g|^2 \, \mathrm{d}m < \infty.$$

It is easily checked that $\langle \alpha f + \beta g, h \rangle = \alpha \langle f, h \rangle + \beta \langle g, h \rangle$ and $\langle f, g \rangle = \overline{\langle g, f \rangle}$ for any $f, g, h \in L^2(E)$ and $\alpha, \beta \in \mathbb{C}$, so $\langle \cdot, \cdot \rangle$ is a sesquilinear form on $L^2(E)$. Finally, we have

$$\langle f, f \rangle = \int_E |f|^2 \, \mathrm{d}m \geq 0$$

and $\langle f, f \rangle = 0$ if and only if $|f|^2 = 0$ almost everywhere, which is equivalent to saying that $f = 0$ almost everywhere. By identifying $L^2(E)$ with $L^2(E)/\sim$ where $\sim$ is the same equivalence relation as above, this is enough to show that $(L^2(E), \langle \cdot, \cdot \rangle)$ is an inner product space. $\square$

Remark 3.10

Let $1 \leq p < \infty$ and $E \in \mathcal{L}$. Analogous to what we have done so far, we can define

$$L^p(E) = \left\{ f : E \to \mathbb{C} : f \text{ is measurable and } \int_E |f|^p \, \mathrm{d}m < \infty \right\}.$$

and define the L^p**-norm** by

$$\|f\|_p = \left(\int_E |f|^p \, \mathrm{d}m \right)^{1/p}.$$

It can be verified that this is indeed a norm by using Hölder's inequality. This makes $(L^p(E), \|\cdot\|_p)$ a normed vector space.

Let $1 \leq p < \infty$. We now show that $L^p(E)$ is complete with respect to the norm $\|\cdot\|_p$ defined as above. In particular, this means that $(L^1(E), \|\cdot\|_1)$ is a Banach space and $(L^2(E), \langle \cdot, \cdot \rangle)$ is a Hilbert space. We recall that if a Cauchy sequence has a convergent subsequence in a metric space, then the Cauchy sequence also converges.

Theorem 3.11

Let $1 \leq p < \infty$ and $E \in \mathcal{L}$. Then $(L^p(E), \|\cdot\|_p)$ is complete.

Proof. Fix $1 \leq p < \infty$, and take a Cauchy sequence $(f_n)_{n=1}^\infty$ in $L^p(E)$ with respect to $\|\cdot\|_p$. Choose a subsequence $n_1 < n_2 < n_3 < \cdots$ such that

$$\|f_{n_{k+1}} - f_{n_k}\|_p < 2^{-k}$$

for all $k \in \mathbb{N}$. By the above remark, it suffices to find a function $f \in L^p(E)$ such that $\|f - f_{n_k}\|_p \to 0$ as $k \to \infty$. We define

$$g_m(x) = |f_{n_1}(x)| + \sum_{k=1}^m |f_{n_{k+1}}(x) - f_{n_k}(x)|.$$

Then we see that each g_m is measurable with $0 \leq g_1 \leq g_2 \leq \cdots$ and

$$\|g_m\|_p \leq \|f_{n_1}\|_p + \sum_{k=1}^m \|f_{n_{k+1}} - f_{n_k}\|_p \leq \|f_{n_1}\|_p + 1 = M.$$

Thus, we have $0 \leq g_1^p \leq g_2^p \leq \cdots$ and

$$\int_E g_m^p \, \mathrm{d}m \leq M^p$$

for all $m \in \mathbb{N}$. Now, define

$$g(x) = \lim_{m\to\infty} g_m(x) = |f_{n_1}(x)| + \sum_{k=1}^\infty |f_{n_{k+1}}(x) - f_{n_k}(x)|,$$

and observe that g is measurable. By Lebesgue's Monotone Convergence Theorem (Theorem 2.48), we have

$$\int_E g^p \, \mathrm{d}m = \lim_{m\to\infty} \int_E g_m^p \, \mathrm{d}m \leq M^p < \infty.$$

Thus, we have $g \in L^p(E)$ and in particular, $g(x) < \infty$ almost everywhere. If $g(x) < \infty$, then we know that the telescoping series

$$f_{n_1}(x) + \sum_{k=1}^\infty (f_{n_{k+1}}(x) - f_{n_k}(x))$$

converges absolutely. This means that $\lim_{m\to\infty} f_{n_m}(x)$ exists for almost every $x \in E$. Finally, define

$$f(x) = \begin{cases} \lim_{m\to\infty} f_{n_m}(x), & \text{if } g(x) < \infty, \\ 0, & \text{otherwise.} \end{cases}$$

Then $f_{n_k} \to f$ almost everywhere, so f is measurable. Moreover, $|f_{n_k}| \leq g_{k-1} \leq g$ implies that $|f| \leq g$. This means that

$$\int_E |f|^p \,\mathrm{d}m \leq \int_E g^p \,\mathrm{d}m < \infty,$$

so $f \in L^p(E)$. Finally, observe that

$$\|f_{n_m} - f\|_p = \left\| \sum_{k=m}^{\infty} (f_{n_{k+1}} - f_{n_k}) \right\|_p \leq \sum_{k=m}^{\infty} \|f_{n_{k+1}} - f_{n_k}\|_p \to 0.$$

Then applying Lebesgue's Dominated Convergence Theorem to $f_{n_{m+1}}^p \to f^p$ with $|f_{n_{m+1}}|^p \leq g^p \in L^1(E)$ implies that $f^p \in L^1(E)$, which is equivalent to saying that $f \in L^p(E)$. This completes the proof. $\square$

Corollary 3.12

Let $1 \leq p < \infty$ and let $(f_n)_{n=1}^\infty \subseteq L^p(E)$ be a Cauchy sequence with limit $f \in L^p(E)$. Then there exists a subsequence $(f_{n_k})_{k=1}^\infty$ such that $f_{n_k}(x) \to f(x)$ for almost every $x \in E$.

Proof. Take the subsequence $(f_{n_k})_{k=1}^\infty$ from the proof of the previous theorem. $\square$

Remark 3.13

We know that if $f_n \to f$ with respect to $\|\cdot\|_p$ for $1 \leq p < \infty$, then $f_{n_k}(x) \to f(x)$ for almost every $x \in E$ for some subsequence $(f_{n_k})_{k=1}^\infty$. But does this imply that $f_n \to f$ for the original sequence?

The answer is no. Define $(f_n)_{n=1}^\infty$ by $f_1 = \chi_{[0,1]}$, $f_2 = \chi_{[0,1/2]}$, $f_3 = \chi_{[1/2,1]}$, $f_4 = \chi_{[0,1/3]}$, $f_5 = \chi_{[1/3,2/3]}$, $f_6 = \chi_{[2/3,1]}$, and so on. Then $\|f_n\|_p \to 0$ for $1 \leq p < \infty$, but f_n does not converge for any $x \in [0,1]$. Therefore, we really need a subsequence $(f_{n_k})_{k=1}^\infty$ in general.

Let $E \in \mathcal{L}$ with $m(E) > 0$. We now consider the space $L^\infty(E)$.

Definition 3.14

A measurable function $f : E \to \mathbb{C}$ is called **essentially bounded** if there exists $A > 0$ such that $|f| \leq A$ almost everywhere. That is, we have $m\{x \in E : |f(x)| > A\} = 0$.

When $f : E \to \mathbb{C}$ is essentially bounded, the **essential supremum** of $|f|$ is defined to be

$$\|f\|_\infty = \operatorname*{ess\,sup}_{x\in E} |f(x)| = \inf\{A \geq 0 \mid m\{x \in E : |f(x)| > A\} = 0\}.$$

Then the space $L^\infty(E)$ is given by

$$L^\infty(E) = \{f : E \to \mathbb{C} \mid f \text{ is measurable and } \|f\|_\infty < \infty\}.$$

Similar to $(L^p(E), \|\cdot\|_p)$ for $1 \leq p < \infty$, we have that $L^\infty(E)$ equipped with the essential supremum is a normed vector space. In fact, it is a Banach space.

Proposition 3.15

$(L^\infty(E), \|\cdot\|_\infty)$ is a normed vector space.

Proof. Let $\alpha \in \mathbb{C}$ and $f, g \in L^\infty(E)$. Note that $|f| > A$ if and only if $|\alpha f| \geq |\alpha| A$ for all $\alpha \in \mathbb{C} \setminus \{0\}$, so $\|\alpha f\|_\infty = |\alpha| \|f\|_\infty$ and thus $\alpha f \in L^\infty(E)$. To see that $f + g \in L^\infty(E)$, there exist $A, B > 0$ such that $|f(x)| \leq A$ and $|g(x)| \leq B$ for almost every $x \in E$. Then for almost every $x \in E$, we get

$$|f(x) + g(x)| \leq |f(x)| + |g(x)| \leq A + B < \infty,$$

which implies that $f + g \in L^\infty(E)$.

Now, we check that $\|\cdot\|_\infty$ is a norm. It is clear that $\|f\|_\infty \geq 0$. We claim that $|f| \leq \|f\|_\infty$ and $|g| \leq \|g\|_\infty$ almost everywhere. Indeed, observe that

$$m\{x \in E : |f(x)| > \|f\|_\infty + 1/n\} = 0$$

by the definition of $\|f\|_\infty$, and thus

$$m\{x \in E : |f(x)| > \|f\|_\infty\} = m\left(\bigcup_{n=1}^{\infty} \{x \in E : |f(x)| > \|f\|_\infty + 1/n\}\right) = 0.$$

In particular, if $\|f\|_\infty = 0$, then $|f| = 0$ almost everywhere and hence $f = 0$ almost everywhere. Moreover, we have

$$|f + g| \leq |f| + |g| \leq \|f\|_\infty + \|g\|_\infty$$

almost everywhere, which implies that

$$m\{x \in E : |f(x) + g(x)| > \|f\|_\infty + \|g\|_\infty\} = 0.$$

This means that $\|f + g\|_\infty \leq \|f\|_\infty + \|g\|_\infty$, so $\|\cdot\|_\infty$ is a norm as desired. □

Theorem 3.16

$(L^\infty(E), \|\cdot\|_\infty)$ is a Banach space.

Proof. Let $(f_n)_{n=1}^\infty$ be a Cauchy sequence in $(L^\infty(E), \|\cdot\|_\infty)$. Pick a subsequence $(f_{n_k})_{k=1}^\infty$ such that

$$\sum_{k=1}^{\infty} \|f_{n_{k+1}} - f_{n_k}\|_\infty < \infty.$$

Then we have

$$g(x) = |f_{n_1}(x)| + \sum_{k=1}^{\infty} |f_{n_{k+1}}(x) - f_{n_k}(x)| < \infty$$

for almost every $x \in E$. This means that

$$\lim_{m\to\infty} f_{n_m}(x) = \lim_{m\to\infty} \left(f_{n_1}(x) + \sum_{k=1}^{m-1} (f_{n_{k+1}}(x) - f_{n_k}(x)) \right)$$

exists for almost every $x \in E$. Define $f : E \to \mathbb{C}$ by

$$f(x) = \begin{cases} \lim_{m\to\infty} f_{n_m}(x), & \text{if it converges,} \\ 0, & \text{otherwise.} \end{cases}$$

Then we have $\|f - f_{n_m}\|_\infty \to 0$ because for $x \in \{y \in E | \lim_{m\to\infty} f_{n_m}(y) = f(y)\} =: A$, we get

$$|f_{n_m}(x) - f(x)| \leq \sum_{k=m}^{\infty} |f_{n_{k+1}}(x) - f_{n_k}(x)|,$$

which converges uniformly for $x \in A$. Since $m(E \setminus A) = 0$, this gives as $\|f - f_{n_m}\|_\infty \to 0$ as claimed. Finally, we obtain

$$0 \leq \lim_{n\to\infty} \|f_n - f\|_\infty \leq \lim_{n\to\infty} (\|f_n - f_{n_m}\|_\infty + \|f_{n_m} - f\|_\infty) = 0.$$ □

Remark 3.17

If $m(E) < \infty$, then $L^\infty(E) \subseteq L^2(E) \subseteq L^1(E)$. Indeed, if $f \in L^\infty(E)$, then

$$\|f\|_2^2 = \int_E |f|^2 \, \mathrm{d}m \leq \|f\|_\infty^2 \int_E 1 \, \mathrm{d}m = \|f\|_\infty^2 m(E),$$

and similarly, we have $\|f\|_1 \leq \|f\|_\infty m(E)$.

This result is not true without the assumption that $m(E) < \infty$. For example, let $E = \mathbb{R}$, and define $f(x) = 1$ for all $x \in \mathbb{R}$, $g(x) = 1/x$ for $x \geq 1$ and $g(x) = 0$ otherwise, and $h(x) = x^{-1/2}$ on $(0,1)$ and $h(x) = 0$ otherwise. Then $f \in L^\infty(\mathbb{R}) \setminus (L^1(\mathbb{R}) \cup L^2(\mathbb{R}))$, $g \in L^2(\mathbb{R}) \setminus L^1(\mathbb{R})$, and $h \in L^1(\mathbb{R}) \setminus (L^2(\mathbb{R}) \cup L^\infty(\mathbb{R}))$.

3.3 Density and Approximation Results

We now consider the density of certain spaces in $L^p(\mathbb{R})$, where $1 \leq p \leq \infty$.

Definition 3.18

The space of compactly supported continuous functions is given by

$$C_c(\mathbb{R}) = \{f : \mathbb{R} \to \mathbb{C} \mid f \text{ is continuous and there exists a compact subset } K \subseteq \mathbb{R} \text{ with } f|_{\mathbb{R}\setminus K} = 0\}.$$

The space of simple and measurable functions is given by

$$S(\mathbb{R}) = \{f : \mathbb{R} \to \mathbb{C} \mid f \text{ is simple and measurable}\}.$$

The space of simple and measurable functions which are also Lebesgue integrable is defined by

$$S_1(\mathbb{R}) = S(\mathbb{R}) \cap L^1(\mathbb{R}) = \left\{ \varphi = \sum_{i=1}^{n} \alpha_i \chi_{E_i} : \alpha_i \in \mathbb{C},\, m(E_i) < \infty \right\}.$$

Note that for $1 \leq p \leq \infty$, we have $C_c(\mathbb{R}) \subseteq L^p(\mathbb{R})$ and $S_1(\mathbb{R}) \subseteq L^p(\mathbb{R})$ as subspaces. Moreover, we have $S(\mathbb{R}) \subseteq L^\infty(\mathbb{R})$ as a subspace. None of these are closed subspaces though.

Theorem 3.19

$S(\mathbb{R})$ is dense in $L^p(\mathbb{R})$ for $1 \leq p \leq \infty$. That is, for $f \in L^p(\mathbb{R})$, there exists a sequence $(\varphi_n)_{n=1}^\infty$ of simple measurable functions in $L^p(\mathbb{R})$ such that $\varphi_n \to f$ with respect to $\|\cdot\|_p$.

PROOF. Let $f \in L^p(\mathbb{R})$. Since f is measurable, we know from Theorem 2.43 that we can approximate f using simple measurable functions. In particular, there exists a sequence $(\varphi_n)_{n=1}^{\infty}$ of simple measurable functions such that $\varphi_n \to f$ pointwise and

$$|\varphi_1| \leq |\varphi_2| \leq \cdots \leq |f|.$$

When $p = \infty$, we have $\varphi_n \to f$ almost uniformly. This means that $\|\varphi_n - f\|_\infty \to 0$. On the other hand, if $1 \leq p < \infty$, then we have $|f - \varphi_n|^p \to 0$ pointwise, and we know that

$$|f - \varphi_n|^p \leq (2|f|)^p \in L^1(\mathbb{R})$$

since $f \in L^p(\mathbb{R})$. This means that $\varphi_n \in L^p(\mathbb{R})$, and applying Lebesgue's Dominated Convergence Theorem (Theorem 2.58) implies that

$$\|f - \varphi_n\|_p^p = \int_{\mathbb{R}} |f - \varphi_n|^p \, \mathrm{d}m \to 0. \qquad \square$$

THEOREM 3.20

$C_c(\mathbb{R})$ is dense in $L^p(\mathbb{R})$ for $1 \leq p < \infty$.

PROOF. Let $\varepsilon > 0$ and $f \in L^p(\mathbb{R})$ where $1 \leq p < \infty$. We need to find $g \in C_c(\mathbb{R})$ such that $\|f - g\|_p < \varepsilon$.

Since $L_p(\mathbb{R})$ is the closure of the simple integrable functions, it suffices to assume that $f = \varphi$ is simple. Indeed, if we can find a simple function φ such that $\|f - \varphi\|_p < \varepsilon/2$ and a function $g \in C_c(\mathbb{R})$ such that $\|\varphi - g\|_p < \varepsilon/2$, then the triangle inequality gives us the result. So we have

$$\varphi = \sum_{i=1}^{n} \alpha_i \chi_{E_i}$$

where $\alpha \in \mathbb{C}$ and $E_i \in \mathcal{L}$ with $m(E_i) < \infty$.

We can further reduce this problem to approximating $\varphi = \chi_E$ for $E \in \mathcal{L}$ with $m(E) < \infty$, because the triangle inequality again assures us that

$$\left\| \sum_{i=1}^{n} \alpha_i \chi_{E_i} - \sum_{i=1}^{n} \alpha_i g_i \right\|_p \leq \sum_{i=1}^{n} |\alpha_i| \|\chi_{E_i} - g_i\|_p$$

for some $g_i \in C_c(\mathbb{R})$.

Finally, fix $E \in \mathcal{L}$ with $m(E) < \infty$. We can write $E = \bigcup_{n\in\mathbb{N}} E \cap (-n, n)$, and observe that $m(E) = \lim_{n\to\infty} m(E_n)$ by the continuity of measure. We leave it as an exercise to show that

$$\chi_E = \lim_{n\to\infty} \chi_{E_n}$$

with respect to $\|\cdot\|_p$. So without loss of generality, we can assume that $E \in \mathcal{L}$ is bounded, say $E \subseteq (-n, n)$ for some $n \in \mathbb{N}$. We will find $g \in C_c(\mathbb{R})$ such that $\|\chi_E - g\|$ is bounded above by ε multiplied by a constant.

Recall from the structure theorem of Lebesgue measurable sets (Theorem 2.22) that there exists a closed set F and an open set G such that $F \subseteq E \subseteq G$ and $m(G \setminus F) < \varepsilon$. We may assume without loss of generality that $G \subseteq (-n, n)$ because we can chop off elements outside of this interval otherwise.

We now design a $g \in C_c(\mathbb{R})$ such that

(i) $g(x) \in [0, 1]$ for all $x \in \mathbb{R}$;

(ii) $g(x) = 1$ for all $x \in F$; and

(iii) $g(x) = 0$ for all $x \in G^c$.

Given such a function $g \in C_c(\mathbb{R})$, we are done because

$$\|\chi_E - g\|_p^p = \int_G |\chi_E - g|^p \, \mathrm{d}m = \int_F 0 \, \mathrm{d}m + \int_{G \setminus F} |\chi_E - g|^p \, \mathrm{d}m \le 0 + m(G \setminus F) 2^p < 2^p \varepsilon.$$

To construct this function, recall that for $S \subseteq \mathbb{R}$ with $S \neq \varnothing$, we define

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

for any fixed $x \in \mathbb{R}$. We leave it as an exercise to show that $d(\cdot, S) : \mathbb{R} \to [0, \infty)$ is continuous. Then define

$$g(x) = \frac{d(x, G^c)}{d(x, G^c) + d(x, F)}.$$

This is well-defined since $F \cap G^c = \varnothing$ and so $d(x, G^c) + d(x, F) \neq 0$. Moreover, it is continuous because it is the composition of continuous functions. Finally, it is easily seen that g satisfies all three properties above, which completes the proof. □

We note that $C_c(\mathbb{R})$ is *not* dense in $L^\infty(\mathbb{R})$; this can be seen by taking the function $f(x) = 1$ for all $x \in \mathbb{R}$.

Corollary 3.21

Let $a < b$ be finite. Then $C[a, b]$ is dense in $L^p[a, b]$ for $1 \le p \le \infty$.

Proof. Repeat the same argument as Theorem 3.20 by reducing it to the case where $f = \chi_E$ with $E \subseteq (a + 1/n, b - 1/n)$. The case where $p = \infty$ is not an issue here because we have $L^1[a, b] \supseteq L^p[a, b] \supseteq L^\infty[a, b] \supseteq C[a, b]$, and being dense in $L^1[a, b]$ implies being dense in $L^\infty[a, b]$ because $L^1[a, b]$ is a larger space. □

4 Fourier Analysis on the Circle

4.1 General Problem of Fourier Analysis

Recall that we defined $\mathbb{T} = \{z \in \mathbb{C} : |z| = 1\}$, which we can identify with the interval $[-\pi, \pi]$ via the map $\theta \mapsto e^{i\theta}$. We also identify the endpoints so that $\pi = -\pi$. We are interested in studying $(L^p(\mathbb{T}), \|\cdot\|_p)$, where

$$\|f\|_p = \left(\frac{1}{2\pi}\int_{-\pi}^{\pi} |f(\theta)|^p \,\mathrm{d}\theta\right)^{1/p}$$

for $1 \leq p < \infty$, and $(L^\infty(\mathbb{T}), \|\cdot\|_\infty)$ is defined as usual. Note that we have introduced a new normalization here, which is merely for cosmetic purposes to ensure that the identity function has norm 1.

EXERCISE 4.1

Show that for all $p \in (1, \infty)$, we have

$$C(\mathbb{T}) \subseteq L^\infty(\mathbb{T}) \subseteq L^p(\mathbb{T}) \subseteq L^1(\mathbb{T}),$$

and that each of these inclusions are dense. Moreover, prove that

$$\|f\|_\infty \geq \|f\|_p \geq \|f\|_1.$$

Recall that at the beginning of the course in Section 1.3, we discussed Fourier coefficients of functions. We now write down a slightly more rigorous definition.

DEFINITION 4.2

Let $f \in L^1(\mathbb{T})$. For each $n \in \mathbb{Z}$, we define the n**-th Fourier coefficient** of f by

$$\hat{f}(n) = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(\theta)e^{-in\theta}\,\mathrm{d}\theta.$$

The **Fourier series** of f is the series $\sum_{n\in\mathbb{Z}} \hat{f}(n)e^{in\theta}$, and we write

$$f \sim \sum_{n\in\mathbb{Z}} \hat{f}(n)e^{in\theta}.$$

The general problem of Fourier analysis is to answer the following questions:

- Given $f \in L^p(\mathbb{T})$ or $f \in C(\mathbb{T})$, to what extent does the sequence $(\hat{f}(n))_{n\in\mathbb{Z}}$ determine f?
- Given a sequence $(a_n)_{n\in\mathbb{Z}}$, are there necessary or sufficient conditions in which there exists a (possibly unique) function $f \in L^p(\mathbb{T})$ such that $a_n = \hat{f}(n)$ for all $n \in \mathbb{Z}$?
- To what extent does the Fourier series represent f? Do we have

  $$f(\theta) = \sum_{n\in\mathbb{Z}} \hat{f}(n)e^{in\theta}$$

  either pointwise, pointwise almost everywhere, or convergence with respect to $\|\cdot\|_p$?

4.2 L^2-convergence of Fourier Series

It turns out that the theory for functions in $L^2(\mathbb{T})$ is the most "beautiful" and "clean". This is because $L^2(\mathbb{T})$ is a Hilbert space! In fact, most of the theory for $L^2(\mathbb{T})$ follows from general Hilbert space theory.

DEFINITION 4.3

Let $(H, \langle \cdot, \cdot \rangle)$ be a Hilbert space.

- Let $x, y \in H$. We say that x and y are **orthogonal** if $\langle x, y \rangle = 0$, and we write $x \perp y$. Note that this implies that $\|x+y\|^2 = \|x\|^2 + \|y\|^2$.

  Given a subset $S \subseteq H$, the **orthogonal complement** of S is defined to be
  $$S^\perp = \{x \in H \mid \langle x, y \rangle = 0 \text{ for all } y \in S\}.$$
  Note that $S^\perp$ is always a closed subspace.

- An **orthonormal system** in H is a family $\{e_n\}_{n \in S} \subseteq H$ such that for all $n, m \in S$, we have
  $$\langle e_n, e_m \rangle = \delta_{nm} = \begin{cases} 1, & \text{if } n = m, \\ 0, & \text{if } n \neq m. \end{cases}$$
  In particular, this means that $\|e_n\| = 1$ for all $n \in S$ and $e_n \perp e_m$ whenever $n \neq m$.

- An **orthonormal basis** in H is an orthonormal system $\{e_n\}_{n \in S}$ such that $\operatorname{span}\{e_n \mid n \in S\}$ is dense in H with respect to the norm $\|\cdot\|$ induced by the inner product.

EXAMPLE 4.4

Let $S \neq \varnothing$ be a countable set. Then the space of square-summable sequences, denoted
$$\ell^2(S) = \left\{ a = (a_n)_{n \in S} : a_n \in \mathbb{C}, \sum_{n \in S} |a_n|^2 < \infty \right\},$$
is a Hilbert space with the inner product
$$\langle a, b \rangle = \sum_{n \in S} a_n \overline{b_n}.$$
The norm induced by the inner product is given by
$$\|a\| = \langle a, a \rangle^{1/2} = \left(\sum_{n \in S} |a_n|^2 \right)^{1/2}.$$
For each $n \in S$, let δ_n be the sequence with 1 in the n-th entry and 0 in every other entry. Then it is easy to see that $(\delta_n)_{n \in S}$ is an orthonormal system in $\ell^2(S)$. In fact, its span is dense in $\ell^2(S)$, so it forms an orthonormal basis.

In this section, our interest is $(L^2(\mathbb{T}), \|\cdot\|_2)$. Note that the sequence of functions $\{e_n\}_{n \in \mathbb{Z}}$ defined by $e_n(\theta) = e^{in\theta}$ is an orthonormal system in $L^2(\mathbb{T})$. We will show later that this is in fact an orthonormal basis for $L^2(\mathbb{T})$. First, we will prove Parseval's Theorem. Recall that a topological space is called separable if it contains a countable dense subset.

Theorem 4.5: Parseval's Theorem

Let $(H, \langle\cdot,\cdot\rangle)$ be a separable Hilbert space. Let $\{e_n\}_{n\in S}$ be an orthonormal system in H, and let

$$K = \overline{\operatorname{span}\{e_n \mid n \in S\}}^{\|\cdot\|} \subseteq H.$$

For $n \in S$ and $x \in H$, set $\hat{x}(n) = \langle x, e_n\rangle$.

(1) For any $x \in H$, we have

$$\sum_{n\in S} |\hat{x}(n)|^2 \le \|x\|^2 < \infty.$$

(2) The map $P : H \to H$ defined by

$$Px = \sum_{n\in S} \hat{x}(n)e_n = \sum_{n\in S}\langle x, e_n\rangle e_n$$

satisfies the following properties:

(a) For any $P \in B(H,H)$, we have $\|P\| \le 1$, and $\|P\| = 1$ if $S \ne \varnothing$.

(b) We have $PH = K$ and $P^2 = P$ (so P is idempotent).

(c) For all $x \in H$, we have $(x - Px) \perp K$.

(d) We have $\ker(P) = K^\perp = \{x \in H \mid \langle x, k\rangle = 0 \text{ for all } k \in K\}$.

Note that properties (b) and (c) imply that P is the orthogonal projection of H onto K.

(3) For all $x \in H$, Px is the closest point in K to x. That is, we have

$$\|x - Px\| = \inf_{k\in K}\|x - k\|.$$

Proof. We first assume that $|S| < \infty$. In this case, we have $K = \operatorname{span}\{e_n \mid n \in S\}$ (meaning that there is no need to take the closure with respect to $\|\cdot\|$), and for all $x \in H$, the map $P : H \to H$ in the theorem is well-defined and linear. To see that $P^2 = P$, note that

$$P^2x = P(Px) = P\left(\sum_{n\in S}\hat{x}(n)e_n\right) = \sum_{n\in S}\hat{x}(n)e_n = Px$$

by pushing the map P to each e_n. Moreover, it is clear that $PH = K$.

Next, we see that

$$\|Px\|^2 = \langle Px, Px\rangle = \sum_{n\in S}|\hat{x}(n)|^2 = \sum_{n\in S}\langle x, e_n\rangle\overline{\langle x, e_n\rangle} = \left\langle x, \sum_{n\in S}\langle x, e_n\rangle e_n\right\rangle = \langle x, Px\rangle \le \|x\|\|Px\|,$$

where the final inequality follows from Cauchy-Schwarz. Rearranging this yields $\|Px\|(\|x\| - \|Px\|) \ge 0$, and hence $\|Px\| \le \|x\|$. This means that $\|P\| \le 1$, and $\|Pe_n\| = \|e_n\| = 1$ implies that $\|P\| = 1$ when $S \ne \varnothing$. Our work above also shows that

$$\sum_{n\in S}|\hat{x}(n)|^2 \le \|x\|\|Px\| \le \|x\|^2 < \infty.$$

For all $n_0 \in S$, notice that we have

$$\langle x - Px, e_{n_0}\rangle = \hat{x}(n_0) - \hat{x}(n_0) = 0,$$

so $(x - Px) \perp \operatorname{span}\{e_n \mid n \in S\} = K$ for all $x \in H$. Note that $x \in \ker(P)$ if and only if $\hat{x}(n) = \langle x, e_n\rangle = 0$ for all $n \in S$. This is equivalent to saying that $x \perp K$; that is, $x \in K^\perp$.

Finally, for any $k \in K$, we have

$$\|x-k\|^2 = \|(x-Px)+(Px-k)\|^2 = \|x-Px\|^2 + \|Px-k\|^2 \geq \|x-Px\|^2,$$

which implies that $\|x-Px\| = \inf_{k\in K}\|x-k\|$.

Suppose now that S is an infinite set. We leave it as an exercise to show that the separability of H implies that $|S| = |\mathbb{N}|$. Without loss of generality, we will assume that $S = \mathbb{N}$, so our orthonormal system is $\{e_n\}_{n=1}^{\infty}$. For each $N \in \mathbb{N}$, let $K_N = \operatorname{span}\{e_1, \ldots, e_n\} \subseteq K$. Let $P_N : H \to K_N$ be the corresponding projection

$$P_N x = \sum_{n=1}^{N} \hat{x}(n) e_n.$$

For all $N \in \mathbb{N}$, our above work shows that

$$\sum_{n=1}^{N} |\hat{x}(n)|^2 \leq \|x\|^2 < \infty,$$

and taking $N \to \infty$ gives us

$$\sum_{n=1}^{\infty} |\hat{x}(n)|^2 \leq \|x\|^2 < \infty.$$

For all $x \in H$, we claim that $(P_N x)_{N=1}^{\infty}$ is a Cauchy sequence. Taking $M < N$, we see that

$$\|P_n x - P_m x\|^2 = \left\| \sum_{n=M+1}^{N} \hat{x}(n) e_n \right\|^2 = \sum_{n=M+1}^{N} |\hat{x}(n)|^2,$$

which converges to 0 as $M, N \to \infty$. Then we see that

$$\lim_{N\to\infty} P_N x = \sum_{n=1}^{\infty} \hat{x}(n) e_n$$

exists in K. We define the map $P : H \to K$ by

$$Px = \lim_{N\to\infty} P_N x = \sum_{n=1}^{\infty} \hat{x}(n) e_n,$$

which is linear. Moreover, we have $\|Px\| = \lim_{N\to\infty} \|P_N x\| \leq \|x\|$, so P is bounded.

Note that $Px = x$ for all $x \in \operatorname{span}\{e_1, \ldots, e_N\}$ and $N \in \mathbb{N}$, so taking $N \to \infty$ gives $Px = x$ for all $x \in K$. Also, we have

$$\langle x - Px, e_n \rangle = \langle x, e_n \rangle - \langle x, e_n \rangle = 0$$

as before, so $(x - Px) \perp \overline{\operatorname{span}\{e_n \mid n \in \mathbb{N}\}}^{\|\cdot\|} = K$. Since $Px \in K$, this gives us $x \perp Px$, and thus

$$\|x\|^2 = \|x - Px\|^2 + \|Px\|^2.$$

The rest of the proof is the same as the $|S| < \infty$ case. □

Corollary 4.6

If $\{e_n\}_{n=1}^{\infty}$ is an orthonormal basis for a separable Hilbert space $(H, \langle\cdot,\cdot\rangle)$, then for all $x \in H$, we have

$$\|x\|^2 = \sum_{n\in S} |\hat{x}(n)|^2.$$

PROOF. Note that $(x - Px) \perp K$ for all $x \in H$ where P and K are as in Parseval's Theorem (Theorem 4.5). But $\{e_n\}_{n=1}^{\infty}$ is an orthonormal basis, which means that $H = K$. Then $x - Px = 0$ and hence $\|x - Px\| = 0$ for all $x \in H$. From this, we obtain

$$\|x\|^2 = \|Px\|^2 + \|x - Px\|^2 = \|Px\|^2 = \sum_{n \in S} |\hat{x}(n)|^2. \qquad \square$$

Now, we return to the case $(L^2(\mathbb{T}), \|\cdot\|_2)$ where

$$\|f\|_2^2 = \frac{1}{2\pi} \int_{-\pi}^{\pi} |f(\theta)|^2 \, \mathrm{d}\theta.$$

We had an orthonormal system $\{e_n\}_{n \in \mathbb{Z}}$ in $L^2(\mathbb{T})$ given by $e_n(\theta) = e^{in\theta}$. For $f \in L^2(\mathbb{T})$, we have

$$\hat{f}(n) = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(\theta) e^{-in\theta} \, \mathrm{d}\theta = \langle f, e_n \rangle.$$

We define the N-th partial sum of the Fourier series of f as

$$S_N f(\theta) = \sum_{n=-N}^{N} \hat{f}(n) e^{in\theta}.$$

In particular, denoting $\mathrm{Pol}(\mathbb{T}) = \mathrm{span}\{e_n \mid n \in \mathbb{Z}\}$ as the "trigonometric functions on the circle", we have $\mathrm{Pol}(\mathbb{T}) \subseteq C(\mathbb{T}) \subseteq L^2(\mathbb{T})$ and $S_N f \in \mathrm{Pol}(\mathbb{T})$ for all $N \in \mathbb{N}$. From Parseval's Theorem, we see that

$$S_N : L^2(\mathbb{T}) \to \mathrm{span}\{e_n \mid n \in \mathbb{Z},\, |n| \leq N\} \subseteq \mathrm{Pol}(\mathbb{T}) \subseteq L^2(\mathbb{T})$$

is the orthogonal projection of $L^2(\mathbb{T})$ onto $\mathrm{span}\{e_n \mid n \in \mathbb{Z},\, |n| \leq N\}$.

THEOREM 4.7

The family $\{e_n\}_{n \in \mathbb{Z}}$ is an orthonormal basis for $L^2(\mathbb{T})$. In particular, if $f \in L^2(\mathbb{T})$, then f is the $\|\cdot\|_2$-limit of its partial sums $(S_N f)_{N=1}^{\infty}$. That is, with respect to $\|\cdot\|_2$, we have

$$f = \sum_{n \in \mathbb{Z}} \hat{f}(n) e_n = \lim_{N \to \infty} S_N f.$$

PROOF. We already know that $\{e_n\}_{n \in \mathbb{Z}}$ is an orthonormal system for $L^2(\mathbb{T})$. Consider the set

$$K = \overline{\mathrm{span}\{e_n \mid n \in \mathbb{Z}\}}^{\|\cdot\|_2} = \overline{\mathrm{Pol}(\mathbb{T})}^{\|\cdot\|_2} \subseteq L^2(\mathbb{T}).$$

To show that $\{e_n\}_{n \in \mathbb{Z}}$ is an orthonormal basis for $L^2(\mathbb{T})$, our aim is to show that $K = L^2(\mathbb{T})$.

It suffices to prove that $\mathrm{Pol}(\mathbb{T}) \subseteq C(\mathbb{T})$ is dense in $C(\mathbb{T})$ with respect to $\|\cdot\|_\infty$. Indeed, suppose that $\mathrm{Pol}(\mathbb{T})$ is dense in $C(\mathbb{T})$ with respect to $\|\cdot\|_\infty$. Let $\varepsilon > 0$ and let $f \in L^2(\mathbb{T})$. From Corollary 3.21, we know that $C[-\pi, \pi]$ is dense in $L^2[-\pi, \pi]$ with respect to $\|\cdot\|_2$. Hence, there exists $g \in C(\mathbb{T})$ such that $\|g - f\|_2 < \varepsilon$. Choose $p \in \mathrm{Pol}(\mathbb{T})$ such that $\|p - g\|_\infty < \varepsilon$, which exists by our above assumption. Then we obtain

$$\|p - g\|_2 \leq \|p - g\|_\infty < \varepsilon,$$

and the triangle inequality implies that $\|f - p\|_2 < 2\varepsilon$.

To show that $\mathrm{Pol}(\mathbb{T})$ is dense in $C(\mathbb{T})$ with respect to $\|\cdot\|_\infty$, we recall the Stone-Weierstrass Theorem. It states that if X is a compact metric space and $\mathcal{A} \subseteq C(X)$ is an algebra which is self-adjoint, separates points, and contains a non-zero constant function, then $\mathcal{A}$ is dense in $C(X)$ with respect to $\|\cdot\|_\infty$. Recall that $\mathcal{A} \subseteq C(X)$ separates points if for any distinct points $x, y \in X$, there exists a function $f \in \mathcal{A}$ such that $f(x) \neq f(y)$.

In our case, we have $X = \mathbb{T} \subseteq \mathbb{C}$ and $\mathcal{A} = \mathrm{Pol}(\mathbb{T})$. It is clear that $\mathrm{Pol}(\mathbb{T})$ is an algebra, and it is self-adjoint since $\overline{e_n} = e_{-n}$. It also contains $1 = e_0 \in \mathrm{Pol}(\mathbb{T})$. Finally, we see that $\mathrm{Pol}(\mathbb{T})$ separates points because if $\theta_1, \theta_2 \in \mathbb{T}$ are distinct, then

$$e_1(\theta_1) = e^{i\theta_1} \neq e^{i\theta_2} = e_1(\theta_2).$$

It follows from Stone-Weierstrass that $\mathrm{Pol}(\mathbb{T})$ is dense in $C(\mathbb{T})$ with respect to $\|\cdot\|_\infty$, which shows that $\{e_n\}_{n\in\mathbb{Z}}$ is an orthonormal basis for $L^2(\mathbb{T})$.

Now, let $f \in L^2(\mathbb{T})$. Our goal is to show that $\|f - S_N f\|_2 \to 0$. Let $\varepsilon > 0$. From our work above, we can find $p \in \mathrm{Pol}(\mathbb{T})$ such that $\|f - p\|_2 < \varepsilon$. By choosing $N \geq \deg(p)$, we have

$$\begin{aligned}
\|f - S_N f\| &\leq \|f - p\|_2 + \|p - S_N f\|_2 \\
&= \|f - p\|_2 + \|S_N p - S_N f\|_2 \\
&\leq \|f - p\|_2 + \|S_N\| \|f - p\|_2 \\
&< 2\varepsilon
\end{aligned}$$

since $\|S_N\| \leq 1$. It follows that $f = \lim_{N\to\infty} S_N f$ with respect to $\|\cdot\|_2$. □

Corollary 4.8: Plancharel Identity

For any $f \in L^2(\mathbb{T})$, we have

$$\|f\|_2^2 = \sum_{n\in\mathbb{Z}} |\hat{f}(n)|^2.$$

Proof. We know that $S_N f \to f$ with respect to $\|\cdot\|_2$. Applying Parseval's Theorem gives us

$$\|f\|_2 = \lim_{N\to\infty} \|S_N f\|_2 = \lim_{N\to\infty} \sum_{n=-N}^{N} |\hat{f}(n)|^2 = \sum_{n\in\mathbb{Z}} |\hat{f}(n)|^2.$$ □

Definition 4.9

Let $(H_1, \|\cdot\|_{H_1})$ and $(H_2, \|\cdot\|_{H_2})$ be Hilbert spaces. A **unitary isomorphism** from H_1 to H_2 is a linear map $U : H_1 \to H_2$ which is surjective and isometric. That is, for all $x \in H_1$, we have

$$\|Ux\|_{H_2} = \|x\|_{H_1}.$$

Note that isometric implies injective, so U is a bijection.

In particular, Plancharel's Theorem (Corollary 4.8) says that the function $U : L^2(\mathbb{T}) \to \ell^2(\mathbb{Z})$ defined by $f \mapsto (\hat{f}(n))_{n\in\mathbb{Z}}$ is a unitary isomorphism. Then for $f, g \in L^2(\mathbb{T})$, we have

$$\langle Uf, Ug\rangle = \sum_{n\in\mathbb{Z}} \hat{f}(n)\overline{\hat{g}(n)} = \lim_{N\to\infty} \sum_{n=-N}^{N} \hat{f}(n)\overline{\hat{g}(n)} = \lim_{N\to\infty} \langle S_n f, g\rangle = \langle f, g\rangle = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(\theta)\overline{g(\theta)}\,\mathrm{d}\theta,$$

so U preserves inner products. In fact, this property is equivalent to being surjective and isometric.

4.3 Pointwise and Uniform Convergence of Fourier Series

So far, we know that if $f \in L^2(\mathbb{T})$, then the Fourier coefficients $(\hat{f}(n))_{n\in\mathbb{Z}}$ allow us to completely recover f in terms of L^2-convergence. Namely, we have

$$S_N f(\theta) = \sum_{n=-N}^{N} \hat{f}(n) e^{in\theta}$$

where $\lim_{N\to\infty} \|f - S_N f\|_2 = 0$. Moreover, by abstract measure theory, we know that for all $f \in L^2(\mathbb{T})$, there exists a subsequence $(N_k)_{k=1}^{\infty}$ with $N_k < N_{k+1}$ for all $k \in \mathbb{N}$ such that

$$f(\theta) = \lim_{k\to\infty} S_{N_k} f(\theta) = \lim_{k\to\infty} \sum_{n=-N_k}^{N_k} \hat{f}(n) e^{in\theta}$$

for almost every $\theta \in [-\pi, \pi]$.

Our fundamental problem is the following: for all $f \in L^2(\mathbb{T})$, do we have $S_N f \to f$ almost everywhere (that is, without passing to a subsequence)?