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 of 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 $u_t = Ku_{xx}$, 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

$$T(t)X(\pi) = T(t)X(-\pi),$$
$$T(t)X'(\pi) = T(t)X'(-\pi).$$

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

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

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 $1 \leq i \leq n$, 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 $L(f, P) \leq U(f, P)$. 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

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

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

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

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 $f_1 \leq f_2 \leq \cdots$ 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

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

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 our 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 extends 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 $E \subsetneq [0,1)$ 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 $E \subsetneq [0,1)$ 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. $\square$

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. □

2.3 Lebesgue Measure

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. □

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 $\mathcal{L}$.