MATH 118ABC Lecture Notes

Professors: Dr. Denis Labutin, Dr. Gustavo Ponce

Fall 2019 - Spring 2020

Bryan Xu

Introduction

These are the lecture notes for MATH 118ABC - Real Analysis I-III, from the 2019-2020 school year taught by Denis Labutin and Gustavo Ponce. This course covers the real number system, elements of set theory, continuity, differentiability, Riemann integral, implicit function theorems, convergence processes, and special topics.

Contents

1 Preliminaries

We review a few concepts from 117 and cover some things more in depth.

1.1 Construction of $\mathbb{R}$

Recall that the real numbers, which we denote $\mathbb{R}$, are an ordered field such that $\mathbb{Q} \subset \mathbb{R}$ and the Axiom of Completeness holds.

Definition 1 (Dense Set)**.** A set $E \subset \mathbb{R}$ is *dense* in $\mathbb{R}$ if given any two real numbers $a < b$, it is possible to find a point $x \in E$ such that $a < x < b$.

Theorem 1 (Nested Segments Principle)**.** Let $[a_1, b_1] \supset [a_2, b_2] \supset \cdots \supset [a_n, b_n] \supset \cdots$, and let $I_n = [a_n, b_n]$, for $a_n, b_n \in \mathbb{R}$, $\forall n$. Then

$$\bigcup_{n=1}^{\infty} I_n \neq \emptyset.$$

This requires closed intervals, because if we take $I_n = (0, \frac{1}{n})$, then $\bigcup_{n=1}^{\infty} I_n = \emptyset$, and it also requries bounded intervals, because if we take $E_n = [n, \infty)$, then $\bigcup_{n=1}^{\infty} E_n = \emptyset$ as well.

Proof. Define $L = \{a_n \mid n \in \mathbb{N}\}$. Nested implies that b_k is an upper bound for L. Thus there exists $\sup L = m$ and $m \leq b_k$, for all k. Let $R = \{b_k \mid k \in \mathbb{N}\}$ that has m has its lower bound. This means $\inf R = M \in \mathbb{R}$, and $m \leq M$. Hence for all n, $a_n \leq m \leq M \leq b_n$ implies that for all n,

$$\bigcup_{n=1}^{\infty} [a_n, b_n] \supset [m, M],$$

where possibly $m = M$. If this is the case, then $[m, M] = m$, but either way, it is not equal to the empty set. $\square$

In contrast to the Axiom of Completeness, we can construct the real numbers in a different way.

Definition 2 (Dedekind Cut)**.** A subset A of the rational numbers is called a *Dedekind cut*, or just simply *cut*, if it has the three properties:

(a) A is nonempty and $A \neq \mathbb{Q}$,

(b) If $r \in A$, then A contains every rational $q < r$,

(c) A does not have a maximum; if $r \in A$, then there exists $s \in A$ such that $r < s$.

Suppose $r \in \mathbb{Q}$ and we let $A = \{t \in \mathbb{Q} \mid t < r\}$. Is A a cut?

(a) Clearly, $A \neq \mathbb{Q}$ and A is nonempty.

(b) If $t_1 \in A$, then $r > t_1 > t_2 \in \mathbb{Q}$, so $t_2 \in A$.

(c) Since $r \notin A$, by the density of $\mathbb{Q}$, there always exists $t_2 \in A$ for all $t_1 \in A$, so A does not have a maximum.

Hence A is a cut. Be careful, however, as not every cut is of this form. The set $B = \{t \in \mathbb{Q} : t \leq 2\}$ is NOT a cut, as it does have a maximum.

Definition 3 (Real Numbers $\mathbb{R}$)**.** We define the real numbers $\mathbb{R}$ to be the set of all cuts in $\mathbb{Q}$.

We want to rigorously define algebraic operations in $\mathbb{R}$ using this idea of cuts.

Definition 4 (Addition)**.** Given A and B in $\mathbb{R}$, define

$$A + B = \{a + b \mid a \in A, \text{ and } b \in B\}.$$

Is $A + B$ a cut? We check our three conditions once again:

(a) Clearly $A + B$ is nonempty. To show $A + B \neq \mathbb{Q}$, since $A \neq \mathbb{Q}$ and $B \neq \mathbb{Q}$, there exists $r \notin A$ and $s \notin B$. For all $a \in A$, we have $r > a$ and for all $b \in B$, we have $s > b$. This gives us $r + s > a + b$ for all $a + b \in A + B$, and in particular, there exists $r + s \notin A + B$. Thus $A + B \neq \mathbb{Q}$.

(b) Let $a + b \in A + B$ be arbitrary and let $s \in \mathbb{Q}$ satisfy $s < a + b$. Then $s - b < a$, and so $s - b \in A$ because A is a cut. But then

$$s = (s - b) + b \in A + B,$$

and so this property is satisfied.

(c) Suppose $A + B$ has a maximum. Then $\max(A + B) = r + s$ for some $r \in A$ and $s \in B$. Since A and B have no maximum, there exists $r' \in A$ and $s' \in B$ such that $r' > r$ and $s' > s$. This gives $r + s < r' + s' \in A + B$, which contradicts the fact that $r + s$ is the maximum.

It is easy to show that the ordered field axioms are satisfied by addition as well, and thus we can see that addition is indeed well-defined in our construction of $\mathbb{R}$. We also want an identity element; that is, some element where $A + O = A$ is satisfied. This will be

$$O = \{p \in \mathbb{Q} \mid p < 0\}.$$

We want to be able to satisfy the inverse property as well, which is that $A + (-A) = O$. How would we define $-A$?

Definition 5 (Additive Inverse)**.** Given $A \in \mathbb{R}$, define

$$-A = \{r \in \mathbb{Q} \mid \exists t \notin A \text{ with } t < -r\}.$$

It is not hard to show that $-A$ is indeed a cut. Finally, we can formally define multiplication.

Definition 6 (Multiplication)**.** Given $A \geq O$ and $B \geq O$, define the product

$$AB = \{ab \mid a \in A, b \in B, \text{ where } a, b \geq 0\} \cup \{q \in \mathbb{Q} \mid q < 0\}.$$

Again, showing that this is a cut is a mere formality:

(a) Clearly $AB \neq \emptyset$ because $\{q \in \mathbb{Q} \mid q < 0\} \neq \emptyset$, and $AB \neq \mathbb{Q}$ since both A and B are not equal to $\mathbb{Q}$.

(b) If $t \in AB$ and $r \in \mathbb{Q}$ such that $0 < r < t$, then $t = ab$ for some $a \in A$ and $b \in B$. Since $(r/a)a < ab$ and $r/a < b$, we have $r = (r/a)a \in AB$.

(c) If $t < 0$ is an element of AB, then by the density of the rationals, there always exists $r > t$ such that $r < 0$. If $t > 0$ then $t = ab$ for some $a \in A$ and $b \in B$, where $a, b > 0$. Because A and B are cuts, there exists $a < c \in A$ and $b < d \in B$ such that $t < cd \in AB$. Hence AB does not have a maximal element.

1.2 Sequences and Series

In this section, we review the concept of sequences.

Definition 7 (Sequence)**.** A *sequence* is any $f : \mathbb{N} \to \mathbb{R}$, where $n \mapsto f(n) = f_n$.

We can study the "end behavior" or "long-term behavior" of a sequence by looking at its convergence.

Definition 8 (Convergence)**.** We say that a sequence a_n *converges* to $L \in \mathbb{R}$ if

$$\forall \epsilon > 0, \exists N \in \mathbb{N} \text{ such that } \forall n \geq N, |a_n - L| < \epsilon.$$

We usually denote this as either $a_n \to L$ or

$$\lim_{n\to\infty} a_n = L.$$

Definition 9 (Infinity)**.** Let E be a set. If E is not bounded above (and consequently, $M \in \mathbb{R}$ is not an upper bound of E, for all M), then we define the supremum of E to be *infinity*:

$$\sup E = \infty.$$

Similarly, if E is not bounded below, then we define the infimum of E to be negative infinity:

$$\inf E = -\infty.$$

How to we prove prove convergence? There are a couple of ways to do this.

(1) Calculus convergence lemmas. If $\exists L_1, L_2 \in \mathbb{R}$, where $a_n \to L_1$ and $b_n \to L_2$, then the following are true:

- $c_1 a_n + c_2 b_n \to c_1 L_1 + c_2 L_2$,
- $a_n b_n \to L_1 L_2$,
- Provided $L_2 \neq 0$, then $\frac{a_n}{b_n} \to \frac{L_1}{L_2}$.

(2) Squeeze Theorem. If both $a_n, b_n \to L$, for $L \in \mathbb{R}$ or $\pm\infty$, and $a_n \leq c_n \leq b_n$, for all n, then $c_n \to L$.

(3) Monotonic sequences. Recall that a sequence a_n is *monotonic increasing* if $a_n \leq a_{n+1}$, $\forall n$. If this is true, then

$$\lim_{n\to\infty} a_n = \sup\{a_n \mid n \in \mathbb{N}\} = L \in \mathbb{R} \text{ or } \infty.$$

This leads us to an important theorem:

Theorem 2 (Monotone Convergence Theorem)**.** If a_n is a monotonic and bounded sequence, then a_n converges.

Example 1. Let $a > 1$. Let $x_n = \frac{n}{a^n}$. Find the limit of x_n.

Proof. First we can see that

$$\frac{x_{n+1}}{x_n} = \frac{n+1}{a^{n+1}} \cdot \frac{a^n}{n} = \left(1 + \frac{1}{n}\right) \cdot \frac{1}{a} \implies x_{n+1} = \frac{1}{a}\left(1 + \frac{1}{n}\right)x_n.$$

Letting $n \to \infty$ on both sides, we get $L = \frac{1}{a} \cdot L$, and then because $a > 1$, we can see that $L = 0$. However, this only proves that if the limit exists, then $L = 0$. We need to show that the limit exists. Then for fixed $\delta > 0$,

$$\frac{x_{n+1}}{x_n} = \left(1 + \frac{1}{n}\right)(1 - \delta), \quad \forall n$$

Then we find N_0 such that $\frac{1}{n} < \delta$ for all $n \geq N_0$. So for all $n \geq N_0$,

$$\frac{x_{n+1}}{x_n} \leq (1 + \delta)(1 - \delta) = 1 - \delta^2 < 1.$$

Thus for all $n \geq N_0$, we have $\frac{x_{n+1}}{x_n} \leq 1$, so $x_{n+1} \leq x_n$, so $x_{N_0} \geq x_{N_0+1} \geq x_{N_0+2} \geq \cdots$. Hence our sequence is monotonic decreasing and $0 \leq x_n \forall n$. Thus our limit exists, and $L \geq 0$. □

Definition 10 (Divergence to $\pm\infty$)**.** We say that a sequence a_n diverges to $\pm\infty$ if

$$\forall T \in \mathbb{R}, \exists N \in \mathbb{N} \text{ such that } n \geq N \implies a_n \geq T \quad (a_n \leq T).$$

Consequently, we say simply say that a sequence *diverges* if it doesn't converge and doesn't diverge to $\pm\infty$.

Definition 11 (Subsequence)**.** We say that b_n is a subsequence of a sequence a_n iff $\exists \sigma : \mathbb{N} \to \mathbb{N}$ such that $\sigma(n) < \sigma(n+1)$ for all n and

$$b_n = a_{\sigma(n)}.$$

Subsequences are very useful in determining if a_n diverges. This is because

$$a_n \to L \implies b_n \to L \text{ for all subsequences } b_n \text{ of } a_n.$$

Theorem 3 (Bolzano-Weierstrass Theorem)**.** If a_n is a sequence such that $\exists M$ where $|a_n| \leq M$ for all n, then there exists $b_n = a_{m_n}$, a subsequence such that $a_{m_n} \to L$.

Definition 12 (Series)**.** Let a_n be a sequence. We say that the *series* converges; that is,

$$\sum_{n=1}^{\infty} a_n = L \in \mathbb{R},$$

if the sequence $s_n = a_1 + \cdots + a_n$ satisfies $s_n \to L$. If $S_n \to \infty$, then the series diverges to ∞.

The easiest series so to speak are those with $\sum_{n=1}^{\infty} a_n$ where $a_n \geq 0$ for all n. This is because $S_1 \leq \cdots S_n \leq S_{n+1} \leq \cdots$. Then the series converges if and only if $\exists M < \infty$ such that $S_n \leq M$ for all n.

Definition 13 (Limit Superior/Inferior)**.** Let a_n be a sequence. Define a new sequence a_N^+ by

$$a_N^+ = \sup a_{n \geq N}.$$

Then we define the *limit superior* of a_n as

$$\limsup_{n\to\infty} a_n = \inf a_N^+.$$

Similarly, we can define another new sequence a_N^- by

$$a_N^- = \inf a_{n \geq N}.$$

Then we define the *limit inferior* of a_n as

$$\liminf_{n\to\infty} a_n = \sup a_N^-.$$

In other words, the limit superior is the largest limit point, while the limit inferior is the smallest limit point.

1.3 Continuity

Definition 14 (Convergence of a Function)**.** We say that a function $f : (a,b) \to \mathbb{R}$ *converges* to limit $L \in \mathbb{R}$ at p iff

$$\forall \epsilon > 0, \exists \delta > 0 : \text{ if } |x - p| < \delta, \text{ then } |f(x) - L| < \epsilon.$$

This is denoted

$$\lim_{x \to p} f(x) = L.$$

There exists an equivalent formulation of this definition in terms of sequences:

Definition 15 (Sequential Convergence of a Function)**.** We say that a function $f : (a, b) \to \mathbb{R}$ *converges* to limit $L \in \mathbb{R}$ at p iff

$$\forall x_n \text{ such that } x_n \to p \text{ with } x_n \neq p,\ f(x_n) \to L,\ \forall n.$$

Definition 16 (Continuity)**.** We say that a function f is *continuous* at p if

$$\lim_{x \to p} f(x) = f(p).$$

In particular, this is equivalent to saying

$$\forall \epsilon > 0, \exists \delta > 0 \text{ such that if } |x - p| < \delta, \text{ then } |f(x) - f(p)| < \epsilon.$$

Some properties of continuous functions:

- If f is continuous at 0, and $f(0) < 0$, then $\exists \delta > 0$ such that $f(x) < 0$, for all $x \in [-\delta, \delta]$.
- f, g continuous at p implies that $f + g$, fg, and $\frac{f}{g}$ are all continuous at p (assuming $g \neq 0$).

Theorem 4. If f is a continuous function, then for all x_n such that $x_n \to p$, we have that $f(x_n) \to f(p)$. In particular, this means that $|f(x_n) - f(p)| \to 0$.

As a consequence of this, this means that f is not continuous at p if there exists some x_n such that $x_n \to p$ but $f(x_n) \not\to f(p)$.

Definition 17 (C^0–Class)**.** Let E be any open/closed/combination interval. Then we say that $f \in C(E)$, the C^0*–class* of functions iff f is continuous at every $p \in E$.

Theorem 5 (Intermediate Value Theorem)**.** Let $f \in C([a, b])$, where $a, b \in \mathbb{R}$, and $f(a) < f(b)$. Then for all L with $f(a) \leq L \leq f(b)$, there exists $c \in [a, b]$ such that $f(c) = L$.

Proof. Suppose $L = 0$ and $f(a) \leq 0 \leq f(b)$. If $f(a) = 0$ or $f(b) = 0$, then we're done. Otherwise, $f(a) < 0 < f(b)$, and we define $I_1 = [a, b]$. Then if $f(\frac{a+b}{2}) = 0$, we're done again. But if $f(\frac{a+b}{2}) > 0$, then I_2 will be the left-side of the interval, and if $f(\frac{a+b}{2}) < 0$, then I_3 will be the right-side of the interval. Thus

$$I_n = [a_n, b_n].$$

If $I_1 \supset I_2 \supset \cdots \supset I_n$ has solution, then we're done. Otherwise, the length will be

$$\ell(I_n) = \frac{|b-a|}{2^n} \cdot 2, \quad f(a_n) < 0,\, f(b_n) > 0.$$

By the Nested Segments Principle, $\bigcup_{n=1}^{\infty} I_n = \{x_*\}$. Thus $a_n \to x_*$ and $b_n \to x_*$, and because $a \le a_n \le b_n$ and $a \le b \le b_n$, it follows that $a \le x_* \le b$, and so $x_* \in [a,b]$. Thus, f is continuous at x_*. Finally,

$$f(a_n) > 0,\, a_n \to x_* \implies f(a_n) \to f(x_*) \le 0, \quad \forall n,$$

$$f(b_n) < 0,\, b_n \to x_* \implies f(b_n) \to f(x_*) \ge 0, \quad \forall n.$$

Hence we conclude that $f(x_*) = 0$.

If $L \neq 0$, then consider a new function $g(x) - f(x) - L$. Then $g(a) \le 0$ and $g(b) \ge 0$. We apply Case 1, and so $g(x_*) = 0 \iff f(x_*) = L$. □

In general, the δ that we pick will depend on f, ϵ, and x. If we want to get rid of this requirement, then we must introduce a new concept.

Definition 18 (Uniform Continuity)**.** We say that a function $f : E \to \mathbb{R}$ is *uniformly continuous* on E iff

$$\forall \epsilon > 0, \exists \delta > 0 \text{ such that if } \forall x, y \in E,\, |x-y| < \delta, \text{ then } |f(x) - f(y)| < \epsilon.$$

We we fix x, then $\lim_{y \to x} f(y) = f(x)$, and so f is continuous at y. Thus f being uniformly continuous will always imply that f is continuous. Much like with continuity, there exists an equivalent sequential formulation of uniform continuity.

Definition 19 (Sequential Uniform Continuity)**.** If f is uniformly continuous on E, then

$$\forall x_n, y_n \in E, \text{ if } |x_n - y_n| \to 0, \text{ then } |f(x_n) - f(y_n)| \to 0.$$

Top rove that a function is not uniformly continuous, we can negate this definition to get that f is not unifomly continuous on E if

$$\exists \epsilon_0 > 0, \exists x_n, y_n \in E \text{ such that } |x_n - y_n| \to 0 \text{ and } |f(x_n) - f(y_n)| \to 0.$$

Example 2. Let $E = (0,1)$ and $f(x) = \frac{1}{x} \in C((0,1))$. If f uniformly continuous on E?

Proof. No. Take $\epsilon_0 = 1$. Let $u_n = \frac{1}{n}$, and we want v_n to be defined by $\frac{1}{u_n} - \frac{1}{v_n} = 1$. Solving for v_n, we get $v_n = \frac{1}{n-1}$. Then

$$|u_n - v_n| = \left|\frac{1}{n} - \frac{1}{n-1}\right| \leq \left|\frac{1}{n}\right| + \left|\frac{1}{n-1}\right| \to 0,$$

but

$$|f(u_n) - f(v_n)| = \left|\frac{1}{u_n} - \frac{1}{v_n}\right| = |n - (n-1)| = 1 \geq 1.$$

Thus $|f(x_n) - f(y_n)| \geq \epsilon_0$. □

Theorem 6. Let $a, b \in \mathbb{R}$, and $f \in C([a,b])$. Then f is uniformly continuous on $[a,b]$.

Proof. Suppose to the contrary that f is not uniformly continuous on $[a,b]$. Then there exists $\epsilon_0 > 0$ and x_n, y_n such that $|x_n - y_n| \to 0$ but $|f(x_n) - f(y_n)| \geq \epsilon_0$. Then because x_n is bounded, we find n_k, a subsequence defined by $n_k = x_{n_k}$, where $n_k \to p$. Then

$$a \leq n_k \leq b \implies a \leq p \leq b \implies p \in [a,b] \implies f(n_k) \to f(p).$$

We take $m_k = y_{m_k}$, $|n_k - m_k| \to 0$. Thus $m_k \to p$ because

$$|m_k - p| = |m_k + n_k - n_k - p| \leq |m_k - n_k| + |n_k - p| \to 0.$$

Hence $f(m_K) \to f(p)$, and it follows that

$$|f(n_k) - f(m_k)| = |f(n_k) - f(p) + f(p) - f(m_k)| \leq |f(n_k) - f(p)| + |f(p) - f(m_k)| \to 0.$$

But $|f(n_k) - f(m_k)| \geq \epsilon_0 > 0$, a contradiction. □

2 Topology of $\mathbb{R}$

2.1 Basic Definitions

Definition 20 (Neighborhood)**.** Let $\epsilon > 0$, and $x \in \mathbb{R}$. The *neighborhood* of x is defined by

$$V_\epsilon(x) = (x - \epsilon, x + \epsilon) = \{z : |z - x| < \epsilon\}.$$

Using this definition, we can rewrite a few definitions:

- $x_n \to p$ if and only if $\forall \epsilon > 0, \exists N$ such that if $n \geq N$, then $x_n \in V_\epsilon(p)$.
- f is continuous at p if and only if $\forall \epsilon > 0, \exists \delta$ such that $f(v_\delta(p)) \subset V_\epsilon(f(p))$.

Definition 21 (Open Set)**.** A set $E \subset \mathbb{R}$ is *open* iff for all $x \in E$, there exists $\epsilon > 0$ such that $V_\epsilon(x) \subset E$.

For example, the set (a, b), for $a, b \in \mathbb{R}$ is open, because we let $x_0 \in (a, b)$ and

$$\epsilon = \frac{1}{2} \min\{|a - x_0|, |b - x_0|\}.$$

Then $V_\epsilon(x) \subset E$. However, the set $(0, 1]$ is not open because for $x_0 = 1$, we have that $\forall \epsilon > 0$, $V_\epsilon(1) \not\subset E$, because $1 + \frac{\epsilon}{10} \in V_\epsilon(1)$ but $1 + \frac{\epsilon}{10} \notin (0, 1]$.

Definition 22 (Interior Point)**.** Let $E \subset R$. A point p is an *interior point* of E iff there exists $\epsilon_0 > 0$ such that $V_{\epsilon_0}(p) \subset E$. We define

$$\text{int}(E) = \{p \mid p \text{ is an interior point of } E\}.$$

If E is open, then $\text{int}(E) = E$.

Definition 23 (Closed Set)**.** A set $E \subset \mathbb{R}$ is *closed* iff the complement, $E^c = \mathbb{R} \setminus E$, is open.

Open and closed do not form a partition of the reals; for example the set $(0,1]$ is not open nor closed. How do unions and intersections affect open/closed sets?

Theorem 7. (a) If $E_i \subset \mathbb{R}$ is open for all $i \in I$, then $\bigcup_{i\in I} E_i$ is open.

(b) If $E_1,\dots,E_N$ are open for $N < \infty$, then $\bigcap_{i=1}^N E_i$ is open.

(c) If $E_i \subset \mathbb{R}$ is closed for all $i \in I$, then $\bigcap_{i\in I} E_i$ is closed.

(d) If $E_1,\dots,E_N$ are closed for $N < \infty$, then $\bigcup_{i=1}^N E_i$ is closed.

Proof. Note that (a) $\implies$ (c) and (b) $\implies$ (d) because of DeMorgan's Laws.

(a) Take any $x_0 \in \bigcup_{i\in I} E_i$. Then $x_0 \in E_{i_0}$ for some $i_0 \in I$. Then

$$\exists \epsilon > 0 : V_\epsilon(x_0) \subset E_{i_0} \implies V_\epsilon(X_0) \subset \bigcup_{i\in I} E_i.$$

(b) Take any $x_0 \in \bigcup_{i=1}^N E_i$. Then $x_0 \in E_1$, $x_0 \in E_2$, $\cdots$, $x_0 \in E_N$. The assumption is that

$$\exists \epsilon_1 > 0 : V_{\epsilon_1}(x_0) \subset E_1, \exists \epsilon_2 > 0 : V_{\epsilon_2}(x_0) \subset E_2, \dots, \exists \epsilon_N > 0 : V_{\epsilon_N}(x_0) \subset E_N.$$

If we take $V_\epsilon = \min\{\epsilon_1,\dots,\epsilon_N\}$, then

$$v_\epsilon(x_0) \in E_n, \quad \forall n \in [1,N].$$

□

Note that (b) and (d) require N to be finite, because if we let $E_n = (-\frac{1}{n}, 1 + \frac{1}{n})$, which is open for all n, we find that

$$\bigcap_{i=1}^{\infty} E_i = [0,1],$$

which is obviously not open.

Definition 24 (Isolated Point)**.** Let $E \subset \mathbb{R}$. A point p is an *isolated point* of E iff $p \in E$ and there exists $\epsilon > 0$ such that $V_\epsilon(p) \cap E = p$.

For example, if we take the set $E = \{1, \frac{1}{2}, \frac{1}{3}, \dots, \frac{1}{n}, \dots\}$, every point is an isolated point.

Definition 25 (Limit Point)**.** Let $E \subset \mathbb{R}$. A point p is a *limit point* of E iff there exists x_n such that

- $x_n \in E$, for all n,
- $x_n \neq p$, for all n,
- $\lim_{n\to\infty} x_n = p$.

Note that if p is an isolated point of E, then p is not a limit point of E, and vice versa. Note also that a limit point can be contained in E, but it does not necessarily have to be in E.

Definition 26 (Closure)**.** Let $E \subset \mathbb{R}$. The *closure* of E is defined as

$$\overline{E} = E \cup \{p \mid p \text{ is a limit point of } E\}.$$

Theorem 8. A set $E \subset \mathbb{R}$ is closed iff E contains all of its limit points. In particular, if $\overline{E} = E$.

Proof. ($\Longrightarrow$): Assume $\mathbb{R} \setminus E$ is open. Take any limit point of E, say p. Suppose $p \notin E$. Then $p \in \mathbb{R} \setminus E$, and so there exists $\epsilon_0 > 0$ such that $V_{\epsilon_0}(p) \cap E = \emptyset$. This is a contradiction, because we can find x_n such that $x_n \in E$ for all n and $x_n \to p$.
($\Longleftarrow$): Suppose E is not closed. Then $R \setminus E$ is not open, and so some point $p \in \mathbb{R} \setminus E$ is not an interior point of $\mathbb{R} \setminus E$. This means for all $\epsilon > 0$, $V_\epsilon(P) \not\subset \mathbb{R} \setminus E$, and so $V_\epsilon(p) \cap E \neq \emptyset$. Then there exists $x_\epsilon \in V_\epsilon(p) \cap E$, with $x_\epsilon \neq p$. So $\epsilon = 1, \frac{1}{2}, \ldots, \frac{1}{n}$. It follwos that we have x_n such that $x_n \in E$ for all n, $x_n \neq p$ for all n, and $x_n \in V_{1/n}(p)$. Thus

$$0 \leq |x_n - p| \leq \frac{1}{n}.$$

By the Squeeze Theorem, $x_n \to p$, so p is a limit point of E. By our assumption, $p \in E$, a contradiction for the choice of p. □

Definition 27 (Boundary Point)**.** Let $E \subset \mathbb{R}$. We say a point p is a *boundary point* iff for all $\epsilon > 0$, $V_\epsilon(x) \cap E \neq \emptyset$ and $V_\epsilon(x) \cap E^c \neq \emptyset$. The set of all boundary points of E is denoted ∂E or $\text{bd}(E)$.

Definition 28 (Perfect Set)**.** A set $P \subset \mathbb{R}$ is *perfect* if it is closed and contains no isolated points.

As it turns out, any nonempty perfect set is uncountable.

Definition 29 (Connected Set)**.** Two nonempty sets $A, B \subset \mathbb{R}$ are *separated* if $\overline{A} \cap B$ and $A \cap \overline{B}$ are both empty. A set $E \subset \mathbb{R}$ is *disconnected* if it can be written as $E = A \cup B$, where A and B are nonempty separated sets. A set that is not disconnected is called a *connected* set.

Theorem 9. A set $E \subset \mathbb{R}$ is connected if and only if

(a) For all disjoint sets A and B satisfying $E = A \cup B$, there always exists a convergent sequence $x_n \to x$ with x_n contained in one of A or B, and x is an element of the other.

(b) Whenever $a < c < b$ for $a, b \in E$, it follows that $c \in E$ as well.

2.2 Compactness

Definition 30 (Open Cover)**.** Let $E \subset \mathbb{R}$. An *open cover* of E is any collection of sets $\mathcal{C} = \{\mathcal{O}_\alpha\}_{\alpha \in A}$ such that

- $\mathcal{O}_\alpha \subset \mathbb{R}$ is open, for all $\alpha \in A$,
- $\bigcup_{\alpha \in A} \mathcal{O}_\alpha \supset E$.

Some examples and nonexamples of open covers:

- The collection $\mathcal{C}_1 = \{(q-1, q+1)\}_{q \in \mathbb{Q}}$ is an open cover of $\mathbb{R}$, because for all $r \in \mathbb{R}$, $\exists q \in \mathbb{Q}$ such that $r \in (q-1, q+1)$ because of the density of $\mathbb{Q}$ in $\mathbb{R}$.
- The collection $\mathcal{C}_2 = \{(x-1), (x+1)\}_{x \in \mathbb{R}}$ is an open cover of $\mathbb{R}$ because for all $r \in \mathbb{R}$, we take $x = r$. Then $r \in (x-1, x+1)$, so $\mathcal{C}_2$ is indeed an open cover.
- For the set $E = [0, 1]$, the collection $\mathcal{C}_3 = \left\{(\frac{1}{n}, 2)\right\}_{n \in \mathbb{N}}$ is not an open cover of E, because for all $n \in \mathbb{N}$, $0 \notin (\frac{1}{n}, 2)$.

Theorem 10 (Heine-Borel)**.** Let $K \subset \mathbb{R}$. Then the following are equivalent:

(1) For all open covers $\mathcal{C}$ of K, there exists a finite subcover. That is, there exists

$$\mathcal{C}' = \{\mathcal{O}_{\alpha_1}, \mathcal{O}_{\alpha_2}, \ldots, \mathcal{O}_{\alpha_N}\},\ \alpha_j \in A,\ N < \infty,\ \bigcup_{i=1}^{N} \mathcal{O}_{\alpha_i} \supset K.$$

(2) K is closed and bounded.

(3) For all x_n such that $x_n \in K$, we can find a subsequence $y_k = x_{n_k}$ such that

$$\lim_{k\to\infty} y_k = a,$$

for some $a \in K$.

(4) K is *compact*.

Theorem 11. If $K \subset \mathbb{R}$ is compact, then $C(K)$ is a vector space.

Theorem 12. Let $K \subset \mathbb{R}$ be compact. Then for all $f \in C(K)$,

$$\sup_k f = \max_k f = f(x_1), \text{ for some } x_1 \in K$$

$$\inf_k f = \min_k f = f(x_2), \text{ for some } x_2 \in K$$

Proof. Let $M = \sup_k f$, where $M \in \mathbb{R} \cup \{\infty\}$. Then there exists x_n such that $x_n \in K$ and $m - \frac{1}{n} \leq f(x_n) \leq M$, if $M < \infty$, or $f(x_n) \geq n$ if $M = \infty$. Compactness tells us that there exists x_{n_k} such that $x_{n_k} \to a \in K$. Thus $f(x_{n_k}) \to f(a)$, so the second case is impossible, and $M = f(a)$. □

2.3 Further Topics

Definition 31 (F_σ/G_δ Set)**.** A set $A \subset \mathbb{R}$ is called an F_σ *set* if it can be written as the countable union of closed sets. A set $B \subset \mathbb{R}$ is called a G_δ *set* if it can be written as a countable intersection of open sets.

Obviously, a set A is a G_δ set if and only if its complement is an F_σ set. The set $[a,b]$ is a G_δ set because we can write it as

$$[a,b] = \bigcap_{n=1}^{\infty} \left(a - \frac{1}{n}, b + \frac{1}{n}\right).$$

Example 3. Show that $\mathbb{Q}$ is an F_σ set and that $\mathbb{I} = \mathbb{R} \setminus \mathbb{Q}$ is a G_δ set.

Proof. We can write $\mathbb{Q}$ as
$$\mathbb{Q} = \bigcap_{q \in \mathbb{Q}} \{q\}.$$
Since $\{q\}$ is closed, and $\mathbb{Q}$ is countable, we can see that $\mathbb{Q}$ is a F_σ set. Then by DeMorgan's Law,
$$\mathbb{I} = \mathbb{R} \setminus \mathbb{Q} = \left(\bigcap_{q \in \mathbb{Q}} \{q\} \right)^c = \bigcup_{q \in \mathbb{Q}} \mathbb{R} \setminus \{q\},$$
and because $\mathbb{R} \setminus \{q\}$ is open, $R \setminus \mathbb{Q}$ is a G_δ set. □

Theorem 13. If $\{G_1, G_2, \dots\}$ is a countable collection of open and dense sets, then the intersection $\bigcap_{n=1}^{\infty} G_n$ is not empty.

Recall that a set G is dense in $\mathbb{R}$ if and only if $\overline{G} = \mathbb{R}$. We have a related definition.

Definition 32 (Nowhere-Dense Set)**.** A set E is *nowhere-dense* if $\overline{E}$ contains no nonempty open intervals.

Theorem 14. A set E is nowhere-dense if and only if the complement of $\overline{E}$ is dense in $\mathbb{R}$.

Proof. $(\Longrightarrow)$: Suppose E is nowhere-dense. Then we choose $a, b \in \mathbb{R}$ with $a < b$. There exists $V_\epsilon(x) \subseteq (a, b)$. Since $\overline{E}$ by assumption does not contain any nonempty open intervals, $V_\epsilon(x) \cap \overline{E}^c \neq \emptyset$, so we choose any $y \in V_\epsilon(x) \cap \overline{E}^c$, and we find $a < y < b$ where $y \in \overline{E}^c$. Hence $\overline{E}^c$ is dense in $\mathbb{R}$.
$(\Longleftarrow)$: Suppose $\overline{E}^c$ is dense in $\mathbb{R}$. Then for any $a, b \in \mathbb{R}$ and $a < b$, there exists $y \in \overline{E}^c$ such that $a < y < b$. So for any open interval $(a, b) \cap \overline{E}^c \neq \emptyset$. This shows that (a, b) cannot be contained in $\overline{E}$, so E is nowhere-dense. □

Theorem 15 (Baire's Theorem)**.** The set of real numbers $\mathbb{R}$ cannot be written as the countable union of nowhere-dense sets.

3 Differentiation

3.1 Introduction

There are two related but distinct definitions for derivatives. The first definition is only really useful in one dimension.

Definition 33 (Derivative)**.** Let $f : (a,b) \to \mathbb{R}$, with $x_0 \in (a,b)$. The *derivative* of f at x_0 is defined as

$$f'(x_0) = \lim_{\Delta x \to 0} \frac{f(x_0 + \Delta x) - f(x_0)}{\Delta x} = \lim_{x \to x_0} \frac{f(x) - f(x_0)}{x - x_0}.$$

Definition 34 (Differentiability)**.** Let $f : (a,b) \to \mathbb{R}$, with $x_0 \in (a,b)$. Then f is *differentiable* at x_0 iff there exists $A \in \mathbb{R}$ such that for all Δx,

$$f(x_0 + \Delta x) = f(x_0) + A\Delta x + r(\Delta x), \quad \frac{|r(\Delta x)}{\Delta x} \to 0, \text{ as } \Delta x \to 0.$$

This means that $r(\Delta x)$ tends to 0 faster than any linear function. For example, $r(\Delta x) = (\Delta x)^2$ or $r(\Delta x) = |\Delta x|^{1+\alpha}$ for $\alpha > 0$ are fine, but $r(\Delta x) = \frac{\Delta x}{10000}$ is not fine.

Theorem 16. Let $f : (a,b) \to \mathbb{R}$, and $x_0 \in (a,b)$. Then

(1) $\exists f'(x_0) \in \mathbb{R} \Longleftrightarrow f$ is differentiable at x_0.

(2) $A = f'(x_0)$ in the definition.

Proof. $(\Longrightarrow)$: We have

$$\frac{f(x_0 + \Delta x) - f(x_0)}{\Delta x} - f'(x_0) = \alpha(\Delta) \to 0, \text{ as } \Delta x \to 0$$

$$\Longrightarrow \forall \Delta x,\ f(x_0 + \Delta x) - f(x_0) = f'(x_0)\Delta x + \Delta x \alpha(\Delta x)$$

$$\Longrightarrow f(x_0 + \Delta x) - f(x_0) = A\Delta x + r(\Delta x).$$

Then

$$\frac{|r(\Delta x)|}{\Delta x} = \frac{|\Delta x \alpha|}{|\Delta x|} = |\alpha(\Delta x)| \to 0.$$

$(\Longleftarrow)$: Suppose f is differentiable at x_0. Then $\forall \Delta x \neq 0$, $f(x_0 + \Delta x) = f(x_0) + A\Delta x + r(\Delta x)$. Then

$$\Longrightarrow \frac{f(x_0) + \Delta x) - f(x_0)}{\Delta x} = A + \frac{r(\Delta x)}{\Delta x}$$

$$\Longrightarrow \lim_{\Delta x \to 0} \frac{f(x_0 + \Delta x) - f(x_0)}{\Delta x} = A \cdot \lim_{\Delta x \to 0} \frac{r(\Delta x)}{\Delta x} = f'(x_0).$$

□

When we compare the definitions of continuity and differentiability, we can see that they are quite similar:

- Continuity:

$$\forall \Delta x : f(x_0 + \Delta x) = f(x_0) + f(x_0 + \Delta x) - f(x_0) = f(x_0) + \rho(\Delta x), \quad \lim_{\Delta x \to 0} \rho(\Delta x) = 0.$$

- Differentiability:

$$\forall \Delta x : \rho(x_0 + \Delta x) = f(x_0) + f'(x_0)\Delta x + r(\Delta x), \quad \lim_{\Delta x \to 0} \frac{|r(\Delta x)|}{\Delta x} = 0.$$

As a direct corollary,

Theorem 17. If $f : (a, b) \to \mathbb{R}$ is differentiable at $x_0 \in (a, b)$ then f is continuous at x_0. The converse is false in general.

Definition 35 (Sequential Derivative)**.** If $f'(x_0)$ exists, then

$$\forall x_n \text{ such that } x_n \neq x_0, \text{ we have } f'(x_0) = \lim_{n \to \infty} \frac{f(x_n) - f(x_0)}{x_n - x_0}.$$

We can use this to prove that $f(x) = |x|$ is not differentiable at $x = 0$. Take $x_n = \frac{1}{n}$ and $y_n = -\frac{1}{n}$. Then

$$f'(0) = \lim_{n \to \infty} \frac{1/n}{1/n} = 1, \quad f'(0) = \lim_{n \to \infty} \frac{|-1/n|}{-1/n} = -1.$$

Since the limits are not equal, f is not differentiable at $x = 0$.

3.2 Derivative Theorems

Theorem 18 (First Derivative Test)**.** Let $f:(a,b)\to\mathbb{R}$. Let $x_*: f(x_*)\geq f(x)$, for all $x\in(a,b)$. Say that $f'(x_*)$ exists. Then $f'(x_*)=0$.

Proof. We know that

$$\lim_{\Delta x\to 0}\frac{f(x_*+\Delta x)-f(x_*)}{\Delta x}$$

exists and is finite. Thus for all x_n such that $\Delta x_n\neq 0$ for all n, $\Delta x\to 0$. So

$$f'(x_*)=\lim_{n\to\infty}\frac{f(x_*+\Delta x_n)-f(x_*)}{\Delta x_n}.$$

Consider Δx_n where $\Delta x_n>0$ and $\Delta\tilde{x}_n$ where $\Delta x_n<0$. Therefore $f'(x_*)=0$. □

Theorem 19 (Mean-Value Theorem)**.** Let $a,b\in\mathbb{R}$, $f\in C([a,b])$, and $f'(x)$ exists for all $x\in(a,b)$. Then there exists some x_0 with $a<x_0<b$ where

$$\frac{f(b)-f(a)}{b-a}=f'(x_0).$$

Proof. Let

$$g(x)=f(x)-\left[f(a)+\frac{f(b)-f(a)}{b-a}(x-a)\right]$$

. Then $g\in C([a,b])$. Since $[a,b]$ is compact, there exists $x_0\in[a.b]$ such that $g(x_0)=\max_{[a,b]}g$. We split into two cases:

- $g(x)=0$ on $[a,b]$. Then $g'(x)=0$, so

  $$f'(x)=\frac{f(b)-f(a)}{b-a}.$$

  Because $x\mapsto\alpha x+\beta$ is trivially differentiable, $f'=\alpha$, so we can take any x_0.

- $g(x)\neq 0$. Then $g(a)=g(b)=0$. Therefore, either

  $$\max_{[a,b]}g=g(x_0)>0,\text{ or }\min_{[a,b]}g=g(x_*)<0.$$

  Assume WLOG that $g(x_0)>0$, with x_0 being the point of the maximum. By the First Derivative Test, $g'(x_0)=0$ implies that

  $$0=f'(x_0)-\frac{f(b)-f(a)}{b-a},\text{ so }f'(x_0)=\frac{f(b)-f(a)}{b-a}.$$

□

Theorem 20 (Corollary to the MVT)**.** Let $f : (a,b) \to \mathbb{R}$. Assume for all $x \in (a,b)$, $f'(x) > 0$. Then f is strictly increasing on (a,b); that is, $a < x_1 < x_2 < b$ implies that $f(x_1) < f(x_2)$.

Proof. Fix $x_1 < x_2$. Because differentiability implies continuity, $f \in C([(x_1,x_2])$. Because f is differentiable at all $x \in (a,b)$, we apply the MVT to get

$$\frac{f(x_2) - f(x_1)}{x_2 - x_1} = f'(\xi), \quad \text{for some } \xi \in (x_1, x_2).$$

Because $f'(x) > 0$, $f'(\xi) > 0$ as well, so

$$\frac{f(x_2) - f(x_1)}{x_2 - x_1} > 0 \implies f(x_2) > f(x_1).$$

□

A special case of the Mean-Value Theorem gets its own name:

Theorem 21 (Rolle's Theorem)**.** Let $f : [a,b] \to \mathbb{R}$ be continuous on $[a,b]$ and differentiable on (a,b). If $f(a) = f(b)$, then there exists a point $c \in (a,b)$ such that $f'(c) = 0$.

For this section, we denote $f'(t), g'(t), \ldots$ as $\dot{f}(t), \dot{g}(t), \ldots$.

Definition 36 (Classical Solution)**.** We say that the function $u(t)$ is a *classical solution* to the following ordinary differential equation:

$$\begin{cases} \dot{u}(t) = F(t), & a \leq t \leq b \\ u|_{t=a} = x_0 \end{cases}$$

if and only if

(1) $u(t) \in C([a,b])$,

(2) There exists $u(t)$ for all $t \in (a,b)$, and $\dot{u}(t) = F(t)$, for all t,

(3) $u(a) = x_0$.

Theorem 22 (Uniqueness of Classical Solution)**.** There exists at most one classical solution $u(t)$ to the above differential equation.

Proof. Suppose $u_1(t)$ and $u_2(t)$ are two classical solutions. Define $v(t) = u_1(t) - u_2(t)$. This satisifes

(1) $v(a) = 0$,

(2) $v \in C([a,T])$, for all $T \leq b$,

(3) $\dot{v}(t) = u_1 \dot{-} u_2(t) = \dot{u}_1(t) - \dot{u}_2(t) = F - F = 0$.

We claim $v(t) = 0$. By the Mean-Value Theorem,

$$\frac{v(t) - v(a)}{t - a} = \dot{v}(t_*) \in (a,t) = 0.$$

So $v(t) = v(a)$ for all t. Thus $V(t) = 0$ for all t, and so $u_1(t) = u_2(t)$. □

Theorem 23 (Inverse Function Theorem)**.** Let $a, b \in \mathbb{R}$, $f \in C([a,b])$, and let $x_1 < x_2$, so that $f(x_1) < f(x_2)$. Then

(1) There exists $f^{-1} : [f(a), f(b)] \to [a,b]$, where $f^{-1} \circ f = \mathrm{id}_{[a,b]}$ and $f \circ f^{-1} = \mathrm{id}_{[f(a),f(b)]}$,

(2) $f^{-1} \in C([f(a), f(b)])$,

(3) If $x_0 \in (a,b)$, and there exists $f'(x_0) \neq 0$, then there exists

$$(f^{-1})'(f(x_0)) = \frac{1}{f'(x_0)}.$$

There are many different notations:

$$(f^{-1})'\Big|_{f(x_0)} = \frac{1}{f'|_{x_0}}, \quad (f^{-1})'\Big|_{y_0} = f'\Big|_{f^{-1}(y_0)} = f'(x_0).$$

Proof. (1) $[a,b] \xrightarrow{f} [\alpha, \beta]$ is injective by monotonicity. Then because f is continuous, by the Intermediate Value Theorem, f is also bijective.

(2) We want to prove that for all $\epsilon > 0$, we can find δ such that

$$|y - y_0| < \delta \implies |f^{-1}(y) - x_0| < \epsilon.$$

We use the Intermediate Value Theorem with $y_1 : f^{-1}(y_1) = x_0 + \epsilon$ and $y_2 : f^{-1}(y_2) = x_0 - \epsilon$. By monotonicity of f,

$$y_2 < y_0 < y_1.$$

By monotonicity of f^{-1}, take $\delta = \min\{|y_1 - y_0|, |y_2 - y_0|\}$. Hence we're done.

(3) We want

$$\lim_{y \to y_0} \frac{f^{-1}(y) - f^{-1}(y_0)}{y - y_0} = \frac{1}{f'(x_0)}.$$

Using sequences, for all y_n with $y_n \neq y_0$ and $y_n \to y_0$, we have

$$\frac{f^{-1}(y_n) - f^{-1}(y_0)}{y_n - y_0} = \frac{x_n - x_0}{f(x_n) - f(x_0)}.$$

Now for all y_n, there exists a unique x_n such that $f(x_n) = y_n$ and $x_n = f^{-1}(y_n)$. By (b), we know that $y_n \to y_0$ if and only if $x_n \to x_0$ and $y_n \neq y_0$ if and only if $x_n \neq x_0$. Thus

$$x_n = f^{-1}(y_n) = \frac{1}{f(x_n) - f(x_0)/x_n - x_0} = \frac{1}{f'(x_0)}.$$

□

Theorem 24 (Calculus Rules)**.** If $f'(x_0)$ and $g'(x_0)$ exist, then

- (*Addition Rule*):
$$(f+g)'(x_0) = f'(x_0) + g'(x_0)$$
.
- (*Product Rule*):
$$(fg)'(x_0) = f'(x_0)g(x_0) + g'(x_0)f(x_0).$$
- (*Quotient Rule*):
$$\left(\frac{f}{g}\right)'(x_0) = \frac{f'(x_0)g(x_0) - g'(x_0)f(x_0)}{[g(x_0)]^2}.$$

Proof. These can be easily proved using the sequential definition for derivatives. For example:

$$\begin{aligned}\lim_{x\to x_0} \frac{f(x)g(x) - f(x_0)g(x_0)}{x - x_0} &= \frac{f'(x_n)g(x_n) - f(x_n)g(x_0) + f(x_n)g(x_0) - f(x_0)g(x_0)}{x_n - x_0} \\ &= f(x_n)\frac{g(x_n) - g(x_0)}{x_n - x_0} + \frac{f(x_0) - f(x_0)}{x_n - x_0}g(x_0) \\ &= f'(x_0)g(x_0) + g'(x_0)f(x_0)\end{aligned}$$

□

Theorem 25 (Chain Rule)**.** Let $f : (a,b) \to \mathbb{R}$, where $f((a,b)) \subset (c,d)$, and $g : (c,d) \to \mathbb{R}$, and say that $f'(x_0)$ and $g'(x_0)$ exist. Then $g \circ f$ is defined in (a,b), and

$$(g \circ f)'(x_0) = g'(f(x_0)) \cdot f'(x_0).$$

Proof. We have that $f(x_0+\Delta x) = f(x_0) + A\Delta x + r(\Delta x)$, where $\frac{|r(\Delta x)|}{|\Delta x|} \to 0$, and $f(x_0) = y_0$. Similarly, $g(y_0+\Delta y) = g(y_0) + B\Delta y + \rho(\Delta y)$. Thus

$$\begin{aligned}(g\circ f)(x_0+\Delta x) - (g\circ f)(x_0) &= g(f(x_0+\Delta x)) - g(f(x_0)) \\ &= g(f(x_0) + A\Delta x + r(\Delta x)) - g(f(x_0)) \\ &= g(f(x_0)) + BA\Delta x + Br(\Delta x) + \rho(A\Delta x + r(\Delta x)) - g(f(x_0)) \\ &= BA\Delta x + R(\Delta x), \quad (R(\Delta x) = Br(\Delta x) + \rho(A\Delta x + r\Delta x))\end{aligned}$$

Then we can easily see that

$$\frac{|R(\Delta x)|}{|\Delta x|} \to 0,$$

so $g\circ f$ is differentiable by definition. □

Theorem 26 (Generalized Mean Value Theorem)**.** Let $f, g \in C([a,b])$, and assume $f'(x)$ and $g'(x)$ exist for all $x \in (a,b)$. Then

$$\frac{f(b)-f(a)}{b-a} = \frac{f'(\xi)}{g'(\xi)},$$

for some $0 < \xi < x$.

We can use this to prove another famous Calculus theorem.

Theorem 27 (L'Hôpital's Rule)**.** Let $f : (a,b) \to \mathbb{R}$, $a, b \in \mathbb{R} \cup (-\infty, \infty)$, and $g : (a,b) \to \mathbb{R}$. Also assume that $f'(x)$ and $g'(x)$ exist for all $x \in (a,b)$, and that $g(x) \neq 0$ for $x \in V_\epsilon(0)$. Suppose $f(x), g(x) \to \alpha$, and $x \to \alpha^+$, where α is either 0 or $\pm\infty$, and suppose that

$$\frac{f'(x)}{g'(x)} = c \in \mathbb{R} \cup \{\pm\infty\}, \text{ as } x \to a^+.$$

Then

$$\frac{f(x)}{g(x)} \to c, \text{ as } x \to a^+.$$

Proof. There are many cases to prove, but we only show one: $a \in \mathbb{R}$, $\alpha = 0$, and c is anything. Define by continuity, $f(0) = 0$, $g(0) = 0$, and $f, g : [a, b/2] \to \mathbb{R}$. By the GMVT,

$$\frac{f(x)}{g(x)} = \frac{f(x)-f(0)}{g(x)-g(0)} = \frac{f'(\xi)}{g'(\xi)}, \quad \text{for some } 0 < \xi < x.$$

We want to evaluate $\lim_{x\to 0^+} \frac{f(x)}{g(x)}$. Given $\epsilon > 0$, there exists $0 < z \le \delta$ such that

$$\left|\frac{f'(z)}{g'(z)} - c\right| \le \epsilon.$$

Then
$$\left|\frac{f(x)}{g(x)} - c\right| = \left|\frac{f'(\xi)}{g'(\xi)} - c\right| \le \epsilon,$$
if $0 \le \xi \le x \le \delta$, as desired. □

3.3 Differentials

Suppose $x \in \mathbb{R}$ (not a function), and fix $x_0 \in \mathbb{R}$. Then

$$dx = (dx)_{x_0} = x - x_0,$$

and so dx is just a number. If f is a fucntion, assume that $f'(x_0)$ exists. Then

$$df = (df)_{x_0} = (df)(x_0, dx) = f'(x_0)\,dx.$$

So $df|_{x_0}$ is a function of dx. Now suppose that $f^{(n)}(x_0)$ exists. Then

$$(d^n f)_{x_0} = f^{(n)}(x_0)(dx)^n.$$

Thus,

- $df = f'(x_0)\,dx$, which is linear in dx,
- $d^2 f = f''(x_0)(dx)^2$, which is quadratic in dx,
- $d^3 f = f'''(x_0)(dx)^3$, which is cubic in dx,
- and so on.

Theorem 28 (Invariance of df)**.** Let f be a function, and assume $f'(x_0)$ exists. Then,

- If x is not a function, and $x \in \mathbb{R}$, then
$$df = f'(x_0)\,dx = f'\Big|_{x_0} dx.$$
- If x is a function, $x(t)$, and $x(t_0) = x_0$, then
$$df = d(f(x(t)) = f'\Big|_{x_0} x'\Big|_{t_0} dt = f'\Big|_{x_0} dx.$$

Notice that either way, df turns into the same form. That is, for x being a variable and x being a function,

$$df = f'(x)\,dx.$$

This is only true for df. We can check this with $d^2 f$.

- For x not a function,

$$d^2 f = f''(x_))(dx)^2 = f''\Big|_{x_0} (dx)^2.$$

- For x being a function,

$$\begin{aligned} d^2 f &= d^2(f(x(t)) = [f(x(t))]''(dt)^2 \\ &= (f'x')'(dt)^2 \\ &= (f''x'x' + f'x'')\,dt\,dt \\ &= f''\Big|_{x_0} (x'dt)^2 + f'x''(dt)^2 \\ &= f''\Big|_{x_0} (dx)^2 + f'\Big|_{x_0} x''\Big|_{t_0} dt^2 \end{aligned}$$

As we can obviously see, $d^2 f$ does not have this invariance property. Differentials follow the basic algebraic properties that we expect them to:

- $d(fg) = (df)g + g(df)$,
- $d(cf) = c\,df$, for $c \in \mathbb{R}$,
- $d(\frac{f}{g}) = \frac{(df)g - f(dg)}{g^2}$.

Example 4. Compute f' for $f(x) = \exp(\sin(x^2 + \ln x))$.

Solution. We have

$$\begin{aligned} df &= d(e^z) = e^z\,dz = e^{\sin(x^2+\ln x)} d(\sin(x^2 + \ln x) \\ &= e^{(\cdots)} d(\sin u) = e^{(\cdots)} \cos u\,du \\ &= e^{(\cdots)} \cos(\cdots) d(x^2 + \ln x) \\ &= e^{(\cdots)} \cos(\cdots) \left[2x\,dx + \frac{1}{x}\,dx\right] \\ &= \boxed{e^{\sin(x^2+\ln x)} \cos(x^2 + \ln x)\left(2x + \frac{1}{x}\right) dx} \end{aligned}$$

□

We can rigorously define differentials as maps. For a fixed point $a \in \mathbb{R}$, the *tangent space* at a, denoted $T_a\mathbb{R} = T_a$, is the line $\mathbb{R}$ with the origin at a. More formally, T_a is a vector space

$$T_a = \{a\} \times \mathbb{R},$$

with the following operations:

$$(a,h_1)+(a,h_2)=(a,h_1+h_2), \quad C(a,h)=(a,Ch).$$

Usually we will just denote this as $h_{1,2} \in T_a$, $h_1+h_2 \in T_a$, and $Ch \in T_a$. Using this idea, we can now actually define what a "differential" really is.

Definition 37 (Differential)**.** Suppose f is differentiable at a. The *differential* of f at a, $(df)_a$, is the linear mapping between the corresponding tangent spaces

$$\begin{aligned}(df)_a : \quad & T_a \to T_{f(a)} \\ & h \mapsto f'(a)h.\end{aligned}$$

Directly from this, the chain rule $(f\circ g)'|_a = f'|_{g(a)}g'|_a$ takes the form

$$d(f\circ g)_a = (df)_{g(a)} \circ (dg)_a,$$

so that $d(f\circ g)_a(h) = f'|_{g(a)}\cdot g'|_a\cdot h$. Then going back to the main formula for the differential of a function $f = f'dx$, we can see that this is actually the identity map $\mathrm{id}_{\mathbb{R}} : x \to x$, which is differentiable at a with $x' = 1$. Its differential will be

$$\begin{aligned}(dx)_a : \quad & T_a \to T_a \\ & h \mapsto h.\end{aligned}$$

Invoking the chain rule, we get

$$(df)_a = (df)_a \circ d(\mathrm{id}_{\mathbb{R}})_a = f'(a)(dx)_a.$$

This gives us the familiar form that we expect; that is,

$$(df)_a = f'(a)(dx)_a.$$

Then for any $h \in T_a$,

$$(df)_a(h) = f'(a)\cdot((dx)_a(h)) = f'(a)h.$$

Tangent spaces are a little preview for how real analysis ties into differential geometry, and we shall explore more of those ideas in a later chapter.

4 Riemann Integration

4.1 Introduction

Definition 38 (Partition)**.** A *partition* of $[a,b]$ is any $P = \{x_0, \dots, x_n\}$, where $a = x_0 < x_1 < \cdots < x_{n-1} < x_n = b$.

We denote the set of all partitions as

$$\mathcal{P}_{[a,b]} = \{P \mid P \text{ is a partition of } [a,b]\}.$$

Definition 39 (Gap)**.** The *gap* of a partition P is defined as $\max |x_{j-1} - x_j|$.

Definition 40 (Riemann Sums)**.** Let $f : (a,b) \to \mathbb{R}$. Then for all $P \in \mathcal{P}_{[a,b]}$, the *lower Riemann sum* is defined to be

$$L(f,P) = \sum_{j=1}^{n} m_j |x_j - x_{j-1}|, \quad \text{where } m_j = \inf_{[x_{j-1},x_j]} f.$$

Similarly, the *upper Riemann* sum is defined to be

$$U(f,P) = \sum_{j=1}^{n} M_j |x_{j-1} - x_j|, \quad \text{where } M_j = \sup_{[x_{j-1},x_j]} f.$$

Because we know that $m_j \le M_j$ for all j, it must be true that

$$L(f,P) \le U(f,P), \quad \forall f, \forall P \in \mathcal{P}_{[a,b]}.$$

Note that if our function f is not bounded, then

$$\sup_{[a,b]} f = \infty \implies \forall P \in \mathcal{P}_{[a,b]}, \text{ we have some } \sup_{[x_{j-1},x_k]} f = \infty \implies U(f,P) = \infty, \forall P.$$

Thus $U(f,P)$ and $L(f,P)$ are only nontrivial for bounded functions, where

$$\sup_{[a,b]}|f| < \infty, \quad \inf_{[a,b]}|f| > -\infty.$$

Definition 41 (Lower Integral)**.** For all $f:[a,b]\to\mathbb{R}$, we define the *lower integral* as

$$\underline{\int_a^b} f = \sup\{L(f,P) \mid P \in \mathcal{P}_{[a,b]}\}.$$

Definition 42 (Upper Integral)**.** For all $f:[a,b]\to\mathbb{R}$, we define the *upper integral* as

$$\overline{\int_a^b} f = \inf\{U(f,P) \mid P \in \mathcal{P}_{[a,b]}\}.$$

Combining these definitions, we can formally define Riemann integration.

Definition 43 (Riemann Integral)**.** We say the function $f:[a,b]\to\mathbb{R}$ is *Riemann integrable* iff

$$\underline{\int_a^b} f = \overline{\int_a^b} f \in \mathbb{R}.$$

In this case, we call

$$\int_a^b f = \underline{\int_a^b} f = \overline{\int_a^b} f$$

the *Riemann integral* of f over $[a,b]$.

4.2 More on Riemann Integral

Definition 44 (Refinement)**.** A partition $P^* \in \mathcal{P}_{[a,b]}$ is a *refinement* of $P \in \mathcal{P}_{[a,b]}$ iff $P^* \supset P$; that is, if $P = \{a = x_0 < x_1 < \cdots < x_{n-1} < x_n = b\}$, then for all j, $P^* \cap [x_{j-1}, x_j]$ is a partition of $[x_{j-1}, x_j]$.

Theorem 29. Let $f : [a,b] \to \mathbb{R}$ be bounded. Then

(1) $L(f,P) \le L(f,P^*)$ and $U(f,P) \ge U(f,P^*)$,

(2) $L(f,P_1) \le U(f,P_2)$, for all $P_1, P_2 \in \mathcal{P}_{[a,b]}$,

(3) For all $P \in \mathcal{P}_{[a,b]}$,

$$L(f,P) \le \underline{\int_a^b} f \le \overline{\int_a^b} f \le U(f,P).$$

Proof. At once, (1) $\implies$ (2). Take P^*, a common refinement of P_1 and P_2 ($P^* = P_1 \cup P_2$). Then

$$L(f,P_1) \le L(f,P^*) \le U(f,P^*) \le U(f,P_2).$$

(2) $\implies$ (3) at once as well because

$$L(f,P_1) \le U(f,P_2).$$

On the left hand side, take $\sup_{P_1 \in \mathcal{P}_{[a,b]}} f$, which gives

$$\underline{\int_a^b} f \le U(f,P_2), \quad \forall P_2 \in \mathcal{P}_{[a,b]}.$$

On the right hand side, take $\inf_{P_2 \in \mathcal{P}_{[a,b]}} f$, which gives

$$\overline{\int_a^b} f \ge L(f,P_1), \quad \forall P_1 \in \mathcal{P}_{[a,b]}.$$

Hence we only need to prove (1). Notice that

$$L(f,P) = \sum_{j=1}^{n} \inf_{[x_{j-1},x_j]} f|x_{j-1} - x_j|.$$

For fixed j, $P^* \cap [x_{j-1}, x_j]$ is a partition of $[x_{j-1}, x_j]$. Because $y_{k-1}, y_k] \subset [x_{j-1}, x_j]$,

$$\inf_{[y_{k-1},y_k]} f \ge \inf_{[x_{j-1},x_j]} f \implies \sum_{k=1}^{n} \inf_{[y_{k-1},y_k]} f|y_{k-1} - y_k| \ge \sum_{k=1}^{n} \inf_{[x_{j-1},y_j]} f|y_{k-1} - y_k|$$

Then

$$\sum_{k=1}^{n} \inf_{[x_{j-1},y_j]} f|y_{k-1} - y_k| = \inf_{[x_{j-1},x_j]} f \sum_{k=1}^{m} |y_{k-1} - y_k| = \inf_{[x_{j-1},x_j]} f|x_{j-1} - x_j|.$$

Hence

$$L(f, P^* \cap [x_{j-1}, x_j]) \ge \sum_{j=1}^{n} \inf_{[x_{j-1},x_j} |.$$

Then the left hand side is equal to $L(f,P^*)$, and the right hand side is equal to $L(f,P)$. □

Consider the Dirichlet function,

$$f_D(x) = \begin{cases} 1, & x \in \mathbb{Q} \\ 0, & x \in \mathbb{R} \setminus \mathbb{Q} \end{cases}$$

on some interval $[a,b]$. We can tell that this is not Riemann integrable, because

$$L(f_D, P) = 0, \quad \forall P, \quad U(f_D, P) = 1, \quad \forall P,$$

and so

$$\underline{\int_a^b} f = 0 < 1 = \overline{\int_a^b} f.$$

We see that even though $f_D(x)$ is bounded, it does not mean it is Riemann integrable.

Theorem 30 (Archimedes-Riemann Theorem)**.** Let $f : [a,b] \to \mathbb{R}$ be bounded. Then

(1) f is Riemann integrable on $[a,b]$ if and only if there exists P_n, a sequence of partitions such that

$$U(f,P_n) - L(f,P_n) \to 0, \quad \text{as } n \to \infty,$$

(2) If the above condition holds, then for P_n, we have

$$\lim_{n\to\infty} U(f,P_n) = \lim_{n\to\infty} L(f,P_n) = \int_a^b f.$$

(3) The first statement is equivalent to:

$$\forall \epsilon > 0, \exists P_\epsilon \in \mathcal{P}_{[a,b]} \text{ such that } U(f,P_\epsilon) - L(f,P_\epsilon) < \epsilon.$$

Proof. We only prove the third statement. Assume $U(f,P_\epsilon) - L(f,P_\epsilon) < \epsilon$. Then for arbitrary $\epsilon > 0$,

$$L(f,P_\epsilon) \le \underline{\int_a^b} f \le \overline{\int_a^b} f \le U(f,P_\epsilon) \implies \overline{\int_a^b} f - \underline{\int_a^b} f < \epsilon.$$

□

Example 5. Prove that $\int_0^1 x^2$ exists and equals $\frac{1}{3}$.

Proof. We have that

$$P_n = \left\{ \frac{k}{n},\, k = 0, \ldots, n \right\},$$

and

$$M_{j+1} = \left(\frac{j+1}{n} \right)^2, \quad m_{j+1} = \left(\frac{j}{n} \right)^2.$$

Thus

$$L(f, P_n) = \sum_{j=1}^{n} \left(\frac{j-1}{n} \right)^2 \cdot \frac{1}{n} = \frac{1}{n^3} \sum_{j=1}^{n} (j-1)^2 = \frac{1}{n^3}(0^2 + 1^2 + \cdots + (n-1)^2.$$

$$U(f, P_n) = \frac{1}{n^3}(0^2 + 1^2 + \cdots + (n-1)^2 + n^2).$$

It follows that

$$0 \leq U(f, P_n) - L(f, P_n) = \frac{n^2}{n^3} = \frac{1}{n} \to 0 \implies x^2 \text{ is Riemann integrable.}$$

Then

$$\int_0^1 x^2 = \lim_{n \to \infty} U(f, P_n) = \lim_{n \to \infty} \frac{1}{n^3} \left(\frac{n(n+1)(2n+1)}{6} \right) = \lim_{n \to \infty} \frac{(1 + \frac{1}{n})(2 + \frac{1}{n})}{6} = \frac{1}{3}.$$

□

Theorem 31. If $f : [a, b] \to \mathbb{R}$ is continuous, then f is Riemann integrable.

Proof. Notice that f is uniformly continuous. We choose a partition P of $[a, b]$ with

$$|P| = \max_{k \in [1,n]} (x_k - x_{k-1}) \leq \frac{\delta}{2}.$$

Then

$$U(f, P) - L(f, P) = \sum_{k=1}^{n} (x_k - x_{k-1}) \left(\sup_{[x_{k-1}, x_k]} f - \inf_{[x_{k-1}, x_k]} f \right) < \epsilon |b - a|.$$

□

4.3 Integral Properties/Theorems

Theorem 32 (Properties of Integrals)**.** (1) For $c \in [a,b]$ and f Riemann integrable on $[a,b]$, then

$$\int_a^b f = \int_a^c f + \int_c^b f.$$

(2) If f, g are Riemann integrable, and $\lambda, \mu \in \mathbb{R}$, then

$$\int_a^b (\lambda f + \mu g) = \lambda \int_a^b f + \mu \int_a^b g.$$

(3) If $f \geq 0$ on $[a,b]$ and f is Riemann integrable, then

$$\int_a^b f \geq 0.$$

(4) If $m \leq f(x) \leq M$ on $[a,b]$, then

$$m(b-a) \leq \int_a^b f \leq M(b-a).$$

(5) If $f(x) \leq g(x)$ on $[a,b]$, then

$$\int_a^b f \leq \int_a^b g.$$

(6) If f is Riemann integrable, then $|f|$ is Riemann integrable as well, and

$$\left| \int_a^b f \right| \leq \int_a^b |f|.$$

Proof. Basic proofs using partitions. □

Theorem 33 (Mean Value Theorem)**.** If f is continuous on $[a,b]$, then there exists $c \in [a,b]$ such that

$$f(c) = \frac{1}{b-a} \int_a^b f.$$

Proof. Let $M = \max_{[a,b]} f = f(\beta)$, and $m = \min_{[a,b]} f = f(\alpha)$. Then

$$m \leq f(x) \leq M \implies f(\alpha) \leq \int_a^b \frac{1}{b-a} f \leq f(\beta).$$

By the Intermediate Value Theorem, there exists $c \in [\alpha, \beta]$ such that

$$f(c) = \frac{1}{b-a}\int_a^b f.$$

□

Theorem 34 (Fundamental Theorem of Calculus). (a) If $f : [a,b] \to \mathbb{R}$ is Riemann integrable, and $F : [a,b] \to \mathbb{R}$ satisfies $F'(x) = f(x)$ for all $x \in [a,b]$, then

$$\int_a^b f = F(b) - F(a).$$

(b) Let $g : [a,b] \to \mathbb{R}$ be Riemann integrable, and for $x \in [a,b]$, define

$$G(x) = \int_a^x g.$$

Then G is continuous on $[a,b]$. If g is continuous at some point $c \in [a,b]$, then G is differentiable at c and $G'(c) = g(c)$.

Proof. (a)

□

Theorem 35 (Integration By Parts). For f and g Riemann integrable,

$$\int_a^b f'g = fg\Big|_a^b - \int_a^b fg'.$$

4.4 Discontinuous Functions

We know that continuous functions are always Riemann integrable, but discontinuous functions like the Dirichlet function are not Riemann integrable. As it turns out, for example, the Heaviside function, defined as

$$H(x) = \begin{cases} 0, & x \leq 0 \\ 1, & x > 0 \end{cases},$$

is Riemann integrable on $[-1,1]$, despite being discontinuous.

Example 6. Prove that

$$\psi(x) = \begin{cases} \frac{1}{q}, & x \in \mathbb{Q}, \\ 0, & x \notin \mathbb{Q} \end{cases}$$

is Riemann integeable on $[0,1]$.

Proof. Let $\alpha \in \mathbb{Q} \cap [0,1]$. Then ψ is discontinuous at α. Let $y_n \to \alpha$, for $y_n \notin \mathbb{Q}$. Then $\psi(y_n) \not\to \psi(\alpha) = 0$. Let $\beta \in \mathbb{Q} \cap [0,1]$. Any sequence $\frac{p_n}{q_n} \in \mathbb{Q}$ where $\lim_{n\to\infty} \frac{p_n}{q_n} = \beta$ has $q_n \to \infty$. Thus

$$\lim_{n\to\infty} \psi\left(\frac{p_n}{q_n}\right) = 0 = \psi(\beta).$$

Thus $\psi(x)$ is Riemann integrable on $[0,1]$. □

Example 7. Prove that $f(x) = \sin\frac{1}{x}$ for $x > 0$ and $f(x) = 0$ for $x = 0$ is Riemann integrable on $[0,1]$.

Proof. Let $P = x_k$, with $0 \le k \le n$, be a partition $x_0 = 0 < x_1 = \epsilon < x_2 < \cdots < x_n = 1$. Then

$$\begin{aligned} U(f,P) - L(f,P) &= \epsilon\left(\sup_{[0,\epsilon]} f - \inf_{[0,\epsilon]} f\right) + U\left(f\Big|_{[\epsilon,1]}, P\right) - L\left(f\Big|_{[\epsilon,1]}, P\right) \\ &\le 2\epsilon + U\left(f\Big|_{[\epsilon,1]}, P\right) - L\left(f\Big|_{[\epsilon,1]}, P\right). \end{aligned}$$

Since the upper and lower Riemann sums are arbitrarily small, we proved that f is Riemann integrable. □

Theorem 36. Let $f : [a,b] \to \mathbb{R}$ be bounded, and suppose f has a finite set of discontinuities $D(f)$ in $[a,b]$. Then f is Riemann integrable.

Proof. Fix $\epsilon > 0$. Then choose a partition $P = x_k$ so that

$$\sum_{D(f)\cap[x_{k-1},x_k]\neq\emptyset} (x_{k-1}, x_k) < \epsilon.$$

Then

$$U(f,P) - L(f,P) = \sum_{D(f)\cap[x_{k-1},x_k]\neq\emptyset} + \sum_{D(f)\cap[x_{k-1},x_k]=\emptyset}.$$

Let $|f(x)| \le M$, $x \in [a,b]$. Then the first term equals $2M\epsilon$, and the second term represents where f is continuous. Hence f is Riemann integrable. □

Theorem 37. Any monotonic function is Riemann integrable.

Proof. Assume $f : [a,b] \to \mathbb{R}$ is non-decreasing; that is, $x \le y \implies f(x) \le f(y)$. let $P = x_k$, where $0 \le k \le n$, be a partition of $[a,b]$. Then

$$\sup_{[x_{k-1},x_k]} f(x) = f(x_k), \qquad \inf_{[x_{k-1},x_k]} f(x) = f(x_{k-1}).$$

Assume $|P| = \max_k(x_k - x_{k-1}) = \epsilon$. Then we have

$$U(f,P) - L(f,P) = \sum_{k=1}^{n}(x_k - x_{k-1})(f(x_k) - f(x_{k-1})) \leq \epsilon(f(b) - f(a)).$$

□

Definition 45 (Zero Content)**.** A subset $A \subseteq [a,b]$ has *zero content* if for all $\epsilon > 0$, there exists $I_1, \dots, I_n$, open finite intervals such that $A \subseteq I_1 \cup \cdots \cup I_n$, and $|I_1| + \cdots + |I_n| < \epsilon$.

Finite sets have zero content.

Theorem 38. If $D(f)$ has zero content, then f is Riemann integrable.

4.5 Change of Variables and Taylor Polynomials

Before, we have omitted the differential when discussing integrals. In actuality, we denote integrals as such:

$$\int_a^b f(x)\,dx.$$

With this, we can rewrite the Fundamental Theorem of Calculus as

$$\int_a^b df(x) = f(b) - f(a).$$

Theorem 39. Let $\varphi : [a,b] \to \mathbb{R}$ be differentiable, and $\varphi' \in C([a,b])$, with $\varphi([a,b]) \subset [c,d]$. Suppose $f \in C([c,d])$. Then

$$\int_a^b (f \circ \varphi) \cdot \varphi' = \int_{\varphi(a)}^{\varphi(b)} f.$$

In differential notation,

$$\int_a^b f(\varphi(x)) \cdot \varphi'(x)\,dx = \int_{\varphi(a)}^{\varphi(b)} f(y)\,dy.$$

Proof. Because all functions f, φ, φ' are continuous, the integrals exist. We want to prove that

$$\int_a^x (f \circ \varphi)\varphi' = \int_{\varphi(a)}^{\varphi(x)} f, \quad \forall x \in [a,b].$$

At $x = a$, the both the left and right hand side will be zero. Then for all $x \in [a,b]$, the derivative of the left hand side will be

$$f(\varphi(x))\varphi'(x),$$

while the derivative of the right hand side will be

$$\frac{d}{dx}\left[\int_{\varphi(a)}^{y} f\bigg|_{y=\varphi(x)}\right] = f(y)\bigg|_{y=\varphi(x)} \cdot \varphi'(x).$$

Thus by the Mean Value Theorem, the left and right hand side will be equal. □

Example 8. Compute $\int_2^3 e^{x^2} x\,dx$.

Solution. We have

$$\begin{aligned}\int_2^3 e^{x^2} x\,dx &= \int_2^3 e^{x^2} \frac{d(x^2)}{2}\\ &= \frac{1}{2}\int_4^9 e^t\,dt\\ &= \frac{1}{2}(e^9 - e^4)\end{aligned}$$

□

Theorem 40 (Taylor's Formula)**.** Let I be some $(a-\delta, a+\delta)$, for $\delta > 0$. Let $f: I \to \mathbb{R}$, and suppose $f', f'', \ldots, f^{(n+1)}$ exist and $f^{(n+1)} \in C(I)$. Then for all $b \in I$,

$$f(b) = f(a) + \frac{f'(a)}{1!}(b-a) + \frac{f''(a)}{2!}(b-a)^2 + \cdots + \frac{f^{(n)}(a)}{n!}(b-a)^n + R_n(b),$$

where the remainder is called *Taylor's polynomial*:

$$R_n(b) = \frac{1}{n!}\int_a^b (b-t)f^{(n+1)}(t)\,dt$$

Proof. We have that

$$\begin{aligned}f(b) = f(a) + \int_a^b f'(x)\,dx &= f(a) + \int_a^b f'(x)d(x-b)\\ &= f(a) + f'(x)(x-b)\bigg|_{x=a}^{x=b} - \int_a^b (x-b)d(f'(x))\\ &+ f(a) + \frac{f'(a)}{1!}(b-a) - \int_a^b \frac{f''(x)d(x-b)^2}{2}\\ &= \cdots\end{aligned}$$

and so on... □

We can use this, for example, to find an infinite series expansion for e. We know that $(e^x)^{(n)} = e^x$. So

$$e = e^1 = 1 + \frac{1}{1!} + \frac{1}{2!}1^2 + \cdots + \frac{1}{n!} + R_n(1) = 1 + \frac{1}{1!} + \frac{1}{2!} + \cdots + \frac{1}{n!} + R_n(1).$$

As n goes to infinity,

$$|R_n(1)| = \left| \frac{1}{n!} \int_0^1 (1-t)^n e^t \, dt \right| \le \frac{1}{n!} \int_0^1 |(1-t)^n| e^t \, dt \le \frac{1}{n!} \int_0^1 2^n c \, dt = \frac{c2^n}{n!} \to 0.$$

Thus

$$e = \sum_{n=0}^{\infty} \frac{1}{n!}.$$

4.6 Generalizations of Riemann Integral

We study two different methods of generalizing the normal Riemann integral to make it applicable in a wider variety of situations: the Henstock-Kurzweil Integral (or Generalized Riemann integral/GR-integral) and the Riemann-Stieltjes integral. They approach the problem in different ways: one generalizes the partition structure, and one generalizes the notion of length.

Henstock-Kurzweil Integral

Definition 46 (Tagged Partition)**.** Let $P = \{x_0, x_1, \ldots, x_n\}$ be a partition of $[a, b]$. A *tagged partition* is one where in addition to P, we have chose points c_k in each of the subintervals $[x_{k-1}, x_k]$.

Definition 47 (Generalized Riemann Sum)**.** Given a function $f : [a, b] \to \mathbb{R}$ and a tagged partition (P, c_k), the *Riemann sum* generated by this partition is given by

$$R(f, P) = \sum_{k=1}^{n} f(c_k)(x_k - x_{k-1}).$$

Evidently, we have

$$L(f, P) \le R(f, P) \le U(f, P).$$

Definition 48 (δ–fine)**.** Let $\delta > 0$. A partition P is *δ–fine* if every subinterval $[x_{k-1}, x_k]$ satisfies $x_k - x_{k-1} < \delta$. In other words, every subinterval has width less than δ.

Theorem 41 (Limit Criterion for Riemann Integrability)**.** A bounded function $f : [a,b] \to \mathbb{R}$ is Riemann-integrable with

$$\int_a^b f = A$$

if and only if for every $\epsilon > 0$, there exists $\delta > 0$ such that for any tagged partition (P, c_k) that is δ–fine, it follows that

$$|R(f, P) - A| < \epsilon.$$

If we allow δ to be a function of x, then we can generalize the Riemann integral.

Definition 49 (Gauge)**.** A function $\delta : [a,b] \to \mathbb{R}$ is called a *gauge* on $[a,b]$ if $\delta(x) > 0$ for all $x \in [a,b]$.

Definition 50 ($\delta(x)$–fine)**.** Given a gauge $\delta(x)$, a tagged partition (P, c_k) is $\delta(x)$*–fine* if every subinterval $[x_{k-1}, x_k]$ satisfies $x_k - x_{k-1} < \delta(c_k)$. In other words, every subinterval has width less than $\delta(c_k)$.

If $\delta(x)$ is constant, then we have the normal Riemann integral. It is when this is not constant, that we have found a way to describe measuring the fineness of partitions that is quite different. Consider the interval $[0,1]$. If $\delta(x) = 1/9$, we want to find a $\delta(x)$–fine partition of $[0,1]$. Take the partition $0 = x_0 < x_1 < \cdots < x_{18} = 1$ where

$$x_k = \frac{k}{18}, \quad k \in [1, 18].$$

We choose any $c_k \in [x_{k-1}, x_k]$, which gives us that $(P, \{c_k\})$ is a $\delta(x)$–fine partition of $[0,1]$

Theorem 42. Given a gauge $\delta(x)$ on an interval $[a,b]$, then there exists a tagged partition (P, c_k) that is $\delta(x)$–fine.

Using this idea, we introduce the GR-Integral (Generalized-Riemann Integral), which also has a special name.

Definition 51 (Henstock-Kurzweil Integral)**.** A function f on $[a,b]$ has *Henstock-Kurzweil integral* A if, for every $\epsilon > 0$, there exists a gauge $\delta(x)$

on $[a,b]$ such that for each tagged partition (P, c_k) that is $\delta(x)$–fine, it is true that

$$|R(f,P) - A| < \epsilon.$$

In this case, we write

$$A = \int_a^b f.$$

Theorem 43. If a function is Henstock-Kurzweil integrable, then the value of said integral is unique.

Proof. Let A_1 and A_2 be two values of the HK-integral of f on $[a,b]$. For each $\epsilon > 0$, there exists gauges $\delta_1(x)$ and $\delta_2(x)$ such that for every tagged partitions $(P_1, \{c_k\}_{k=1}^{n_1})$ and $(P_2, \{c_k\}_{k=1}^{n_2})$, we have

$$|R(f,P_1) - A_1| < \frac{\epsilon}{2}, \quad |R(f,P_2) - A_2| < \frac{\epsilon}{2}.$$

Consider the gauge $\delta(x) = \min\{\delta_1(x), \delta_2(x)\}$. Then there exists a tagged partition $(P, \{c_k\}_{k=1}^{n}$ that is $\delta(x)$–fine. Hence

$$|R(f,P) - A_1| < \frac{\epsilon}{2}, \quad |R(f,P) - A_2| < \frac{\epsilon}{2},$$

and so

$$|A_1 - A_2| \leq |A_1 - R(f,P)| + |R(f,P) - A_2| < \epsilon.$$

Thus, $A_1 = A_2$. □

It should be obvious to see that

$$\{\{\text{Riemann integrable}\} \implies \text{Henstock-Kurzweil integrable}\}.$$

We just take $\delta(x) = \delta$, and so the tagged partition $(P, \{c_k\})$ will be δ–fine. The converse is not true. Indeed, Dirichlet's function, while not Riemann integrable, is Henstock-Kurzweil integrable, and

$$\int_0^1 f_D(x) = 0.$$

Proof. Let $\epsilon > 0$. We want to construct a gauge $\delta(x)$ on $[0,1]$ such that whenever $(P, \{c_k\}_{k=1}^{n})$ is $\delta(x)$-fine, it follows that

$$0 \leq \sum_{k=1}^{n} f_D(c_k)(x_k - x_{k-1}) < \epsilon.$$

The gauge represents a restriction of the size of $\Delta x_k = x_k - x_{k-1}$ in the sense that $\Delta x_k < \delta(c_k)$. The Riemann sum will consist of products of the form $f_D(c_k)\Delta x_k$.

Then for irrational tags, $f_D(c_k) = 0$, so we have nothing to worry about. We just need to ensure that if c_k is rational, it comes from a suitably thin interval.

Let $\{r_1, r_2, \dots\}$ be an enumeration of the countable set of rational numbers in $[0,1]$. For each r_k, let $\delta(r_k) = \epsilon/2^{k+1}$. For $x \notin \mathbb{Q}$, set $\delta(x) = 1$. Hence we have constructed a gauge that meets our requirements. $\square$

All of our other integral properties and theorems, for example the Fundamental Theorem of Calculus or Change of Variables, will also work for HK-integrable functions.

Riemann-Stieltjes Integral

Definition 52 (α–length)**.** Let I be a bounded interval, and let $\alpha : X \to \mathbb{R}$ be a function where $I \subset X$. Then the *α–length* $\alpha[I]$ of I is defined to be:

$$\alpha[I] = \begin{cases} 0, & \text{if } I \text{ is a point or } \emptyset \\ \alpha(b) - \alpha(a), & \text{if } I \text{ is one of } [a,b], (a,b), (a,b], [a,b) \end{cases}$$

As an example, for $\alpha : \mathbb{R} \to \mathbb{R}$ given by $\alpha(x) = x^2$, we have that

$$\alpha[[2,3]] = \alpha(3) - \alpha(2) = 9 - 4 = 5, \quad \alpha[(-3,-2)] = -5.$$

As it turns out, our intuitive sense of what defines "length" is a special case of α–length if we take $\alpha(x) = x$, the identity function.

Theorem 44. Let I be a bounded interval, and let $\alpha : X \to \mathbb{R}$ be a function defined on some domain where $I \subset X$. Let $P \in \mathcal{P}_I$ be a partition of I. Then

$$\alpha[I] = \sum_{j \in P} \alpha[j].$$

Definition 53 (Riemann-Stieltjes Integral)**.**

5 Series

5.1 Convergence Tests

Definition 54 (Partial Sum)**.** Given a pair of sequences S_n and a_n, where $S_n = a_1 + \cdots + c_n$, we call S_n the *partial sum*.

In general, for any number $N \in \mathbb{N}$,

$$\sum_{n=1}^{\infty} a_n \text{ converges} \implies \sum_{n=N}^{\infty} a_n \text{ converges.}$$

The proof of this is trivial; if $S_m = a_1 + \cdots + a_m$, and $\tilde{S}_m = a_N + a_{N+1} + \cdots + a_m = S_m - (a_1 + \cdots + a_{N-1})$, which is $S_m - k$, a constant, which is convergent.

Theorem 45. If $\sum_{n=1}^{\infty} a_n$ converges, then $\lim_{n\to\infty} a_n = 0$.

Proof. Let $S_n = 1_1 + \cdots + a_n$. Then $S_n \to L \in \mathbb{R}$. Let $R_n = S_{n+1} = a_1 + \cdots + a_n + a_{n+1}$. Then $R_n \to L$. Hence $R_n - S_n = a_{n+1} \to 0$. □

The converse is false. The harmonic series,

$$\sum_{n=1}^{\infty} \frac{1}{n} = 1 + \frac{1}{2} + \frac{1}{3} + \cdots$$

has $\lim_{n\to\infty} \frac{1}{n} = 0$, but it does not converge.

For series with positive terms, meaning $a_n \geq 0$ for all n, we have that $S_{n+1} \geq S_n$. Hence the partial sums are monotonically increasing. So for these types of series:

Theorem 46. *Let $a_n \geq 0$. Then $\sum_{n=0}^{\infty} a_n$ converges if and only if S_n is bounded.*

This leads to one of the most common convergence tests:

Theorem 47 (Comparison Test)**.** Let $0 \leq a_n \leq b_n$, for all n. Then

$$\sum_{n=0}^{\infty} b_n \text{ converges} \implies \sum_{n=0}^{\infty} a_n \text{ converges.}$$

Proof. Let $S_n = a_1 + \cdots + a_n$ and $R_n = b_1 + \cdots + b_n$. We know $\sum_{n=1}^{\infty} b_n = L < \infty$. Therefore, because $b_n \geq 0$ for all n, $R_n \leq L$ for all n. Since we assumed that $S_n \leq R_n \leq l$, S_n converges. □

When does the series

$$\sum_{n=1}^{\infty} \frac{1}{n^{\alpha}},$$

for fixed α, converge? For $\alpha \leq 0$, this will diverge, because it fails the divergence test. We call this the p–series. Then we use another convergence test to figure out its convergence for other α.

Theorem 48 (Integral Test)**.** Let $f : [1, \infty) \to \mathbb{R}$ satisfy (i) $f \geq 0$ and (ii) f is non-strictly decreasing. Let $a_k = f(k)$, for all k. Then

$$\sum_{n=1}^{\infty} a_n \text{ converges} \iff \lim_{N\to\infty} \int_1^N f(x)\,dx < \infty.$$

Proof. $(\Longrightarrow)$: By monotonicity, S_n is bounded. Then

$$\int_1^N f(x)\,dx = \int_1^2 f\,dx + \cdots + \int_{n-1}^{n} f\,dx \leq a_1 + a_2 + \cdots + a_{n-1} \leq M,\ \forall n.$$

Hence $\int_1^N f(x)\,dx$ is increasing and bounded.
$(\Longleftarrow)$: Trivial. □

Thus, we use the integral test on the p–series, where $f(x) = \frac{1}{x^{\alpha}}$ for $x \geq 1$. Both conditions hold, so

$$I_n = \int)1^N \frac{dx}{x^{\alpha}} = \frac{1}{1-\alpha} x^{1-\alpha}\Big|_1^N = \frac{1}{1-\alpha} \cdot [N^{1-\alpha} - 1].$$

So $\lim_{n\to\infty} I_n$ is finite only if $\alpha > 1$. Thus

$$\sum_{n=1}^{\infty} \frac{1}{n^{\alpha}}$$

converges for $\alpha > 1$, and diverges for $\alpha < 1$. At $\alpha = 1$,

$$\int_1^N \frac{dx}{x} = \ln N \xrightarrow{n\to\infty} \infty.$$

Thus the p–series converges only for α strictly greater than 1.

Example 9. Prove that $\sum_{n=3}^{\infty} \frac{1}{n \ln n}$ converges.

Solution. Since $\ln n \geq 1$, $\frac{1}{n} \geq \frac{1}{n \ln n}$. We use the Integral Test on $f(x) = \frac{1}{x \ln x}$ for $x \geq 3$. Since this is positive and monotone, we can then find that

$$\lim_{N \to \infty} \int_3^N \frac{1}{x \ln x} \, dx$$

can be computed to find a numerical solution. □

Theorem 49 (Leibniz Alternating Series Test)**.** Suppose $\lim_{n \to \infty} a_n = 0$ and a_n is monotone decreasing. Then $a_1 - a_2 + a_3 - a_4 + \cdots = \sum_{n=1}^{\infty} (-1)^{n+1} a_n$ converges.

For example, consider the alternating Harmonic series,

$$\sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n}.$$

This converges, despite the fact that the non-alternating Harmonic series does not.

Definition 55 (ϵ-Definition of Convergence)**.** A series $\sum_{n=1}^{\infty} a_n$ converges to some $S \in \mathbb{R}$ if and only if

$$\forall \epsilon >, \exists N \text{ such that } \forall n \geq N, |S_n - S| < \epsilon.$$

In sequences, recall that the following is true:

$$a_n \text{ converges } \implies |a_n| \text{ converges.}$$

In series, the opposite is true:

$$\sum_{n=1}^{\infty} |a_n| \text{ converges } \implies \sum_{n=1}^{\infty} a_n \text{ converges.}$$

Definition 56 (Cauchy Sequence)**.** A sequence a_n is *Cauchy* iff

$$\forall \epsilon > 0, \exists N \text{ such that if } n, m \geq N, \text{ then } |a_n - a_m| < \epsilon.$$

Theorem 50 (Cauchy Convergence Criterion)**.** A sequence a_n converges to $L \in \mathbb{R}$ if and only if a_n is a Cauchy sequence.

Proof. $(\implies)$: By definition, if a_n converges to L, then

$$\forall \epsilon > 0, \exists N \text{ such that if } n \geq N, \text{ then } |a_n - L| < \epsilon.$$

Then slightly modifying this,

$$n, m \geq N \implies |a_n - a_m| = |a_n - L + L - a_m| \leq |a_n + L| + |a_m - L| < 2\epsilon.$$

$(\Longleftarrow)$: Take $\epsilon = 1$ in the Cauchy definition. Then there exists some N_1 such that if $n, m \geq N_0$, then $|a_n - a_m| \leq 1$. Then for all $n \geq N_0$,

$$|a_n| = |a_n - a_{N_0} + a_{N_0}| \leq |a_n - a_{N_0}| + |a_{N_0}| \leq 1 + |a_{N_0}|.$$

Therefore for all n,

$$|a_n| \leq \max\{|a_1|, |a_2|, \ldots, |a_{N_0-1}|, |a_{N_0}|\}.$$

Thus $|a_n|$ is bounded. Now by Bolzano-Weierstrass, there exists a subsequence $b_k = a_{n_k}$ such that $b_k \to L \in \mathbb{R}$ as $k \to \infty$. Now take $\epsilon > 0$. Then

$$\exists N_\epsilon \text{ such that if } n, m \geq N_\epsilon, \text{ then } |a_n - a_m| < \epsilon,$$

$$\exists M_\epsilon \text{ such that if } m \geq M_\epsilon, \text{ then } |b_m - L| < \epsilon.$$

But $b_m = a_{n_m}$. We find m_0 such that $n_{m_0} \geq N_\epsilon$ and $m_0 \geq M_\epsilon$. Then for all $n \geq N_\epsilon$, we have

$$|a_n - L| = |a_n - b_{m_0} + b_{m_0} - L| \leq |a_m - b_{m_0}| + |b_{m_0} - L| < 2\epsilon.$$

□

Definition 57 (Absolute/Conditional Convergence)**.** A series $\sum_{n=1}^{\infty} a_n$ converges *absolutely* iff $\sum_{n=1}^{\infty} |a_n|$ converges. A series $\sum_{n=1}^{\infty} a_n$ converges *conditionally* iff $\sum_{n=1}^{\infty} a_n$ converges, but $\sum_{n=1}^{\infty} |a_n|$ does not.

5.2 Special Topics

Definition 58 (Rearrangement)**.** Let $\sigma : \mathbb{N} \to \mathbb{N}$ be a bijection. The series $\sum_{n=1}^{\infty} b_n = \sum_{n=1}^{\infty} a_{\sigma(n)}$ is called a *rearrangement* (or *permutation*) of the series $\sum_{n=1}^{\infty} a_n$.

The main question we want to ask is if $\sum_{n=1}^{\infty} a_n = S$, then what is $\sum_{n=1}^{\infty} a_{\sigma(n)}$?

Theorem 51. Assume $\sum_{n=1}^{\infty} a_n$ converges to some $S \in \mathbb{R}$ absolutely. Then for all $\sigma : \mathbb{N} \to \mathbb{N}$, $\sum_{n=1}^{\infty} a_{\sigma(n)} = S$.

Proof. We want to estimate

$$\left|\sum_{k=1}^{n} a_{\sigma(k)} - S\right|, \text{ for large } n.$$

We have that

$$\sigma^{-1}(\{1,\ldots,N\}) = \{\sigma^{-1}(1),\ldots,\sigma^{-1}(N)\}.$$

Let $k = \max\{\sigma^{-1}(1),\ldots,\sigma^{-1}(N)\}$. Then $\{1,\ldots,N\} \subset \{\sigma^{-1}(1),\ldots,\sigma^{-1}(N)\}$. For $m \geq k$,

$$\left|\sum_{n=1}^{m} a_{\sigma(n)} - S\right| = \left|\sum_{j=1}^{N} a_j + \sum_{j\geq N} a_j - S\right| \leq \left|\sum_{j=1}^{N} a_j - S\right| + \left|\sum_{j\geq N} a_j\right| \leq \epsilon + \sum_{j\geq N} |a_j|$$

$$\leq \sum_{j=N}^{\infty} |a_j| + \epsilon < 2\epsilon.$$

□

Theorem 52 (Riemann Rearrangement Theorem)**.** Let $\sum_{n=1}^{\infty} a_n$ converge conditionally. Let $L \in \mathbb{R} \cup \{\pm\infty\}$. Then there exists $\sigma : \mathbb{N} \to \mathbb{N}$ such that

$$\sum_{n=1}^{\infty} a_{\sigma(n)} = L.$$

Proof. Let b_n be the sequence containing all a_n such that $a_n \geq 0$, preserving order. Let c_n be the sequence containing all a_n such that $a_n \leq 0$, preserving the order. We claim that

$$\sum_{n=1}^{\infty} b_n = \sum_{n=1}^{\infty} c_n \to \infty.$$

They can't converge, because $\sum_{n=1}^{\infty} a_n$ converges conditionally. Let $S_n = \sum_{k=1}^{n} |a_k|$. Then for all n,

$$0 \leq S_n = |a_1| + \cdots + |a_n| = \sum_{k} b_k + \sum_{k} c_k \leq \sum_{k=1}^{\infty} b_k + \sum_{k=1}^{\infty} |c_k| = c < \infty.$$

Thus $0 \leq S_1 \leq \cdots \leq S_n \leq C$, for all n. Thus there exists some $\lim_{n\to\infty} S_n < \infty$, so $\sum_{n=1}^{\infty} |a_n|$ converges, a contradiction. Now assume that $\sum_{n=1}^{\infty} c_n = -\infty$, and $\sum_{n=1}^{\infty} b_n < \infty$. Denote R_n and T_n as the partial sums respectively. Then $T_n \to -\infty$, $R_n \to c$, so

$$\sum_{k=1}^{n} a_k = Tm_1 + R_{m_1} \leq T_{m_1} + c.$$

As $m,n \to \infty$, $T_m \to -\infty$, and $\sum_{n=1}^{\infty} a_n = -\infty$, a contradiction. Hence $\sum_{n=1}^{\infty} b_n = \infty$ and $\sum_{n=1}^{\infty} c_n = -\infty$, for $b_n \geq 0$ and $c_n < 0$. If we fix any $L > 0, L \in \mathbb{R}$, and fix $\epsilon > 0$. We find N_ϵ such that for $n \geq N$, $|b_n|,|c_n| \leq \epsilon$. Then there exists a unique n_1 such that $b_1 + \cdots + b_{n_1} \leq L$, and there exists a unique m_1 such that $b_1 + \cdots + b_{n_1} + c_n + \cdots + c_{m_1} \geq L$. We repeat this process until we form a new sequence $(b_1, \dots, b_{n+1}, c_n, \dots, c_{m_1}, b_{n_1+2})$ which is a rearrangement of a_n under some σ. Denote $a_{\sigma(n)}$ as $\tilde{S}_n$. We take M such that for$N \geq M$, $\tilde{S}_n$ contains all $b_1, \dots, b_{N_\epsilon}, c_1, \dots, c_{N_\epsilon}, \dots$, so

$$|\tilde{S}_n - L| \leq \sup_{k \geq n \geq M} |a_{\sigma(k)}| \leq \sup_{j \geq N_\epsilon} |b_j|, |c_j| < \epsilon.$$

□

Suppose a series $\sum_{n=1}^{\infty} a_n$ converges absolutely. Denote

$$a_n^+ = \max\{0, a_n\}, \quad a_n^- = \min\{0, a_n\}.$$

Obviously $\sum_{n=1}^{\infty} a_n^+$ converges to some $P \in \mathbb{R}$, and $\sum_{n=1}^{\infty} a_n^-$ converges to some $N \in \mathbb{R}$. But does $\sum_{n=1}^{\infty} a_n = P + N$?

Definition 59 (Regrouping)**.** The *regrouping* of the series $\sum_{n=1}^{\infty}$ is another series $\sum_{n=1}^{\infty} b_n$ such that $b_1 = (a_1 + \cdots + a_{n_1})$, $b_2 = a_{n_{j+1}} + \cdots + a_{n_2}$, with $n_1 < n_2 < \cdots$.

Theorem 53. Suppose $\sum_{n=1}^{\infty} a_n = S \in \mathbb{R}$. Then any regrouping of this series will also equal S.

Proof. Let $\sum_{n=1}^{\infty} b_n \sim S_n$, and $\sum_{n=1}^{\infty} a_n \sim R_n$. Then

$$S_m = b_1 + \cdots + b_m = a_1 + \cdots + a_{n_m} = R_{n_m} \implies S_{m_1} \text{ is a subsequence of } R_n.$$

Then if R_n converges, that means S_m converges to the same limit. □

6 Elementary Functions

6.1 Exponential/Logarithm Function

We know that $x \mapsto x^{1/n}$ for $(0,\infty) \to (0,\infty)$ and $n \in \mathbb{N}$ is the Inverse Function Theorem applied to $x \mapsto x^n$. This means

$$(x^{1/n})^n \equiv x, \quad (x^n)^{1/n} \equiv x.$$

We can extend this to $n \in \mathbb{Z}$ by saying that if $n < 0$, then

$$x^n = \frac{1}{x^{(-n)}}, \quad x^{1/n} = \frac{1}{x^{(-1/n)}}.$$

We can then extend to $r = \frac{p}{q} \in \mathbb{Q}$ where $x \mapsto (x^{1/q})^p$. We can use the properties of $x \cdots x$ and its inverse $x^{1/n}$ to derive the usual properties of x^r for $x > 0$:

- $x^{r_1} x^{r_2} = x^{r_1 + r_2}$,
- $(x^{r_1})^{r_2} = x^{r_1 r_2}$.

Theorem 54. Define $F : (0,\infty) \to \mathbb{R}$ as

$$x \mapsto \int_1^x \frac{1}{t}\,dt.$$

Then

(1) F is infinitely differentiable, strictly increasing, and $F(x) \to -\infty$ as $x \to 0^+$, $F(x) \to \infty$ as $x \to \infty$.

(2) For all $a, b > 0$, $F(ab) = F(a) + F(b)$.

(3) For all $a > 0$ and $r \in \mathbb{Q}$, $F(a^r) = rF(a)$.

Proof. (1) Follows immediately by FTC and properties of differentiable functinos.

(2) Without loss of generality, assume $0 < a < b$. Then

$$F(ab) = \int_0^{ab} \frac{dx}{x} = \left(\int_1^a \frac{dx}{x} + \int_a^{ab} \frac{dx}{x} \right) = F(a) + \int_a^{ab} \frac{d(xa)}{xa} = F(a) + F(b).$$

(3) Consider $h : (0,\infty) \to \mathbb{R}$ given by $x \mapsto F(x^r) - rF(x)$. Then $h(1) = 0$ and

$$h'(x) = F'\Big|_{x^r} rx^{r-1} - rF'\Big|_{x} = \frac{1}{x^r} rx^{r-1} - \frac{r}{x} = 0.$$

By MVT, $h(x) = 0$, for all x. By the Proposition, $(0,\infty) \overset{F}{\longleftrightarrow} (-\infty,\infty)$, and since F is strictly increasing, so is $\frac{1}{F}$ on $\mathbb{R}$. Indeed, consider $a, b \in \mathbb{R}$ such that $a < b$. Find $x, y > 0$ such that $F(x) = a$, $F(x) = b$, and so $x = F^{-1}(a)$ and $y = F^{-1}(b)$. Now $a < b \iff F(x) < F(y)$ must have $x < y$, because otherwise $x \geq y \implies F(x) \geq F(y)$, a contradiction. But $x = F^{-1}(a)$, $y = F^{-1}(b)$, so $a < b \iff F^{-1}(a) < F^{-1}(b)$.

□

Definition 60 (Exponential/Logarithm Function)**.** For all $x > 0$, we define $F(x) = \ln x$, and for all $x \in \mathbb{R}$, define $F^{-1}(x) = \exp(x)$, or e^x.

Some notes:

- There exists a unique number, called e, such that $\ln e = 1$ by the IVT.
- Both $\ln x$ and $\exp(x)$ are infinitely differentiable on their respective domains.
- $\ln(\exp(x)) \equiv x$, for all $x \in \mathbb{R}$.

The derivatives of these functions will be

$$(\ln x)' = \frac{1}{x}, \quad x > 0 \text{ (by FTC)}.$$

$$(\ln(\exp x))' = \ln'\Big|_{\exp x} \cdot (\exp x)' \equiv 1 \implies (\exp x)' = \exp x.$$

Finally, we can use $\ln x$, $\exp x$, and a^r for $r \in \mathbb{Q}$ to define exponents for real numbers; that is, a^x for $x \in \mathbb{R}$. We know

$$a^r = \exp(\ln(a^r)) = \exp(r \ln a).$$

If $a > 0$, and $x \in \mathbb{R}$, we define $a^x = \exp(x \ln a)$. a^x is also infinitely differentiable. Some properties of real exponents:

- If $a = e$, then $e^x = \exp(x \cdot 1) = \exp(x)$.
- $(a^x)' = \exp(x \ln a)' = \exp(\ln xa) \cdot \ln a = a^x \ln a$.

Theorem 55. Suppose $c, k \in \mathbb{R}$. Then $u(x) = ke^{cx}$ will satisfy the following differential equation:

$$\begin{cases} u' = ku, & x \in \mathbb{R} \\ u|_{x=0} = c \end{cases}$$

Proof. $(\Longleftarrow)$: Follows by Chain Rule.
$(\Longrightarrow)$: Let $v(x)$ be a another solution. Define $h(x) = \frac{v(x)}{e^{kx}} - c$; $h(0) = \frac{v(0)}{1} - c = 0$. Then

$$h'(x) = \frac{v'e^{kx} - (e^{kx})'v}{(e^{kx})^2} = \frac{kve^{kx} - ke^{kx}v}{e^{2kx}} = 0, \quad \forall x \in \mathbb{R}.$$

□

This theorem also tells us that

$$x^\alpha = \exp(\alpha \ln x), \quad \text{for } x > 0, \ \alpha \in \mathbb{R}.$$

6.2 Sine and Cosine

Definition 61 (Sine/Cosine Function)**.** The solution to the differential equation

$$\begin{cases} u'' + u = 0, & x \in \mathbb{R} \\ u|_{x=0} = 0 \\ u'|_{x=0} = 1 \end{cases}$$

is defined as $u(x) = \sin x$, the *sine function*. Similarly, the solution to the differential equation

$$\begin{cases} v'' + v = 0, & x \in \mathbb{R} \\ v|_{x=0} = 1 \\ v'|_{x=0} = 0 \end{cases}$$

is defined as $v(x) = \cos x$, the *cosine function*.

There is a theorem that tells us that these $u(x)$ and $v(x)$ are indeed unique solutions to the differential equations.

6.3 Factorial Function

We want to be able to extend the factorial function to the real numbers. Recall that we define the *factorial* of some $n \in \mathbb{N}$ as

$$n! = n \cdot (n-1) \cdot (n-2) \cdots 3 \cdot 2 \cdot 1.$$

To do this, we introduce a new function.

Definition 62 (Gamma Function)**.** For $0 < x < \infty$, the *Gamma function* of x is defined as

$$\Gamma(x) = \int_0^x t^{x-1} e^{-t}\, dt.$$

The integral converges for $x \in (0, \infty)$.

Theorem 56. The following are true for $\Gamma(x)$:

(a) The functional equation

$$\Gamma(x+1) = x\Gamma(x)$$

holds if $0 < x < \infty$.

(b) $\Gamma(n+1) = n!$ for $n = 1, 2, \ldots$.

Proof. (a) follows directly from integration by parts, and since $\Gamma(1) = 1$, by induction, we find (a) $\implies$ (b). □

Theorem 57 (Bohr-Mollerup Theorem)**.** If f is a positive function on $(0, \infty)$ such that

(1) $f(x+1) = f(x)$,

(2) $f(1) = 1$,

(3) $\log f$ is convex,

then $f(x) = \Gamma(x)$.

Proof. We must show that (1), (2), and (3) uniquely determines the Gamma function. By (1), it is enough to do this for $x \in (0, 1)$. Let $\phi = \log f$. Then

$$\phi(x+1) = \phi(x) + \log x, \quad (0 < x < \infty).$$

We also have $\phi(1) = 0$ and ϕ is convex. Suppose $x \in (0, 1)$ and $n \in \mathbb{N}$. Then $\phi(n+1) = \log(n!)$ by above. We consider the difference quotients of ϕ on the intervals $[n, n+1], [n+1, n+1+x], [n+1, n+2]$. Since ϕ is convex,

$$\log n \leq \frac{\phi(n+1+x) - \phi(n+1)}{x} \leq \log(n+1).$$

We repeatedly apply the ϕ identity to get

$$\phi(n+1+x) = \phi(x) + \log[x(x+1)\cdots(x+n)].$$

Hence

$$0 \le \phi(x) - \log\left[\frac{n!n^x}{x(x+1)\cdots(x+n)}\right] \le x\log\left(1+\frac{1}{n}\right).$$

Since the latter expression goes to 0 for $n \to \infty$, $\phi(x)$ is uniquely determined. □

As a direct consequence of this proof, we arrive at *Euler's limit* definition of the Gamma function:

$$\Gamma(x) = \lim_{n\to\infty} \frac{n!n^x}{x(x+1)\cdots(x+n)}.$$

Definition 63 (Beta Function)**.** If $x > 0$ and $y > 0$, then the *Beta function* of x and y is defined as

$$B(x,y) = \int_0^1 t^{x-1}(1-t)^{y-1}\,dt = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}.$$

If we substitute $t = \sin^2\theta$, then the have another form of the Beta function:

$$2\int_0^{\pi/2} (\sin\theta)^{2x-1}(\cos\theta)^{2y-1}\,d\theta = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}.$$

This is one way that we can see that $\Gamma(\frac{1}{2}) = \sqrt{\pi}$; we simply plug in $x = y = \frac{1}{2}$ into the Beta function of this form. Likewise, we can do the substitution $t = u^2$ and get another form of the Gamma function:

$$\Gamma(x) = 2\int_0^\infty u^{2x-1}e^{-x^2}\,du, \quad (0 < x < \infty).$$

This gives rise to the well-known so-called *Gaussian integral* in the case that $x = \frac{1}{2}$:

$$\int_{-\infty}^{\infty} e^{-u^2}\,du = \sqrt{\pi}.$$

7 Sequences of Functions

7.1 Pointwise Convergence

A sequence of functions f_n is exactly what it sounds like. More formally, it means that for all $n = 1, 2, \ldots$, we have $f_n : E \to \mathbb{R}$. For example, consider

$$f_n(x) = x^n, \quad x \in \mathbb{R}, n = 1, 2, \ldots.$$

This has graph:

Definition 64 (Pointwise Convergence)**.** Suppose we have f_n, a sequence of functions on E. We say f_n *converges pointwisely* to f on E as $n \to \infty$ if and only if for every $x_0 \in E$, the sequence of numbers $f_n(x_0)$ converges to $f(x_0)$. In other words, for all $x_0 \in E$, we have $f_n(x_0) \to f(x_0)$.

We usually denote this as either

$$f_n \to f \text{ p.w. on } E, \text{ or p.w.} \lim_{n\to\infty} f_n(x) = f(x).$$

Consider $f_n(x) = x^n$ for $E = (-1, 1)$. Intuitvely, we should have $x^n \to 0$ p.w. on E as $n \to \infty$. Indeed, fix any $x_0 \in (-1, 1)$. Then $|x_0| < 1$, and so $|x_0^n| = |x_0|^n \to 0$ by properties of sequences. If we changed $E = [0, 1]$, then $x^n \not\to 0$ p.w. on E anymore, because for $x_0 = 1$, $x_0^n = 1^n \to 1$, as $n \to \infty$. In this case, $x^n \to f(x)$ p.w. on E, where $f(x)$ is defined by

$$f(x) = \begin{cases} 0, & 0 \le x < 1 \\ 1, & x = 1 \end{cases}$$

When $x_0 = -1$, then $x_0^n = (-1)^n$ is divergent, so x^n is pointwise divergent on $[-1, 1]$. Hence there does not exist an f that f_n converges pointwisely to.

Definition 65. Let $f_n : E \to \mathbb{R}$ for $n = 1, 2, \ldots$, we say

$$\sum_{n=1}^{\infty} f_n(x) = S(x) \text{ p.w. on } E.$$

In other words, for all $x_0 \in E$,

$$\sum_{n=1}^{\infty} f_n(x_0) = S(x_0).$$

For example,

$$\sum_{n=0}^{\infty} x^n = \frac{1}{1-x} \text{ p.w. on } (-1,1).$$

However for all x_0 such that $|x_0| > 1$, $\sum_{n=0}^{\infty} x_0^n$ diverges because $x_0^n \not\to 0$.

Definition 66 (C−norm)**.** Let $f : E \to \mathbb{R}$. Then the C*−norm* of f is

$$\|f\|_{C(E)} = \sup_{x\in E} |f(x)|.$$

This is also known as the *sup-norm* or L^∞*−norm*.

Properties of this C−norm:

(1) $\|f\|_{C(E)} = 0 \iff f(x) = 0$, for all $x \in E$

(2) $\|kf\|_{C(E)} = |k| \cdot \|f\|_{C(E)}$

(3) $\|f+g\|_{C(E)} \le \|f\|_{C(E)} + \|g\|_{C(E)}$

(4) For all $x \in E$, $|f(x)| \le \|f\|_{C(E)}$

Proof. Proof of (2):

$$\begin{aligned}\|kf\|_{C(E)} = |k| \cdot \|f\|_{C(E)} &= \sup_{x\in E} |kf(x)| \\ &= \sup_{x\in E}\{|k| \cdot |f(x)| \mid x \in E\} \\ &= |k| \sup\{|f(x)| \mid x \in E\} \\ &= |k| \cdot \|f\|_{C(E)}\end{aligned}$$

Proof of (3): Since $\|f+g\|_{C(E)} = \sup\{|f(x)+g(x)| \mid x \in E\}$. Notice that

$$|f(x)+g(x)| \le |f(x)| + |g(x)| \le \|f\|_{C(E)} + \|g\|_{C(E)}.$$

□

These properties are analogous to properties of $|x|$, for $x \in \mathbb{R}$.

7.2 Uniform Convergence

Pointwise convergence does not really tell us much about a function. Another form of convergence, uniform convergence, is much more interesting to study.

Definition 67 (Uniform Convergence)**.** Consider a sequence of functions f_n, where $f_n : E \to \mathbb{R}$ and $n = 1, 2, \ldots$. We say f_n *converges uniformly* to f iff

$$\|f_n - f\|_{C(E)} \to 0, \quad n \to \infty.$$

Equivalently,

$$\forall \epsilon > 0, \exists N \text{ such that if } n \geq N, \text{ then } \|f_n - f\|_{C(E)} < \epsilon.$$

If $f_n \to f$ converges uniformly on E, then

$$\forall x_0 \in E, |f_n(x_0) - f(x_0)| < \|f_n - f\|_{C(E)} \to 0, \quad n \to \infty.$$

This shows us that uniform convergence implies pointwise convergence. The converse is not true, however. When we compare the ϵ–definitions of pointwise and uniform convergence, we can see that they are quite similar.

- Pointwise:

  $$\forall x \in E, \forall \epsilon > 0, \exists N \text{ such that if } n \geq N, \text{ then } |f_n(x) - f(x)| < \epsilon.$$

  Here, N is reliant on x and ϵ.

- Uniform:

  $$\forall \epsilon > 0, \exists N \text{ such that if } n \geq N, \text{ then } |f_n(x) - f(x)| < \epsilon, \forall x \in E.$$

  In comparison, N is only reliant on ϵ.

We can also formulate a Cauchy definition for uniform convergence.

Theorem 58. Let $f_n : E \to \mathbb{R}$, for $n = 1, 2, \ldots$. Then f_n converges uniformly to some $f : E \to \mathbb{R}$ iff

$$\forall \epsilon > 0, \exists N \text{ such that if } n, m \geq N, \text{ then } \|f_n - f_m\|_{C(E)} < \epsilon.$$

Proof. $(\Longrightarrow)$: We have some $f : E \to \mathbb{R}$ such that $\|f_n - f\|_{C(E)} \to 0$. Then for all $\epsilon > 0$, there exists N such that if $n \geq N$, then $\|f\_n - f\|_{C(E)} < \epsilon$. This implies that for all $\epsilon > 0$, there exists N such that for $n, m \geq N$, we have

$$\|f_n - f_m\|_{C(E)} = \|f_n - f + f - f_m\|_{C(E)} < \|f_n - f\|_{C(E)} + \|f - f_m\|_{C(E)} < \epsilon + \epsilon = 2\epsilon.$$

($\Longleftarrow$) : We first try to find $f : E \to \mathbb{R}$. Our assumption tells us that f_n is a Cauchy sequence, and so there exists

$$\lim_{n\to\infty} f_n(x_0) = L_{x_0}.$$

Nowe we have $f : E \to \mathbb{R}$ and $f_n \to f$ pointwisely on E ($x \to L_{x_0}$). Now we want to show that $f_n \to f$ uniformly on E. Again, for all $\epsilon > 0$, we take $m \to \infty$, and so there exists N such that if $n, m \geq N$, then for all x, $|f_n(x) - f(x)| < \epsilon$. Hence $\|f - n - f\|_{C(E)} < \epsilon$, and so we conclude that

$$\lim_{n\to\infty} \|f_n - f\|_{C(E)} = 0.$$

□

7.3 Properties of Uniform Convergence

Suppose that all $f_n : E \to \mathbb{R}$ are continuous on E. Let $f - n \to f$ in some sense on E as $n \to \infty$. Is f necessarily continuous on E? For pointwise convergence, this is not true, but for uniform convergence, this is true.

Theorem 59. Let $f_n : E \to \mathbb{R}$ be continuous on E for $n = 1, 2, \ldots$. Let $f_n \to f$ uniformly on E as $n \to \infty$. Then f is continuous on E.

As a quick note, f being continuous at $x_0 \in E$ means that

$$\lim_{x\to x_0} f(x) = f(x_0).$$

Thus

$$\lim_{x\to x_0}\left(\lim_{n\to\infty} f_n(x)\right) = \lim_{n\to\infty} f_n(x_0) = \lim_{n\to\infty}\left(\lim_{x\to x_0} f_n(x)\right).$$

Hence, for pointwise convergence,

$$\lim_{x\to x_0}\left(\lim_{n\to\infty} f_n(x)\right) \neq \lim_{n\to\infty}\left(\lim_{x\to x_0} f_n(x)\right),$$

but it is true for uniform convergence.

Proof. Our goal is to show that, fixing any $x_0 \in E$,

$$\forall \epsilon > 0, \exists \delta > 0 \text{ such that if } |\Delta x| \leq \delta, \text{ then } |f(x_0 + \Delta x) - f(x_0)| < \epsilon.$$

For this ϵ, uniform convergence implies that there exists N such that $\|f_n - f\|_{C(E)} < \epsilon$. This f_n is continuous at x_0, so

$$\exists \delta \text{ such that if } |\Delta x| \leq \delta, \text{ then } |f_N(x_0) - f_N(x_0 + \Delta x)| < \epsilon.$$

Take this δ. Then for $|\Delta x| \leq \delta$, we have

$$\begin{aligned}
|f(x_0+\Delta x)+f(x_0)| &= |f(x_0+\Delta x) - f_N(x_0+\Delta x + f_N(x_0+\Delta x) + f_N(x_0) - f_N(x_0) + f(x_0)| \\
&\leq |f_N(x_0+\Delta x) - f_N(x_0)| + |f(x_0+\Delta x) - f_N(x_0+\Delta x)| + |f_N(x_0) - f(x_0)| \\
&\leq \epsilon + \epsilon + \epsilon \\
&= 3\epsilon.
\end{aligned}$$

□

This shows us that $f_n \xrightarrow{\text{unif on } E}$, then for all $p \in E$,

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

How does uniform convergence interact with integration or integrable functions?

Theorem 60 (Integrable Limit Theorem)**.** Let $f_n : [a,b] \to \mathbb{R}$ be Riemann integrable for all n. Let $f_n \to f$ uniformly on E. Then f is integrable on $[a,b]$, and

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

Proof. For all $\epsilon > 0$, then there exists N_ϵ such that for $n \geq N_\epsilon$, and for all $x \in [a,b]$, $|f_n(x) - f(x)| < \epsilon$. Let P be a partition, and $I_k = [x_{k-1}, x_k]$, for $1 \leq k \leq n$. Then

$$U(f,P) - L(f,P) = \sum_{k=1}^{n} |I_k| \left(\sup_{I_k} f - \inf_{I_k} f \right).$$

Let $J = I_k$. Then

$$\left[\sup_{x\in J} f(x) - \inf_{x\in J} f(x) \right] - \left[\sup_{x\in J} f_n(x) - \inf_{x\in J} f_n(x) \right].$$

Because $n \geq N_\epsilon$ gives $|f(x) - f_n(x)| < \epsilon$, we have that

$$f(x) < f_n(x) + \epsilon, \quad f_n(x) - \epsilon < f(x) \implies \left| \sup_j f - \sup_j f_n \right| < \epsilon.$$

So

$$\begin{aligned}
|U(f,P) - L(f,P) - [U(f_n,P) - L(f_n,P)]| &= \left| \sum_{k=1}^{n} |I_k| \left[\left(\sup_{I_k} f - \sup I_k f_n \right) - \left(\inf_{I_k} f - \inf_{I_k} f_n \right) \right] \right| \\
&\leq 2\epsilon \sum_{k=1}^{n} |I_k| \\
&= 2\epsilon(b-a).
\end{aligned}$$

So $U(f,P) - L(f,P) \leq U(f_n,P) - L(f_n,P) + 2\epsilon(b-a)$. We choose P such that $U(f_n,P) - L(f_n,P) < \epsilon$. Hence f is Riemann integrable. Now it follows that

$$\left|\int_a^b f_n - \int_a^b f\right| = \left|\int_a^b (f_n - f)\right| \leq \int_a^b |f_n - f|.$$

Let $\epsilon > 0$ be aribtrary. Because $f_n \to f$ uniformly, there exists an N such that

$$|f_n(x) - f(x)| < \epsilon/(b-a),$$

for all $n \geq N$ and $x \in [a,b]$. Thus for $n \geq N$ we have that

$$\begin{aligned}\left|\int_a^b f_n - \int_a^b f\right| &\leq \int_a^b |f_n - -f| \\ &\leq \int_a^b \frac{\epsilon}{b-a} \\ &= \epsilon.\end{aligned}$$

□

Uniform convergence is too strong for many important applications. For example, consider $f_n(x) = x^n$ on $[0,1]$. Then

$$\lim_{n\to\infty} f_n(x) = \begin{cases} 1, & x = 1 \\ 0, & x \neq 1 \end{cases} = f(x)$$

is Riemann integrable. Here we have

$$\int_0^1 f_n(x) = \frac{1}{n+1} \implies \lim_{n\to\infty} \frac{1}{n+1} = 0,$$

and

$$\int_0^1 f\,dx = \lim_{n\to\infty} \int_0^1 f_n(x)\,dx = 0.$$

But f_n does not converge uniformly to f, because $\|x^n - f(x)\|_{C([0,1])} = 1 \neq 0$. Here we can switch the limit and integral.

Example 10. Let $f(x) = \tan^{-1}(x)$. We know that $f'(x) = \frac{1}{1+x^2}$, so

$$\int_0^1 \frac{1}{1+x^2} = \frac{\pi}{4}.$$

We know

$$\frac{1}{1+x^2} = 1 - x^2 + x^4 - x^6 + \cdots = \lim_{N\to\infty} \sum_{n=0}^{N} (-1)^n x^{2n}.$$

So is it true that

$$\int_0^1 \left(\lim_{N\to\infty} \sum_{n=0}^{N} (-1)^n x^{2n} \right) dx = \lim_{N\to\infty} \left(\int_0^1 \sum_{n=1}^{N} (-1)^n x^{2n}\, dx \right) = \lim_{N\to\infty} \left(\sum_{n=0}^{\infty} \frac{(-1)^n}{2n+1} \right).$$

If so, then we could show that $\frac{\pi}{4} = 1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + \cdots$. However, we can't use the Integrable Limit Theorem; $S_n(x) = \sum_{j=0}^{n} (-1)^n x^{2n}$ does not converge uniformly on $[0,1]$. This gives us another example of why uniform convergence is too strong.

Definition 68 (Convergence of Series)**.** We say that $\sum_{n=1}^{\infty} f_n(x)$ converges uniformly/pointwisely to $S(x)$ iff for the sequence of partial sums,

$$S_n(x) = \sum_{k=1}^{n} f_k(x),$$

we have that $S_n \to S$ uniformly/pointwisely.

We focus on uniform convergence of series. We have many analagous results to sequences of functions as we do with series:

- C–norm:
$$\left\| \sum_{k=1}^{n} f_n(x) - S(x) \right\|_{C(E)} \to 0, \quad n \to \infty.$$
- Cauchy Criterion:
$$\forall \epsilon > 0, \exists N \text{ such that if } n, m \geq N, \text{ then } \left\| \sum_{k=n}^{m} f_k(x) \right\| < \epsilon.$$
- Divergence Criterion:
$$\sum_{n=1}^{\infty} a_n < \infty \implies \lim_{n\to\infty} |a_n| = 0.$$
- Assume $\sum_{k=1}^{\infty} f_n(x)$ converges uniformly on some E to $S(x)$. Then
$$\lim_{n\to\infty} \|f_n\|_{C(E)} = 0,$$
because $\|S_n - S\|_{C(E)} \to 0$ and hence $\|S_{n+1} - S\|_{C(E)} \to 0$.

The last point is because

$$\|f_n\|_{C(E)} = \|S_{n+1} - S_n\|_{C(E)} = \|S_{n+1} - S + S - S_n\|_{C(E)} \leq \|S_{n+1} - S\|_{C(E)} + \|S - S_n\|_{C(E)} \to 0.$$

Theorem 61 (Differential Limit Theorem)**.** Let $f_n : (a,b) \to \mathbb{R}$ be differentiable for all n. Let f_n' be continuous on (a,b). Suppose that $f_n' \to g$ uniformly on (a,b) for some $g : (a,b) \to \mathbb{R}$, and let there exist $p \in (a,b)$ such that $f_n(p)$ converges. Then there exists some $f : (a,b) \to \mathbb{R}$ such that $f_n \to f$ uniformly on (a,b), and $f'(x) = g(x)$.

To put it in a different way,

$$f'(x) = g(x) \Longleftrightarrow \frac{d}{dx}\left(\lim_{n\to\infty} f_n(x)\right) = \lim_{n\to\infty}\left(\frac{d}{dx} f_n(x)\right).$$

As we can see, this is just an analagous result for derivatives to the Integrable Limit Theorem.

Proof. Step 1: We want to show that f_n' continuous implies that $f^{n\prime}$ is integrable. By the FTC,

$$f_n(x) = f_n(p) + \int_p^x f_n'(t)\,dt.$$

Then $f_n' \to g$ uniformly on (a,b) and f_n' continuous tells us that g is continuous on (a,b). Then

$$\{\text{integral uniformly conv.}\} \implies \int_p^x f_n'(t)\,dt \to \int_p^x g(t)\,dt \text{ on } [p,x].$$

Now we define $f : (a,b) \to \mathbb{R}$ by

$$f(x) = \alpha + \int_p^x g(t)\,dt.$$

Step 2: g is continuous, and by the FTC,

$$f'(x) = 0 + \left(\int_p^x g(t)\,dt\right)' = g(x).$$

Step 3: We want to show that $f_n \to f$ uniformly on (a,b). To do this, we want to show that $\|f_n - f\|_{C(a,b)} = \sup_{x\in(a,b)} |f_n(x) - f(x)| \to 0$, as $n \to \infty$. We have

$$\begin{aligned}
|f_n(x) - f(x)| &= \left| f_n(p) - \alpha + \int_p^x f_n'(t)\,dt - \int_p^x g(t)\,dt \right| \\
&\le |f_n(p) - \alpha| + \left| \int_p^x (f_n'(t) - g(t))\,dt \right| \\
&\le |f_n(p) - \alpha| + \int_p^x |f_n'(t) - g(t)|\,dt \\
&\le |f_n(p) - \alpha| + \int_p^x \sup_{(a,b)} |f_n'(t) - g(t)|\,dt \\
&\le |f_n(P) - \alpha| + \|f_n' - g\|_{C((a,b))} \cdot |b - a|
\end{aligned}$$

Hence

$$\sup_{x\in(a,b)} |f_n(x) - f(x)| \le |f_n(p) - \alpha| + \|f_n' - g\|_{C((a,b))} \to 0, \quad n \to \infty.$$

□

All theorems about uniformly convergent sequences can be reformulated for series.

Theorem 62. Let f_n be integrable on $[a,b]$ for all n. Suppose $\sum_{n=1}^{\infty} f_n(x) \to S(x)$ uniformly on $[a,b]$. Then $f(x)$ is integrable on $[a,b]$, and

$$\int_a^b \sum_{k=1}^{\infty} f_k(x)\,dx = \sum_{k=1}^{\infty} \int_a^b f_k(x)\,dx.$$

Proof. We know that $S_n(x) \to f(x)$ uniformly on $[a,b]$. By the sequential version of the theorem,

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

Thus

$$\lim_{n\to\infty}\left(\int_a^b \sum_{k=1}^{n} f_k(x)\,dx\right) = \lim_{n\to\infty}\left(\sum_{k=1}^{n} \int_a^b f_k(x)\,dx\right) = \sum_{k=1}^{\infty}\left(\int_a^b f_k(x)\,dx\right).$$

□

So how do we prove uniform convergence for series?

- Cauchy Criterion:

$$\forall \epsilon > 0, \exists N \text{ such that if } n, m \ge N, \text{ then } \left\|\sum_{k=n}^{m} f_k\right\| < \epsilon.$$

- Weierstrass M-Test

Theorem 63 (Weierstrass M-Test)**.** Let $f_n : E \to \mathbb{R}$ for all n. Let $M_n = \sup_E |f_n|$, and $\sum_{k=1}^{\infty} M_n < \infty$. Then $\sum_{n=1}^{\infty} f_n(x)$ converges uniformly on E to some $S(x)$, for $S : E \to \mathbb{R}$.

In other words, this tells us that we can prove uniform convergence of a series just by showing that

$$\sum_{n=1}^{\infty} \|f_n\|_{C(E)} < \infty.$$

Proof. Fix $\epsilon > 0$. We have that there exists N such that for $n, m \geq N$, $\sum_{k=n}^{m} M_k < \epsilon$. But then for $n, m \geq N$,

$$\begin{aligned}\|f_n + \cdots + f_m\|_{C(E)} &= \sup_{x\in E} |f_n(x) + \cdots + f_m(x)| \\ &\leq \sup_{x\in E} |f_n(x)| + \cdots + \sup_{x\in E} |f_m(x)| \\ &\leq M_n + \cdots + M_m \\ &\leq \epsilon.\end{aligned}$$

Hence the Cauchy Criterion holds for $\sum_{n=1}^{\infty} f_n(x)$. □

For example, consider the series

$$\sum_{n=1}^{\infty} \frac{\sin(nx)}{n^\alpha},$$

for $\alpha > 1$. Does this series converge uniformly on $\mathbb{R}$? Yes, because

$$\sup \left| \frac{\sin(nx)}{n^\alpha} \right| \leq \frac{1}{n^\alpha}.$$

By the Weierstrass M-test, $\sum_{n=1}^{\infty} \frac{1}{n^\alpha} < \infty$ for $\alpha > 1$, so indeed it is uniformly convergent.

Definition 69 (L^1–norm)**.** Let $f : [a, b] \to \mathbb{R}$ be integrable. We define the L^1*–norm* of f as

$$\|f\|_1 = \int_a^b |f(x)| \, dx.$$

Much like the L^∞/C–norm, we have the expected few properties:

(1) Triangle Inequality:

$$\begin{aligned}\|f + g\|_1 &= \int_a^b |f(x) + g(x)| \, dx \\ &\leq \int_a^b |f(x)| + |g(x)| \, dx \\ &= \|f\|_1 + \|g\|_1\end{aligned}$$

(2) Absolutely Homogeneous:

$$\begin{aligned}\|cf\|_1 &= \int_a^b |cf(x)| \, dx \\ &= |c| \cdot \|f\|_1\end{aligned}$$

(3) Positive Definite:

$$\text{If } \|f\|_1 \geq 0, \text{ then } f = 0 \implies \|f\|_1 = 0, \text{ but } \|f\|_1 = 0 \implies f(x) = 0.$$

If we consider the space $(C([a,b]), \|\cdot\|_1)$, all of our properties hold. This shows us that the space of continuous functions are a subspace of the space of integrable functions.

Definition 70 (L^1−convergence)**.** Let $f_n : [a,b] \to \mathbb{R}$ be integrable for all n. Let $f : [a,b] \to \mathbb{R}$ be integrable. We say $f_n \xrightarrow{L^1} f$ as $n \to \infty$ iff $\|f_n - f\|_1 \to 0$.

As it turns out, uniform convergence implies both L^1 and pointwise convergence, but pointwise and L^1 convergence don't imply anything else.

7.4 Weierstrass Approximation Theorem

Theorem 64 (Weierstrass Approximation Theorem)**.** Let $f : [a,b] \to \mathbb{R}$ be continuous. Given $\epsilon > 0$, there exists a polynomial $p(x)$ satisfying, for all $x \in [a,b]$,

$$|f(x) - p(x)| < \epsilon.$$

Essentially, this tells us that every continuous function in a closed interval can be uniformly approximated by a polynomial. This may seem obvious for many smooth functions, but what about a continuous, nowhere differentiable function? It doesn't seem possible to come up with a polynomial that approximates this. To further explore this, we need to use a different idea.

Definition 71 (Polygonal Function)**.** A continuous function $\phi : [a,b] \to \mathbb{R}$ is *polygonal* if there is a partition

$$a = x_0 < x_1 < \cdots < x_n = b$$

of $[a,b]$ such that ϕ is linear on each subinterval $[x_{i-1}, x_i]$, for $i \in [1,n]$.

Polygonal functions help us define the idea of "linear interpolation". Given a set of points, we want to find a function whose graph passes through those points. If we consider the points

$$(0,1), \left(\frac{1}{4}, \frac{\sqrt{3}}{2}\right), \left(\frac{3}{4}, \frac{1}{2}\right), (1,0),$$

then one polygonal function that works is a piecewise function that simply connects the dots. This is not the only one, however. The function $\sqrt{1-x}$ also works.

Theorem 65. Let $f : [a,b] \to \mathbb{R}$ be continuous. Given $\epsilon > 0$, there exists a polygonal function ϕ satisfying, for all $x \in [a,b]$,

$$|f(x) - \phi(x)| < \epsilon.$$

Proof. Since f is continuous on a compact interval, f is uniformly continuous. Then given $\epsilon > 0$, there exists $\delta > 0$ such that if $|x-y| < \delta$, we have $|f(x)-f(y)| < \frac{\epsilon}{2}$, for all $x \in [a,b]$. Now choose $a = x_0 < x_1 < \cdots < x_n = b$ such that $x_{k+1} - x_k < \delta$ for all $k \in [0, n-1]$. Let ϕ be a polygonal function such that $\phi(x_k) = f(x_k)$. For any interval $[x_k, x_{k+1}]$,

$$\begin{aligned} |\phi(t) - f(t)| &= |\phi(t) - \phi(x_k) + \phi(x_k) - f(t)| \\ &\leq |\phi(t) - \phi(x_k)| + |f(x_k) - f(t)| \\ &< \epsilon. \end{aligned}$$

□

We can generalize the Weierstrass Theorem to a be applicable in a broader variety of situations.

Definition 72 (Algebra)**.** A family $\mathcal{A}$ of functions defined on a set E is said to be an *algebra* if

(i) $f + g \in \mathcal{A}$,

(ii) $fg \in \mathcal{A}$,

(iii) $cf \in \mathcal{A}$ for all $f, g \in \mathcal{A}$ and $c \in \mathbb{R}$.

In abstract algebra lingo, this just means that $\mathcal{A}$ is closed under addition, multiplication, and scalar multiplication.

Definition 73 (Uniformly Closed)**.** If $\mathcal{A}$ has the property that $f \in \mathcal{A}$ whenever $f_n \in \mathcal{A}$ for $n = 1, 2, \ldots$, and $f_n \to f$ uniformly on E, then $\mathcal{A}$ is said to be *uniformly closed*.

If $\mathcal{B}$ is the set of all functions which are limits of uniformly convergent sequences of members of $\mathcal{A}$, then $\mathcal{B}$ is known as the *uniform closure* of $\mathcal{A}$. As an example, the set of all polynomials forms an algebra, and we can restate the Weierstrass Approximation Theorem as saying that $C([a,b])$ is the uniform closure of the set of polynomials on $[a,b]$.

Theorem 66. Let $\mathcal{B}$ be the uniform closure of an algebra $\mathcal{A}$ of bounded functions. Then $\mathcal{B}$ is a uniformly closed algebra.

Proof. If $f, g \in \mathcal{B}$, then there exists uniformly convergent sequences f_n and g_n such that $f_n \to f$ and $g_n \to g$, and $f_n, g_n \in \mathcal{A}$. Since these are all bounded functions, it follows that each of the following uniformly:

$$f_n + g_n \to f + g, \quad f_n g \to f g, \quad c f_n \to c f.$$

Hence $f + g \in \mathcal{B}$, $fg \in \mathcal{B}$, and $cf \in \mathcal{B}$, so $\mathcal{B}$ is an algebra. □

Definition 74 (Separate Points)**.** Let $\mathcal{A}$ be a family of functions on a set E. Then $\mathcal{A}$ is said to *separate points* on E if to every pair of distinct points $x_1, x_2 \in E$, there corresponds a function $f \in \mathcal{A}$ such that $f(x_1) \neq f(x_2)$.

If to each $x \in E$ there corresponds a function $g \in \mathcal{A}$ such that $g(x) \neq 0$, we say that $\mathcal{A}$ *vanishes at no point of* E.

The algebra of all polynomials in one variable always has these properties on $\mathbb{R}$.

Theorem 67. Suppose $\mathcal{A}$ is an algebra of functions on a set E, $\mathcal{A}$ separates points on E, and $\mathcal{A}$ vanishes at no point of E. Suppose x_1, x_2 are distinct points of E, and c_1, c_2 are constants, then $\mathcal{A}$ contains a function f such that

$$f(x_1) = c_1, \quad f(x_2) = c_2.$$

Proof. The assumptions show that $\mathcal{A}$ contains functions g, h, k such that

$$g(x_1) \neq g(x_2), \quad h(x_1) \neq 0, \quad k(x_2) \neq 0.$$

Let

$$u = gk - g(x_1)k, \quad v = gh - g(x_2)h.$$

Then evidently, $u, v \in \mathcal{A}$, and $u(x_1) = v(x_2) = 0$, $u(x_2) \neq 0$, and $v(x_1) \neq 0$. Thus

$$f = \frac{c_1 v}{v(x_1)} + \frac{c_2 u}{u(x_2)}$$

will have our desired properties. □

Theorem 68 (Stone-Weierstrass Theorem)**.** Let $\mathcal{A}$ be an algebra of real continuous functions on a compact set K. If $\mathcal{A}$ separates points on K and if $\mathcal{A}$ vanishes at no point of K, then the uniform closure $\mathcal{B}$ of $\mathcal{A}$ consists of all real continuous functions on K.

8 Analytic Functions

Theorem 69. Assume $\sum_{n=0}^{\infty} c_n x^n$ converges for $x = p$. Then for all $\delta > 0$, $\sum_{n=0}^{\infty} c_n x^n$ converges uniformly on $|x| \leq |p| - \delta$. In other words, for all $x \in [-|p| + \delta, |p| - \delta]$.

Proof. Following the same procedure as the theorem above, we bound $|c_n||p|^n$ from above by some $N < \infty$. Then

$$\begin{aligned} \sup |c_n||p|^n \left(\frac{|x|}{|p|}\right)^n &= N \sup \left(\frac{|x|}{|p|}\right)^n \\ &\leq N \left(\frac{|p| - \delta}{|p|}\right)^n \\ &= N \left(1 - \frac{\delta}{|p|}\right)^n \end{aligned}$$

Then because $|1 - \frac{\delta}{|p|}| < 1$, the series $\sum_{n=0}^{\infty} N(1 - \frac{\delta}{|p|})^n < \infty$, and so by the Weierstrass M-test, it is uniformly convergent. $\square$

Theorem 70 (Abel's Theorem)**.** For any $\sum_{n=0}^{\infty} c_n x^n$, there exists some $R \in [0, \infty)$ such that

(a) $\sum_{n=0}^{\infty} c_n x^n$ converges absolutely for all $x : |x| < R$,

(b) $\sum_{n=0}^{\infty} c_n x^n$ diverges for all $x : |x| > R$,

(c) No information in general about $|x| = R$,

(d) For all $0 < \rho < R$, $\sum_{n=0}^{\infty} c_n x^n$ converges uniformly on $[-\rho, \rho]$.

Proof. R is defined to be $\{|x| : \sum_{n=0}^{\infty} c_n x^n \text{ converges}\}$, and this will be nonempty because 0 is a possibility for R.

(a) Follows from Theorem above.

(b) If it converges for p, and $|x| > R$, then this contradicts the definition of R.

(c) Consider $\sum_{n=0}^{\infty} \frac{x^n}{n}$.

(d) Follows from Theorem above.

□

8.1 Analyticity and Properties

We can indeed calculate this R using a formula derived from complex analysis.

Theorem 71 (Cauchy-Hadamard Formula)**.** Given a series $\sum_{n=0}^{\infty} c_n x^n$, we can calculate R using the formula

$$R^{-1} = \limsup_{n\to\infty}\left(\sqrt[n]{|c_n|}\right).$$

Definition 75 (Radius of Convergence)**.** Given a series $\sum_{n=0}^{\infty} c_n x^n$, we call such an R the *radius of convergence* of our series. Similarly, we call the interval $(-R, R)$ the *interval of convergence*.

In general, we can extend Abel's Theorem to complex numbers as well. For $z = x + iy$, the series $\sum_{n=0}^{\infty} c_n z^n$ converges absolutely for all $z : |z| < |p|$, and uniformly on $E = \{z : |z| \leq |p| - \delta\}$, for any $\delta > 0$. The reason is because $\sum_{n=0}^{\infty} z^n$ converges if $|z| < 1$, even if $z \in \mathbb{C}$.

Definition 76 (Real-Analytic)**.** If $f(x) = \sum_{n=0}^{\infty} c_n x^n$, with some radius of convergence $R > 0$, then f is called *real-analytic* (or simply just *analytic*) in $(-R, R)$.

The set of all analytic functions on an interval of convergence is denoted

$$C^{\omega}[(-R,R)] = \{f : (-R,R) \to \mathbb{R} \mid f \text{ is analytic in } (-R,R)\}.$$

Obviously, we can integrate analytic functions term-by-term, so

$$\int_a^b f(x)\,dx$$

exists, and we can calculate it to be

$$\int_a^b \left(\sum_{n=0}^{\infty} c_n x^n\right) dx = \sum_{n=0}^{\infty} \int_a^b c_n x^n\,dx = \sum_{n=0}^{\infty} \frac{b^{n+1} - a^{n+1}}{n+1} c_n.$$

Can we also differentiate term-by-term for analytic functions?

Theorem 72. Let $f(x) = \sum_{n=0}^{\infty} c_n x^n$, convergent in $(-R, R)$ for $R > 0$. Then

(1) There exists $f'(x)$ for all $x \in (-R, R)$, and

$$f'(x) = \frac{d}{dx}\left(\sum_{n=0}^{\infty} c_n x^n\right) = \sum_{n=0}^{\infty} n c_n x^{n-1},$$

(2) f' is analytic in $(-R, R)$,

(3) There exists $f^{(k)}(x)$ for all $x \in (-R, R)$ and $k = 1, 2, \ldots$, and

$$f^{(k)}(x) = \sum_{n=k}^{\infty} (c_n x^n)^{(k)}.$$

Proof. We have that (1) $\implies$ (2) and (3) at once. So we only have to prove (1). Fix some $x \in (-R, R)$. We need to show that $\sum_{n=0}^{\infty} f_n'$ converges uniformly on (a, b). We use the $M - test$. For $\rho < a < R$, $\sum_{n=0}^{\infty} |c_n x^n| < \infty$, and $\frac{|x|}{|a|} < 1$. For all $x \in [-\rho, \rho]$,

$$\begin{aligned}
\sum_{n=0}^{\infty} |n c_n x^{n-1}| &\leq \sum_{n=0}^{\infty} \sup_{[-\rho,\rho]} |n c_n x^{n-1}| \\
&= \sum_{n=1}^{\infty} n |c_n| \rho^{n-1} \\
&= \sum_{n=1}^{\infty} |c_n \rho^n| \cdot \frac{n}{a} \cdot \frac{|\rho|^{n-1}}{|a|^{n-1}} \\
&\leq \sum_{n=1}^{\infty} |c_n p^n| \cdot k \\
&< \infty.
\end{aligned}$$

This is because

$$\frac{1}{a} \cdot n \cdot \frac{|\rho|^{n-1}}{|a|^{n-1}} = \frac{1}{a} \cdot n(1-\delta)^{n-1} = \frac{1}{a} \frac{n}{(1+\epsilon)^{n+1}} \to 0, \quad n \to \infty.$$

□

Using this theorem, we can notice a few things:

- If $f \in C^{\omega}((-R, R))$, then all derivatives exist in $(-R, R)$, and so $f \in C^{\infty}((-R, R))$ as well.
- $C^{\omega}((-R, R))$ is a vector space.

- We can calculate any coefficient of $f(x)$ by taking

$$c_n = \frac{f^{(n)}(0)}{n!}.$$

This begs the question: does $C^\omega((-R,R)) \subseteq C^\infty((-R,R))$, or is it a strict subspace?

Example 11. Let

$$f(x) = \begin{cases} e^{-1/x}, & x > 0 \\ 0, & x \le 0 \end{cases}.$$

Notice that $f \notin C^\omega((-R,R))$. However, we claim that $f \in C^\infty((-R,R))$. Notice that

$$\lim_{x\to 0^+} \frac{f(x)-f(0)}{x} = \lim_{x\to 0^+} \frac{f(x)}{x} = \lim_{x\to 0^+} \frac{e^{-1/x}}{x}.$$

Let $t = 1/x$. Then

$$\lim_{t\to\infty} e^{-t} t = \lim_{t\to\infty} \frac{t}{e^t} = 0,$$

so $f'(x) = e^{-1/x} \cdot \frac{1}{x^2}$. For the second derivative, we will find that

$$\lim_{x\to 0^+} \frac{f'(x)-f'(0)}{x} = \lim_{x\to 0^+} e^{-1/x} p_x\left(\frac{1}{x}\right) = \lim_{t\to\infty} \frac{p_x(t)}{e^t} \to 0.$$

This holds for the n^{th} derivative, because the exponential will always grow faster than a power.

From this, we can see that $C^\omega((-R,R)) \subset C^\infty((-R,R))$, a strict subset.

8.2 Analyticity Criterion

Recall Taylor's Formula. We want to exmaine behavior of $R_n(f,x)$.

- First, we fix n and let $x \to 0$. Then

$$f(x) = \left(\sum_{k=0}^{\infty} \frac{f^{(k)}(0)}{k!} x^k\right) + \text{terms bounded by } x^{n+1} = |R_n(x)|.$$

Indeed,

$$\begin{aligned} |R_n(x)| &= \frac{1}{n!}\left|\int_0^x (x-t)^n f^{(n+1)}(t)\,dt\right| \\ &\le \frac{1}{n!}\int_0^x |x-t|^n |f^{(n+1)}(t)|\,dt \\ &\le \frac{\|f^{(n+1)}\|_{C([-R/2,R/2])}}{n!} \cdot |x|^n \cdot |x| \\ &= k|x|^{n+1}. \end{aligned}$$

- Now we fix x_0, or equivalently, let $x \in [-\epsilon, \epsilon] \subset (-R, R)$. If $R_n(f, x_0) \to 0$ as $n \to \infty$, then

$$f(x_0) = \sum_{k=0}^{n} \frac{f^{(k)}(0)}{k!} x^k + R_n(f, x_0) \longrightarrow \sum_{k=0}^{\infty} \frac{f^{(k)}(0)}{k!} x^k,$$

  which is convergent. Thus if $R_n(f, x_0) \to 0$ as $n \to \infty$, then for all $x_0 \in (-\epsilon, \epsilon)$, $f(x) = \sum_{k=0}^{\infty} \frac{f^{(k)}(0)}{k!} x^k$. Thus f is analytic.

Using this, we can find an easier way to determine if a function is analytic.

Theorem 73 (Analyticity Criterion)**.** Let $f \in C^\infty((-R, R))$, and suppose there exists $M < \infty$ such that for all n,

$$\|f^{(n)}\|_{C((-R,R)} = \sup_{|x|<R} |f^{(n)}(x)| \leq M^n.$$

Then

$$f(x) = \sum_{k=0}^{\infty} \frac{f'(k)}{k!} x^k, \quad \forall |x| < R.$$

Proof. By Taylor's Formula,

$$f(x) = \sum_{k=0}^{n} \frac{f^{(k)}(0)}{k!} x^k + \frac{1}{n!} \int_0^x (x-t)^n f^{(n+1)}(t)\, dt.$$

Let the second term be $|I_n|$. Then

$$\begin{aligned} |I_n| &\leq \frac{1}{n!} r^n \int_0^x |f^{(n+1)}(t)|\, dt \\ &\leq \frac{r^n}{n!} M^{n+1} \cdot r \\ &= \frac{(rM)^n}{n!} \cdot rM \\ &= \frac{k^n}{n!} \to 0, \quad n \to \infty. \end{aligned}$$

□

Now we wish to examine analyticity of common elementary functions.

- Consider $f(x) = e^x$. We know $(e^x)^{(n)} = e^x$, for all n. Then

$$e^x = \sum_{k=0}^{n} \frac{x^k}{k!} + \left| \frac{1}{n!} \int_0^x (x-t)^n e^t\, dt \right|,$$

We take $n \to \infty$, and apply the criterion for $[-R, R]$. Thus

$$\begin{aligned}\sup_{|x|\leq R} |R_n(x)| &\leq \sup_{|x|\leq R} \frac{1}{n!}\int_0^x |x-t|^n e^t \, dt \\ &\leq \frac{1}{n!} R^{n+1} e^R \\ &\leq (e^R R) R^n \\ &= k \cdot M^n\end{aligned}$$

Thus by the criterion,

$$e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!}, \quad \forall x : |x| \leq R < \infty.$$

- Consider now $g(x) = \ln(1+x)$. By induction, for $x \in (-1, 1)$,

  $$(\ln(1+x))^{(n)}\Big|_{x=0} = (-1)^{n+1}(n-1)!$$

  Then

  $$\ln(1+x) = \sum_{k=0}^{n} \frac{(-1)^{k+1}(k-1)!}{k!} \cdot x^k + \frac{1}{n!}\int_0^x (x-t)^n \frac{(-1)^n n!}{(1+t)^{n+1}} \, dt.$$

  So

  $$R_n(x) = \int_0^x \frac{(x-t)^n}{(1+t)^n} \frac{dt}{1+t}.$$

  Fix any $x \in (-1, 1), |x| = 1-\delta, \delta > 0$. Then $|R_n(x)| \to 0$ as $n \to \infty$, and so

  $$\ln(1+x) = \sum_{n=0}^{\infty} (-1)^n \frac{x^n}{n}, \quad \forall x : |x| < 1.$$

8.3 Asymptotic Analysis

Definition 77 (Small-o Notation)**.** Suppose f and g are always defined in some $(a-R, a+R)$ and are not zero for $x \neq 0$. Then we say $f(x) = o(g(x))$ as $x \to a$ if $\frac{f(x)}{g(x)} \to 0$, as $x \to a$.

This is basically saying that f vanishes much faster than g does. Some examples:

- $x^2 = o(x)$, $x \to 0 \iff \frac{x^2}{x} = x \to 0$.
- $x = o(x^2)$, $x \to \infty \iff \frac{x}{x^2} = \frac{1}{x} \to 0$.

- $\frac{1}{x} = o(\frac{1}{x^2}),\ x \to 0 \Longleftrightarrow \frac{1/x}{1/x^2} = x \to 0.$
- If $f(x) = o(1)$ as $x \to a$, then $\lim_{x \to a} f(x) = 0$.

We can also reformulate some definitions using o notation.

- Continuity at x_0:
$$f(x) = f(x_0) + o(1),\ x \to x_0.$$
- Differentiability at x_0:
$$f(x) = f(x_0) + f'(x_0)(x - x_0) + o(x - x_0),\ x \to x_0.$$

Definition 78 (Big-$\mathcal{O}$ Notation)**.** Suppose f and g are always defined in some $(a-\epsilon, a+\epsilon)$ for $\epsilon > 0$. Then we say $f(x) = \mathcal{O}(g(x))$ if there exists some $C < \infty$ such that $|f(x)| \leq C|g(x)|$, for all $x \in (a-\epsilon, a+\epsilon)$, and $x \neq a$.

Theorem 74. If $f(x) = o(g(x))$ as $x \to a$, then $f(x) = \mathcal{O}(g(x))$ as $x \to a$.

Proof. If $\frac{|f(x)|}{|g(x)|} \to 0$ as $x \to a$, then $\frac{|f(x)|}{|g(x)|} \leq 1$ in some $(a-\delta, a+\delta)$, $\delta > 0$ and $x \neq a$. It follows that $|f| \leq |g|$ near a, and so $f = \mathcal{O}(g(x))$ as $x \to a$. □

For the case of powers, we have that

- $x^m = o(x^n),\ x \to 0 \Longleftrightarrow m > n,$
- $x^m = \mathcal{O}(x^n),\ x \to 0 \Longleftrightarrow m \geq n,$
- $f(x) = \mathcal{O}(x^n),\ x \to 0 \Longleftrightarrow f(x) = o(x^m),\ x \to 0$ for $m > n$.

We just say for $m > n$, $\mathcal{O}(x^n) = o(x^m)$, $x \to 0$. We have many other properties of small-o notation:

- If $f(x) = o(g)$ as $x \to a$, then $h(x)f(x) = o(hg)$, as $x \to a$.
- If $f, g = o(h)$ as $x \to a$, then $f + g = o(h)$, as $x \to a$.
- If $f(x) = o(g)$ and $h(x) = o(k)$ as $x \to a$, then $fh = o(gk)$, as $x \to a$.
- $x^m o(x^n) = o(x^{m+n})$, as $x \to 0$.
- $o(x^n) + o(x^n) = o(x^n)$, as $x \to 0$.
- $o(x^n) + o(x^m) = o(x^{m+n})$, as $x \to 0$.

Informally, we can state the first three these facts in a simpler way:

- $ho(g) = o(hg)$ as $x \to a$,
- $o(h) + o(h) = o(h)$, as $x \to a$,
- $o(g) \cdot o(k) = o(gk)$, as $x \to a$.

However, be careful about certain properties. Not everything acts the way we expect it to. Consider $f = o(g)$ and $h = o(k)$ as $x \to a$. This does NOT imply in general that

$$f + h = o(g + k), \quad x \to a,$$

or in other words, $o(g) + o(k) \neq o(g + k)$, as $x \to a$. Still, for $m > n$,

$$o(x^m) = o(x^n) = o(x^n) + o(x^n) = o(x^n), \quad x \to 0,$$

and thus

$$o(x^m) + o(x^n) = o(x^{\min\{m,n\}}), \quad x \to 0.$$

Example 12. Prove that $\mathcal{O}(g) \cdot o(f) = o(fg)$, as $x \to a$.

Proof. Let $F_1 = \mathcal{O}(g)$ as $x \to a$ and $F_2 = o(f)$ as $x \to a$. Then $F_1F_2 = o(fg)$ as $x \to a$. Indeed,

$$\frac{F_1F_2}{fg} = \left|\frac{F_1}{g}\right| \cdot \left|\frac{F_2}{f}\right| \to 0, \quad x \to a.$$

□

Example 13. Prove that if $f = o(g)$ as $x \to 0$, and $g = o(x^n)$ as $x \to 0$, then $f = o(x^n)$ as $x \to 0$.

Proof. We have

$$\frac{f}{x^n} = \frac{f}{g} \cdot \frac{g}{x^n} \to 0, \quad x \to 0.$$

□

Another two important properties are as follows (prove it!):

- $o(o(x^n)) = o(x^n)$, as $x \to 0$,
- $\mathcal{O}(o(x^n)) = o(x^n)$, as $x \to 0$.

8.4 Applications of Asymptotics

For the Taylor Polynomial, we claim that $R(x) = \mathcal{O}(x^{n+1})$. This is because

$$|R(x)| \le \left|\frac{1}{n!}\int_0^x (x-t)^n f^{(n+1)}(t)\,dt\right| \le \frac{1}{n!}\int_0^x |x-t|^n |f^{(n+1)}(t)|\,dt$$
$$\le \frac{\|f\|_{C([-R/2,R/2])}}{n!}|x|^{n+1}, \quad \forall x \in [-R/2, R/2].$$

Then because $f^{(n+1)}$ exists and is continuous near 0,

$$f(x) = \sum_{k=0}^{n} \frac{f^{(k)}(0)}{k!}x^k + o(x^n),\ x \to 0,$$

because $\mathcal{O}(x^{n+1}) = o(x^n)$, $x \to 0$. We can also represent power series using asymptotics.

(a) Exponential:

$$e^x = 1 + x + \frac{x^2}{2!} + \cdots + \frac{x^n}{n!} + o(x^n), \quad x \to 0.$$

(b) Sine:

$$\sin x = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \cdots + \frac{(-1)^n}{(2n+1)!}x^{2n+1} + o(x^{2n+2}), \quad x \to 0.$$

(c) Cosine:

$$\cos x = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \cdots + \frac{(-1)^n}{(2n)!}x^{2n} + o(x^{2n+1}), \quad x \to 0.$$

(d) Logarithm:

$$\ln(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \cdots + \frac{(-1)^{n+1}}{n}x^n + o(x^n), \quad x \to 0.$$

(e) Binomial:

$$(1+x)^\alpha = \sum_{n=0}^{N} C_\alpha^n x^n + o(x^N), \quad x \to 0.$$

Here, $\alpha \notin \mathbb{N}$, and

$$C_\alpha^n = \frac{\alpha(\alpha-1)\cdots(\alpha-n+1)}{n!}, \quad C_\alpha^0 = 1.$$

This gives us, for example, that

$$\frac{1}{1+x} = (1+x)^{-1} = 1 - x + x^2 - \cdots + (-1)^n x^n + o(x^n), \quad x \to 0.$$

Theorem 75. Assume that in a neighborhood of 0 we have

$$a_0 + a_1 x + \cdots + a_N x^N + o(x^N) \equiv b_0 + b_1 x + \cdots + b_N x^N + o(x^N).$$

Then $a_i = b_i$, for all i.

Using this theorem, we can do computations that would otherwise be very cumbersome.

Example 14. Compute $\tan^{(5)}(x)\Big|_{x=0}$.

Solution. Since we are always near 0, $o(\cdots)$ will all be $x \to 0$. Then

$$\tan x = \frac{\sin x}{\cos x} = \frac{x - \frac{x^3}{3!} + \cdots + o(x^6)}{1 - \frac{x^2}{2!} + \cdots + o(x^5)} = ax + bx^3 + cx^5 + o(x^5).$$

Notice that

$$\frac{x - \frac{x^3}{3!} + \cdots + o(x^6)}{1 - \frac{x^2}{2!} + \cdots + o(x^5)} = \frac{x - \frac{x^3}{3!} + \cdots + o(x^6)}{1 - [\frac{x^2}{2!} + \cdots + o(x^5)]},$$

and so we can use the expansion for $\frac{1}{1-x}$. Let

$$[\cdots] = \left[\frac{x^2}{2!} - \frac{x^4}{4!} + \cdots\right].$$

Then

$$\tan x = \left[x - \frac{x^3}{3!} + \frac{x^5}{5!} + o(x^6)\right]\left[1 + [\cdots] + [\cdots]^2 + \cdots\right]$$

The second term will be equal to

$$\left[1 - \frac{x^2}{4} + \frac{x^4}{24} + \cdots\right]\left[1 + \frac{x^4}{4} + \cdots\right] = \left(1 + \frac{x^2}{4} + \frac{5x^4}{24} + o(x^5)\right).$$

Thus we can multiply our two infinite sums together, disregarding any terms that are above the sixth power, and so

$$\tan x = x + \frac{1}{3}x^3 + \frac{2}{15}x^5 + o(x^6).$$

Hence we conclude that

$$\tan^{(5)}(x)\Big|_{x=0} = 5! \cdot \frac{2}{15} = \boxed{16}.$$

□

Example 15. Suppose now that we wanted to find

$$\arcsin^{(5)}(x)\Big|_{x=0}.$$

Solution. We can do this in two different ways.

- Method 1: Notice that arcsine is odd, and by Taylor,

$$\arcsin x = x + ax^3 + bx^5 + o(x^5).$$

  Our goal is to find b. Then we can take the sine of both sides to get

$$x = \sin(\arcsin x) = \sin(x + ax^3 + bx^5 + o(x^5)).$$

  Hence we use the sine expansion to calculate our derivative:

$$\sin t = (x + ax^3 + bx^5 + o(x^5)) + \frac{(x + ax^3 + bx^5 + o(x^5))^3}{3!} + \cdots$$

  However, this is a long and arduous task involving many calculations. Is there a simpler way to do this?

- Method 2: We take the derivative of arcsine first. This gives us

$$(\arcsin x)' = \frac{1}{\sqrt{1-x^2}} = (1-x^2)^{-1/2} = 1 + ax^2 + bx^4 + o(x^4),$$

  because the function is even. Then

$$\arcsin^{(5)}(x)\Big|_{x=0} = \left((1-x^2)^{-1/2}\right)^{(4)}\Big|_{x=0}$$

  Now we use the Binomial expansion with $\alpha = -1/2$ and do a substitution of $t = (-x)^2$, so that $o(t^2) = o(x^4)$, and so

$$(1-x^2)^{-1/2} = 1 + -\frac{1}{2}\cdot -x^2 + \frac{1}{2!}\cdot\frac{1}{2}\cdot\frac{3}{2}x^4 + o(x^4) = 1 + \frac{1}{2}x^2 + \frac{3}{8}x^4 + o(x^4).$$

  Then we conclude that

$$\arcsin^{(5)}(x)\Big|_{x=0} = \frac{3}{8}\cdot 4! = \boxed{9}.$$

□

We can also compute seemingly impossible limits using asymptotic analysis as well. Take the limit

$$\lim_{x\to 2}\left(\sqrt{3-x} + \ln(x/2)\right)^{1/\sin^2(x-2)}.$$

This gives us the indeterminate 1^∞ when we plug in 0. We can't seem to use traditional methods, such as L'Hôpital's Rule, to solve this limit.

Theorem 76. If

$$\frac{bx^n + o(x^n)}{ax^n + o(x^n)}, \quad x \to 0,$$

and $a \neq 0$, then

$$\lim_{x\to 0} \frac{bx^n + o(x^n)}{ax^n + o(x^n)} = \frac{b + o(1)}{a + o(1)} = \frac{b}{a}.$$

Example 16. Solve

$$\lim_{x\to 2} \left(\sqrt{3-x} + \ln(x/2)\right)^{1/\sin^2(x-2)}.$$

Solution. We first do a change of variables through $z = x-2$. The limit becomes

$$L = \lim_{z\to 0} \left(\sqrt{1-z} + \ln\left(\frac{z+2}{2}\right)\right)^{1/\sin^2 z}$$

Consider $\ln L$, which is

$$\ln L = \lim_{z\to 0} \frac{\ln(\sqrt{1-z} + \ln(1+\frac{z}{2}))}{\sin^2 z}$$

First we examine the denominator. We want to write $\sin^2 z = az^n + o(z^n)$. Hence, we can see that

$$\sin^2 z = z^2 + o(z^2)$$

Next we examine the numerator. We can expand $\sqrt{1-z}$ as

$$\sqrt{1-z} = 1 - \frac{1}{2}z - \frac{1}{8}z^2 + o(z^2)$$

We know that $\ln(1+\frac{z}{2})$ can be expanded through its formula:

$$\ln\left(1 + \frac{z}{2}\right) = \frac{1}{2}z - \frac{1}{8}z^2 + o(z^2)$$

So $\sqrt{1-z} + \ln(1+\frac{z}{2}) = 1 - \frac{1}{4}z^2 + o(z^2)$. Then using the ln expansion again,

$$\ln\left(\sqrt{1-z} + \ln\left(1 + \frac{z}{2}\right)\right) = -\frac{1}{4}z^2 + o(z^2)$$

Finally, our limit turns into the following expression, which we can easily evalute:

$$\lim_{z\to 0} \frac{-\frac{1}{4}z^2 + o(z^2)}{z^2 + o(z^2)} = -\frac{1}{4}$$

Then $L = e^{\ln L}$, and so

$$\ln L = -\frac{1}{4} \implies L = \boxed{e^{-\frac{1}{4}}}$$

□

9 Analysis in Metric Spaces

9.1 Introduction

We apply uniform convergence with integration and differentiation to $f_n(x) = c_n x^n$ and $\sum_{n=0}^{\infty} f_n(x) = \sum_{n=0}^{\infty} c_n x^n = c_0 + c_1 x + c_2 x^2 + \cdots$. Each $c_n x^n$ is infinitely differentiable at all x. What about $\sum_{n=0}^{\infty} c_n x^n$?

Theorem 77. Assume $\sum_{n=0}^{\infty} c_n x^n$ converges for $x = p$. Then $\sum_{n=0}^{\infty} c_n x^n$ converges for all x such that $|x| < |p|$. In other words, for all $x \in (-|p|, |p|)$.

This doesn't tell us anything about $|x| = |p|$, however. If we consider $\sum_{n=1}^{\infty} \frac{x^n}{n}$ at $x = -1$, then this is the alternating harmonic series, which converges, but if we consider it at $x = 1$, it will be the harmonic series, which diverges.

Proof. The key fact that we use is that $\sum_{n=0}^{\infty} q^n$ converges absolutely for $|q| < 1$, and $\sum_{n=1}^{\infty} q^n = \frac{1}{1-q}$.
For $x = 0$, we get that $c_0 + c_1 x + c_2 x^2 + \cdots = c_0$, and so it converges.
For $x \neq 0$, consider $\sum_{n=0}^{\infty} |c_n||x|^n$. We know that convergence at p implies that $c_n p^n \to 0$, and so $|c_n||p|^n \to 0$ as well. So

$$\begin{aligned} \sum_{n=0}^{\infty} |c_n||x|^n &= \sum_{n=0}^{\infty} |c_n||p|^n \cdot \frac{|x|^n}{|p|^n} \\ &\leq M \sum_{n=0}^{\infty} \left(\frac{|x|}{|p|}\right)^n \\ &< \infty, \end{aligned}$$

provided that $\frac{|x|}{|p|} < 1$, or equivalently $|x| < |p|$. Thus for all x such that $|x| < |p|$, our series converges absolutely. □

Up until now, we have studied $f : E \to \mathbb{R}$ for $E \subset \mathbb{R}$. Now we want to study $f(x_1, \ldots, x_n)$ and $E \subset \mathbb{R}^n$, such that $f : E \to \mathbb{R}^m$.

9.2 Metric and Normed Spaces

Definition 79 (Metric Spaces)**.** A *metric space* (M,d) is an ordered pair where M is a set and d, the *distance function*, is a function $d : M \times M \to \mathbb{R}$ that satisfies the following properties:

(a) Non-negativity:

$$d(x,y) \geq 0, \quad \forall x,y \text{ and } d(x,y) = 0 \iff x = y.$$

(b) Symmetry:

$$d(x,y) = d(y,x), \quad \forall x,y.$$

(c) Triangle Inequality:

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

In $\mathbb{R}$, $d(x,y) = |x-y|$.

Definition 80 (Ball)**.** An *open ball*, $B_r(a)$, is a ball centered at a with radius $r > a$, defined by

$$B_r(a) = \{x \in M \mid d(a,r) < r\}.$$

A *closed ball* is defined similarly:

$$\overline{B}_r(a) = \{x \in M \mid d(a,r) \leq r\}.$$

Definition 81 (Normed Space)**.** A *normed space* is an ordered pair $(V, \|\cdot\|)$, where V is a vector space, and $\|\cdot\|$, the *norm*, is a function $\|\cdot\| : V \to \mathbb{R}$ that satisfies the following properties:

(a) Positive-Definitive:

$$\|x\| = 0 \iff x = 0_V.$$

(b) Absolutely Homogeneous/Scalable:

$$\|\alpha x\| = |\alpha| \cdot \|x\|, \quad \forall \alpha \in \mathbb{R}, \forall x \in V.$$

(c) Triangle Inequality:

$$\|x+y\| \leq \|x\| + \|y\|, \quad \forall x,y.$$

Observe that $(V, \|\cdot\|)$ is a metric space where $d(u,v) = \|u-v\|$. Indeed,

$$d(x,y) = \|x-y\| = \|(x-z)+(z-y)\| \leq \|x-z\| + \|z-y\| = d(x,z) + d(z,y).$$

Some more examples:

- In $\mathbb{R}^n$, if $(V, \|\cdot\|)$ is a normed space, and $S \subset V$ is a subset but not necessarily a subspace. then (S, d), where $d(x, y) = \|x - y\|$ is also a metric space.
- If E is any set, define the vector space

$$B(E) = \left\{ f : E \to \mathbb{R} \mid \sup_E |f| < \infty \right\}.$$

  Then $(B(E), \|\cdot\|_{C(E)})$ is a normed space of infinite dimensions.
- For $[a, b] \in \mathbb{R}$, $C([a, b])$ is an infinite-dimensional vector space, and then $(C([a, b]), \|\cdot\|_1)$ is a normed space.

Definition 82 (L^p−norm)**.** For some $p \in \mathbb{R}$, we define the L^p−*norm* of x is defined by

$$\|x\|_p = (|x_1|^p + |x_2|^p + \cdots + |x_n|^p)^{1/p}.$$

The importance of normed spaces in $\mathbb{R}^n$ is that $\|x\|_p$ is a norm on $\mathbb{R}^n$ for all $1 \leq p < \infty$, including ∞, and that $(\mathbb{R}^n, \|\cdot\|_p)$ is a metric space for all $1 \leq p < \infty$, including ∞. We mainly focus on three L^p−norms:

(1) The L^1−norm:

$$\|x\|_1 = |x_1| + \cdots + |x_n|, \quad x \in \mathbb{R}^n,\ x = (x_1, \ldots, x_n).$$

(2) The L^2−norm:

$$\|x\|_2 = \sqrt{|x_1|^2 + \cdots + |x_n|^2}.$$

(3) The L^∞−norm:

$$\|x\|_\infty = \max_{i=1,\ldots,n} |x_i|.$$

Each of these norms, as expected, follow the triangle inequality:

- $p = 1$:

$$\begin{aligned} \|x + y\|_1 &= \|x_1 + y_1\| + \cdots + \|x_n + y_n\| \\ &\leq \|x_1\| + \|y_1\| + \cdots \\ &= \|x\|_1 + \|y\|_1 \end{aligned}$$

- $p = \infty$:

$$\begin{aligned} \|x+y\|_\infty &= \max_{i=1,\dots,n} (\|x_i + y_i\|) \\ &\le \max_{i=1,\dots,n} (\|x_i\| + \|y_i\|) \\ &\le \max_{i=1,\dots,n} \|x_i\| + \max_{i=1,\dots,n} \|y_i\| \\ &= \|x\|_\infty + \|y\|_\infty \end{aligned}$$

- $p = 2$: We introduce

$$x \cdot y = x_1 y_1 + \cdots + x_n y_n$$

as the *inner product*. Then we define $\|x\|_2 = \sqrt{x \cdot x}$. Finally, we use Cauchy-Schwartz:

$$\|x \cdot y\| \le \|x\|_2 \|y\|_2.$$

Hence

$$\{\text{metric spaces}\} \subset \{\text{normed vector spaces}\} \subset \{\text{spaces of } \mathbb{R}^n \text{ with } \|\cdot\|_p\}$$

9.3 Sequences in Metric Spaces

Definition 83 (Sequence)**.** Let (M, d) be a metric space. A *sequence* in M is any $x : \mathbb{N} \to M$, denoted x_n. Then we say that

$$\lim_{n\to\infty} x_n \to L,$$

iff

$$\forall \epsilon > 0, \exists N \text{ such that for all } n \ge N,\ d(x_n, L) < \epsilon.$$

This is equivalent to saying

$$\forall \epsilon > 0, \exists N \text{ such that for all } n \ge N,\ x_n \in B_\epsilon(L).$$

Theorem 78. If the limit to a sequence exists in (M, d), then it is unique.

Proof. Assume that there exists two limits L_1, L_2 such that $L_1 \ne L_2$. Then $d(L_1, L_2) = R > 0$. We claim that $B_{R/10}(L_1) \cap B_{R/10}(L_2) = \emptyset$. Suppose to the contrary that there exists a point in both balls. Then by the Triangle Inequality,

$$\begin{aligned} d(L_1, L_2) &\le d(L_1, p) + d(L_2, p) \\ &\le \frac{R}{10} + \frac{R}{10} \\ &= \frac{R}{5} < R. \end{aligned}$$

Let $\epsilon = R/10 > 0$. Then

$$\exists N_1 : n \geq N_1 \implies x_n \in B_{R/10}(L_1), \quad \exists N_2 : n \geq N_2 \implies x_n \in B_{R/10}(L_2).$$

Take $n_0 \geq N_1, N_2$. Then $x_{n_0} \in B_{R/10}(L_{1,2})$, a contradiction, since $B_{R/10}(L_1) \cap B_{R/10}(L_2) = \emptyset$. This completes our proof. □

Definition 84 (Bounded Set)**.** Let (M, d) be a metric space, and $S \subset M$. Then S is bounded if $S \subset \overline{B}_R(a)$ for some $R < \infty$, $a \in M$.

Theorem 79. If x_n converges in (M, d), then x_n is bounded in (M, d).

Proof. Fix $\epsilon = 1$. Let $\lim_{x\to\infty} x_n = L \in M$. By definition of $d(x_n, L) \to 0$ means that there exists some N such taht for $n \geq N$, $x_n \in \overline{B}_1(L)$. Let

$$R = \max\{1, d(L, x_1), d(L, x_2), \ldots, d(L, x_n)\} \neq \infty.$$

Then for all n, $x_n \in \overline{B}_R(L)$. □

Theorem 80. (a) For all $x \in \mathbb{R}^n$,

$$\|x\|_0 \leq \|x\|_1 \leq \sqrt{N}\|x\|_2 \leq N\|x\|_0.$$

(b) Let x_n be a sequence in $\mathbb{R}^n$ defined as

$$x_n = \begin{pmatrix} x_{1n} \\ x_{2n} \\ \vdots \\ x_{Nn} \end{pmatrix}.$$

Then

$$|x_n - p|_0 \to 0 \iff \{x_{j_n} \to p_j,\ n \to \infty,\ \forall j \in [1, N]\}.$$

(c) x_n converges in $\|\cdot\|_0$ iff it converges in $\|\cdot\|_1$ iff it converges in $\|\cdot\|_2$.

Proof. We have that (1) $\implies$ (3) at once. □

9.4 Compact Metric Spaces

Theorem 81. Let (M,d) be a metric space, and $K \subset M$. Then the following are equivalent:

(i) Every open cover of K has a finite subcover.

(ii) For all $S \subset K$, where $|S|$ is countable, there exists a limit point $p \in K$. (called *limit point compactness*)

(iii) For all x_n, for $x_n \in K$ for all n, there exists a subsequence $y_m = x_{n_m}$ such that $y_m \to a$ and $a \in K$. (called *sequential compactness*)

As a corollary, if (M,d) is a metric space, and K is compact, then K is closed and bounded. Unlike in $\mathbb{R}$, the converse is NOT necessarily true in general metric spaces!

Now we want to study the normed space $(C(K), \|\cdot\|_{C(K)})$. We know:

- $f_n \to f$ in $C(K)$ iff $f_n \to f$ uniformly on K,
- $f_n \to f$ in $C(K)$ iff $\|f_n - f\|_{C(K)} \to 0$ iff $\forall \epsilon > 0, \exists N$ such that if $n, m \geq N$,then $\|f_n - f_m\|_{C(K)} < \epsilon$. (Cauchy Criterion)

Before, for $K = E \subset \mathbb{R}$, these did not require that K be compact.

Theorem 82. Let (M,d) be a metric space, and $K \subset M$ be compact. Let $f \in C(K)$. Then

(i) $\sup_K |f| < \infty$, so $\|f\|_{C(K)} < \infty$.

(ii) There exists $x_* \in K$ such that $f(x_*) = \sup_K f$, and there exists $y_* \in K$ such that $f(y_*) = \inf_K f$.

In particular, if K is compact and f is continuous on K, then $\sup f = \max f$ and $\inf f = \min f$.

Proof. (i) Suppose to the contrary that $\sup_K |f| = \infty$. Then there exists some $x_n \in K$ such that $|f(x_n)| \to \infty$. By sequential compactness, we find $y_m = x_{n_m}$, a subsequence such that $y_m \to p \in K$. On one hand, $|f(y_m)| \to \infty$. On the other hand, f is continuous at p, so $f(y_n) \to f(p) \in \mathbb{R}$, a contradiction. □

Definition 85 (Pointwise Bounded)**.** We say that a sequence of functions $f_n : E \to \mathbb{R}$ is *pointwise bounded* on E if the sequence $f_n(x)$ is bounded for every $x \in E$. That is, if there exists a finite-valued function ϕ defined on

E such that

$$|f_n(x)| < \phi(x), \quad x \in E, n = 1, 2, \ldots.$$

Definition 86 (Uniform Bounded)**.** We say that a family of sets $\mathcal{F} = \{f_\alpha\}_{\alpha \in \mathbb{R}}$ is *uniformly bounded* if there exists $C < \infty$ such that

$$\sup_K |f_\alpha| = \max_K |f_\alpha| \le C, \quad \forall \alpha \in \mathcal{A}.$$

Definition 87 (Equicontinuity)**.** We say that a family of sets $\mathcal{F} = \{f_\alpha\}_{\alpha \in \mathbb{R}}$ is *equicontinuous* if

$$\forall \epsilon > 0, \exists \delta \text{ such that if } d(x,y) \le \delta, \ \alpha \in \mathcal{A}, \text{ then } |f_\alpha(x) - f_\alpha(y)| < \epsilon.$$

Example 17. Let A be a bounded subset of $C([0,1])$. Prove that the set of functions

$$F(x) = \int_0^x f(t)\,dt, \quad f \in A,$$

is equicontinuous in $C([0,1])$.

Proof. Denote

$$\mathcal{F} = \left\{ \int_0^x f(t)\,dt \mid f \in A \right\}.$$

We choose $M > 0$ such that for each $f \in A$, we have $\|f\|_\infty \le M$. Given $\epsilon > 0$, we let $\delta = \epsilon/M$. Then whenever $x, y \in [0,1]$ are such that $|x-y| < \delta$ and $f \in A$, it follows that

$$\left| \int_0^x f(t)\,dt - \int_0^y f(t)\,dt \right| = \left| \int_y^x f(t)\,dt \right| \le M|x-y| < M\delta = \epsilon.$$

Hence we conclude that $\mathcal{F}$ is equicontinuous. □

Example 18. Prove that the sequence of functions $\sin(nx)$ for $n \ge 1$ is not equicontinuous in $C([0,1])$.

Proof. Suppose to the contrary that this sequence is equicontinuous. In particular, there exists some δ so that for any n and $x, y \in \mathbb{R}$ with $|x-y| < \delta$,

$$|\sin nx - \sin ny| < 1.$$

Let n be large enough so that $\pi/2n < \delta$. Then $|\pi/2n - 0| < \delta$, yet

$$|\sin(n \cdot \pi/2n) - \sin(n \cdot 0)| = 1,$$

a contradiction. Hence this family of functions is not equicontinuous. □

Theorem 83 (Arzelà-Ascoli Theorem)**.** Let (M,d) be a metric space, and let $K \subset M$ be compact. Suppose $\mathcal{F} = \{f_\alpha\}_{\alpha\in\mathbb{R}}$ is a family of continuous functions; that is, $\mathcal{F} \subset C(K)$. Then $\mathcal{F}$ is compact if and only if $\mathcal{F}$ is uniformly bounded and equicontinuous.

While this is easy to state, this version of the Arzelà-Ascoli Theorem is difficult to prove. Instead, we reformulate the theorem into an equivalent, but different, way. To do this, we must introduce a new notion of compactness.

Definition 88 (Precompactness)**.** Let (M,d) be a metric space. We say $S \subset M$ is *precompact* iff $\overline{S}$ is compact.

This is a "weaker" form of compactness, as it gives us the condition that S must be closed. Some other notes:

- Formally, S is precompact iff for all $x_n \in S$, there exists a subsequence $y_m = x_{n_m}$ that converges to some $p \in M$. Notice that $p \notin S$ is possible.
- If $M = C(E)$, and $d(f,g) = \|f-g\|_{C(E)}$, then we have the Cauchy Criterion: f_n converges in $\|\cdot\|_{C(E)}$ iff f_n is Cauchy. Hence $S \in C(E)$ is precompact iff for all $f_n \in S$, there exists a Cauchy subsequence $g_m = f_{n_m}$ such that

$$\forall \epsilon > 0, \exists N \text{ such that for } n,m \geq N, \text{ we have } \|g_m - g_n\|_{C(E)} < \epsilon.$$

Now we can restate Arzelà-Ascoli using this idea.

Theorem 84 (Arzelà-Ascoli Theorem Version 2)**.** Let (M,d) be a metric space, andlet $K \subset M$ be compact. Suppose $\mathcal{F} = \{f_\alpha\}_{\alpha\in\mathbb{R}}$ is a family of continuous functions; that is, $\mathcal{F} \subset C(K)$. Then $\mathcal{F}$ is precompact if and only if $\mathcal{F}$ is bounded in $\|\cdot\|_{C(K)}$ and is equicontinuous.

10 Differentiability in $\mathbb{R}^n$

10.1 Introduction

Definition 89 (Multivariable Differentiability)**.** Let $\Omega \subset \mathbb{R}^n$ be open, and $f : \Omega \to \mathbb{R}^m$ be a function. Let $p \in \Omega$. We say f is differentiable at p iff there exists $A : \mathbb{R}^m \to \mathbb{R}^n$, a linear matrix (that is, an $m \times n$ matrix), such that

$$f(p+h) = f(p) + Ah + R(h), \qquad \frac{|R(h)|_2}{|h|_2} \to 0, \text{ as } h \to 0.$$

We use the notation

$$A \in \mathrm{Hom}(\mathbb{R}^n, \mathbb{R}^m) = \mathrm{Hom}(\mathbb{R}^m, \mathbb{R}^n)$$

to denote the set of all such matrices. Recall from linear algebra that the set of all linear transformations from two vector spaces V and W is denoted $\mathcal{L}(V, W)$. Indeed, by definition,

$$\mathrm{Hom}(\mathbb{R}^n, \mathbb{R}^m) = \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m) = M_{n\times m}.$$

We claim that this A is always unique.

Proof. Suppose that A and B both satisfy the definition. Then $(A-B)h = R_1(h) + R_2(h)$. Let

$$A - B = C = \begin{pmatrix} c_{11} & \cdots & c_{1m} \\ \vdots & \ddots & \vdots \\ c_{n1} & \cdots & c_{nm} \end{pmatrix}.$$

Assume, say $c_{11} \neq 0$. Take

$$h = \begin{pmatrix} t \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

Then

$$(A-B)h = \begin{pmatrix} c_{11}t \\ 0 \\ \vdots \\ 0 \end{pmatrix} = R_1(h) + R_2(h) \implies \begin{pmatrix} c_{11} \\ 0 \\ \vdots \\ 0 \end{pmatrix} = \frac{R_1(te_1) + R_2(te_2)}{t} \to 0,$$

by assumption. Thus $c_{11} = 0$, a contradiction. □

Definition 90 (Differential)**.** We call such an A in the definition of differentiability the *differential* of f at p, denoted

$$df_p \in \mathrm{Hom}(\mathbb{R}^n, \mathbb{R}^m).$$

There are many notations for this, including

$$A = \mathcal{D}f_p = df_p = f_p' = f_{*,p}.$$

Much like in one variable, differentiability at p will always imply continuity at p as well. Much like in a single variable, we have many expected properties:

- If df_p and dg_p exist, and $f, g : \Omega \to \mathbb{R}^m$, then $d(f+g)_p = df_p + dg_p$.
- If $f \in \mathrm{Hom}(\mathbb{R}^n, \mathbb{R}^m)$, then for all p, $(df)_p = f$.
- For $p_{r_j} \in \mathrm{Hom}(\mathbb{R}^n, \mathbb{R})$, $p_{r_j} : \mathbb{R}^n\ to\mathbb{R}$, we have

$$\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \mapsto x_j.$$

- Recall from linear algebra that

$$e_j \in \mathbb{R}^m = \begin{pmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{pmatrix}, \quad \text{where 1 is in the } j^{th} \text{ row.}$$

  For all $f : \Omega \to \mathbb{R}^m$, where $\Omega \subset \mathbb{R}^m$, we have

$$f = \sum_{j=1}^{m} (p_{r_j} \circ f) e_j,$$

  and $p_{r_j} \circ f : \Omega \to \mathbb{R}$.

Theorem 85. Let $\Omega \subset \mathbb{R}^n$ be open, with $p \in \Omega$ and $f : \Omega \to \mathbb{R}^m$ differentiable at p. Let $G \subset \mathbb{R}^m$ be open, $f(\Omega) \subset G$, and $g : G \to \mathbb{R}^k$ be differentiable at $f(p)$. Then $g \circ f$ is differentiable at p, and

$$d(g \circ f)_p = (dg)_{f(p)} \circ (df)_p.$$

Pictorially,

$$\mathbb{R}^n \xrightarrow{df_p} \mathbb{R}^m \xrightarrow{dg_p} \mathbb{R}^k,$$

and the arrow from $\mathbb{R}^n \to \mathbb{R}^k$ is $dg_{f(p)} \circ df_p$, a product of matrices.

Proof. Let $p \in \Omega$ and $h \in \mathbb{R}^n$. Then

$$\forall h,\ f(p+h) = f(p) + df_{p(h)} + r_1(h), \quad \forall z,\ g(f(p)+z) = g(f(p)) + d_{g(p)}(z) + r_2(z).$$

Now

$$\begin{aligned}(g \circ f)(p+h) = g(f(p+h)) = g(f(p) + \frac{d(f_p)h + r_1(h)}{z} \\ &= g(f(p)) + (dg_{f(p)} \circ df_p)(h) + dg_{f(p)}(r_1(h)) + r_2((df)_p h + r_1(h)) \\ &= (g \circ f)(p) + Ah + R(h)\end{aligned}$$

It follows that

$$\begin{cases} \frac{|r_1(h)|}{|h|} \to 0, & h \to 0 \\ \frac{|r_2(z)|}{|z|} \to 0, & z \to 0 \\ z = (df_p)h + r_1(h) \end{cases} \implies \frac{|R(h)|}{|h|} \to 0, \quad h \to 0,$$

because

$$\frac{|(dg_{f(p)})(r_1(h))|}{|h|} \leq \frac{\|dg_{f(p)}\| \cdot |r_1(h)|}{|h|} \to 0, \quad h \to 0.$$

The process for r_2 is similar. □

For $A \in \mathrm{Hom}(\mathbb{R}^n, \mathbb{R}^m)$, we define a norm on A as

$$\|A\| = \|[a_{ij}]\| = \sum_{i,j=1}^{n,m} |a_{ij}|.$$

Using this norm, $(\mathrm{Hom}(\mathbb{R}^n, \mathbb{R}^m), \|\cdot\|)$ is a normed space.

Theorem 86. $f : \Omega \to \mathbb{R}^m$ is differentiable at $p \in \Omega$ if and only if for all $j \in [1, n]$, the funcutions $p_{r_j} \circ f : \Omega \to \mathbb{R}$ are differentiable at p.

Proof. At once from sum rule and chain rule. □

Definition 91 (Partial Derivative)**.** Let $f : \Omega \to \mathbb{R}$, $p \in \Omega$, and $v \in \mathbb{R}^n$ for $v \neq 0$. We define the *partial derivative* of f at p to be

$$\left.\frac{\partial f}{\partial v}\right|_p = \lim_{t \to 0} \frac{f(p+tv) - f(p)}{t}.$$

How are the differential and partial derivative related?

$$\begin{aligned}\left.\frac{\partial f}{\partial v}\right|_p &= \lim_{t \to 0} \frac{f(p+tv) - f(p)}{t} \\ &= \lim_{t \to 0} \frac{f(p) + df_p(tv) + (tv) - f(p)}{t} \\ &= (df_p)v + \lim_{t \to 0} \frac{r(tv)}{t}\end{aligned}$$

Thus if df_p exists, then

$$\left.\frac{\partial f}{\partial v}\right|_p = df_p(v), \quad v \in V \neq 0.$$

10.2 Multivariable Derivative Theorems

10.3 Prelude to Differential Geometry

11 Multivariate Integration

12 Fourier Analysis

13 Measure Theory

13.1 Outer Measure

How would we formally define the notion of the length of an interval? This should be relatively simple to do.

Definition 92 (Length of Open Interval)**.** The *length* $\ell(I)$ of an open interval I is defined to be

$$\ell(I) = \begin{cases} b-a, & \text{if } I = (a,b),\ a,b \in \mathbb{R},\ a<b \\ 0, & \text{if } I = \emptyset \\ \infty, & \text{if } I = (-\infty, a) \text{ or } I = (a,\infty),\ a \in \mathbb{R} \\ \infty, & \text{if } I = (-\infty,\infty) \end{cases}$$

Given a subset of $\mathbb{R}$, $A \subset \mathbb{R}$, we would expect that the size of A cannot be larger than the sum of the lengths of a sequence of open intervals whose union contains A. We can take the infimum of all such sums to get a "size" of A.

Definition 93 (Outer Measure)**.** The *outer measure* $m^*(A)$ of a set $A \subset \mathbb{R}$ is defined by

$$m^*(A) = \inf\left\{\sum_{k=1}^{\infty} \ell(I_k) : I_1, I_2, \ldots \text{ are open intervals such that } A \subset \bigcup_{k=1}^{\infty} I_k\right\}.$$

What will be the outer measure of an aribtrary finite set? Suppose $A = \{a_1, \ldots, a_n\}$ is a finite subset of $\mathbb{R}$. Let $\epsilon > 0$. We define a sequence of open intervals $I_1, I_2, \ldots$ by

$$I_k = \begin{cases} (a_k - \epsilon, a_k + \epsilon), & k \le n \\ \emptyset, & k > n \end{cases}.$$

The union of our sequence of open intervals will contain A. Then $\sum_{k=1}^{\infty} \ell(I_k) = 2\epsilon n$. Hence $m^*(A) \le 2\epsilon n$, but since ϵ is arbitrary, this tells us that $m^*(A) = 0$. So we know that finite sets will have outer measure zero. This is not the strongest argument we can make; in fact, countable sets will have outer measure zero as well.

Theorem 87. Every countable subset of $\mathbb{R}$ has outer measure zero.

Proof. Let $A = \{a_1, a_2, \dots\}$ be a countable subset of $\mathbb{R}$. Let $\epsilon > 0$. Then for $k \in \mathbb{N}$, let

$$I_k = \left(a_k - \frac{\epsilon}{2^k}, a_k + \frac{\epsilon}{2^k}\right).$$

The union of our sequence of open intervals will contain A. Then $\sum_{k=1}^{\infty} \ell(I_k) = 2\epsilon$, and so $m^*(A) \leq 2\epsilon$, which tells us that indeed $m^*(A) = 0$. □

What about set inclusion? Does taking the outer measure preserve order like we expect it to? Indeed, it does.

Theorem 88. Suppose A and B are subsets of $\mathbb{R}$ such that $A \subset B$. Then $m^*(A) \leq m^*(B)$.

Proof. Let $I_1, I_2, \dots$ be a sequence of open intervals whose union contains B, and as a result, also contain A. This means that

$$m^*(A) \leq \sum_{k=1}^{\infty} \ell(I_k).$$

When we take the infimum over all sequences of open intervals whose union contains B, then $m^*(A) \leq m^*(B)$. □

Definition 94 (Translation)**.** If $t \in \mathbb{R}$ and $A \subset \mathbb{R}$, the *translation* $t + A$ is defined by

$$t + A = \{t + a \mid a \in A\}.$$

What about translation-invariance? It would make sense that shifting a set left or right shouldn't affect its outer measure.

Theorem 89. Suppose $t \in \mathbb{R}$ and $A \subset \mathbb{R}$. Then $m^*(t + A) = m^*(A)$.

Proof. Let $I_1, I_2, \dots$ be a sequence of open intervals whose union contains A. Then $t + I_1, t + I_2, \dots$ is a sequence of open intervals whose union contains $t + A$. Thus

$$m^*(t + A) \leq \sum_{k=1}^{\infty} \ell(t + I_k) = \sum_{k=1}^{\infty} I_k.$$

When we take the infimum of the last term over all sequences, we have $m^*(t + A) \leq m^*(A)$. For the other direction, notice that $A = -t + (t + A)$. We apply our inequality from above, replacing A with $t + A$ and t with $-t$. Hence $m^*(A) = m^*(-t + (t + A)) \leq m^*(t + A)$. Combining both sides, we have our desired result. □

Consider the union of intervals of $(1,4)$ and $(3,5)$. This will be $(1,5)$. Notice that

$$4 = \ell((1,4)\cup(3,5)) < \ell(((1,4)+\ell(3,5)) = 5,$$

and this makes sense because $(3,4)$ is counted twice on the right side. This is the idea of *countable subadditivity*, another property that holds for outer measures.

Theorem 90. Suppose $A_1, A_2, \dots$ is a sequence of subsets of $\mathbb{R}$. Then

$$m^*\left(\bigcup_{k=1}^{\infty} A_k\right) \le \sum_{k=1}^{\infty} m^*(A_k).$$

Proof. Assume $m^*(A_k) < \infty$ for all $k \in \mathbb{Z}^+$, as if this is not true, the inequality always holds. Let $\epsilon > 0$. For each $k \in \mathbb{Z}^+$, let $I_{1,k}, I_{2,k}, \dots$ be a sequence of open intervals whose union contains A_k such that

$$\sum_{j=1}^{\infty} \ell(I_{j,k}) \le \frac{\epsilon}{2^k} + m^*(A_k).$$

Hence

$$\sum_{k=1}^{\infty}\sum_{j=1}^{\infty} \ell(I_{j,k}) \le \epsilon + \sum_{k=1}^{\infty} m^*(A_k).$$

We can rearrange $\{I_{j,k} \mid j,k \in \mathbb{Z}^+\}$ into a sequence of open intervals whose union contains $\bigcup_{k=1}^{\infty} A_k$ by adjoining $k-1$ intervals whose indices add up to k for step k. That is,

$$(I_{1,1}), (I_{2,1}, I_{1,2}), (I_{1,3}, I_{2,2}, I_{3,1}), \dots$$

The inequality above tells us that the sum of the intervals above is less than or equal to $\epsilon + \sum_{k=1}^{\infty} m^*(A_k)$. Thus,

$$m^*\left(\bigcup_{k=1}^{\infty} A_k\right) \le \epsilon + \sum_{k=1}^{\infty} m^*(A_k),$$

but since ϵ is arbitrary, we achieve our desired result. ☐

This property also implies another property of interest, known as *finite subadditivity*, which means that

$$|A_1 \cup \cdots \cup A_n| \le |A_1| + \cdots + |A_n|,$$

for all $A_1, \dots, A_n \subset \mathbb{R}$. This is because we take $A_k = \emptyset$ for $k > n$ in the proof above.

What about closed intervals? We expect that $m^*([a,b]) = b - a$. The $\le$ side of the equality is obvious, how how about the $\ge$ side? This requires a bit more thinking, and requires the use of our old friend Heine-Borel.

Theorem 91. Suppose $a, b \in \mathbb{R}$, with $a < b$. Then $m^*([a,b]) = b - a$.

Proof. Let $\epsilon > 0$. Then $(a-\epsilon, b+\epsilon), \emptyset, \emptyset, \ldots$ is a sequence of open intervals whose union contains $[a,b]$. Thus $m^*([a,b]) \leq b - a + 2\epsilon$, and since ϵ is arbitrary, we conclude that $m^*([a,b]) \leq b - a$.
Now let $I_1, I_2, \ldots$ be a sequence of open intervals such that $[a,b] \subset \bigcup_{k=1}^{\infty} I_k$. By Heine-Borel, there exists $n \in \mathbb{Z}^+$ such that $[a,b] \subset I_1 \cup \cdots \cup I_n$. We proceed with induction that

$$\sum_{k=1}^{n} \ell(I_k) \geq b - a.$$

We have already proved the base case of $n = 1$. For $n > 1$, suppose $I_1, \ldots, I_n, I_{n+1}$ are open intervals such that $[a,b] \subset I_1 \cup \cdots \cup I_n \cup I_{n+1}$. This means b is in at least one of our intervals. By relabeling, assume $b \in I_{n+1}$. Let $I_{n+1} = (c,d)$. If $c \leq a$, then $\ell(I_{n+1}) \geq b - a$ and we're done. Hence suppose that $a < c < b < d$. Then $[a,c] \subset I_1 \cup \cdots \cup I_n$. By our inductive hypothesis, $\sum_{k=1}^{n} \ell(I_k) \geq c - a$. Hence

$$\begin{aligned}\sum_{k=1}^{n+1} \ell(I_k) &\geq (c-a) + \ell(I_{n+1}) \\ &= (c-a) + (d-c) \\ &= d - a \\ &\geq b - a.\end{aligned}$$

□

As a direct corollary, we find out that nontrivial intervals are uncountable. That is, every interval in $\mathbb{R}$ that contains at least two distinct elements is uncountable.

Proof. Let I be an interval that contains $a, b \in \mathbb{R}$ with $a < b$. Then

$$m^*(I) \geq m^*([a,b]) = b - a > 0.$$

Since every countable subset of $\mathbb{R}$ has outer measure 0, we can conclude that I is uncountable. □

Not every property that we expect to be satisfied will hold for the outer measure; namely, the idea of *additivity*.

Theorem 92. There exist disjoint sets A and B of $\mathbb{R}$ such that

$$m^*(A \cup B) \neq m^*(A) + m^*(B).$$

13.2 Measurability

We have shown that the outer measure doesn't have all of the desired properties that we would expect for our idea of size of a set. Does there exist a function or "measure" that satisfies some properties for all subsets of $\mathbb{R}$?

Theorem 93 (Nonexistence of Measure)**.** There does not exist a function μ with all of the following properties:

(a) μ is a function from the set of subsets of $\mathbb{R}$ to $[0, \infty]$.

(b) $\mu(I) = \ell(I)$ for every open interval I of $\mathbb{R}$.

(c) Countable Additivity:

$$\mu\left(\bigcup_{k=1}^{\infty} A_k\right) = \sum_{k=1}^{\infty} \mu(A_k),$$

for every disjoint sequence $A_1, A_2, \ldots$ of subsets of $\mathbb{R}$.

(d) Translation Invariance:

$$\mu(t + A) = \mu(A),$$

for every $t \in \mathbb{R}$ and $A \subset \mathbb{R}$.

Proof. Suppose that there did exist a function μ with all of the properties. Observe that $\mu(\emptyset) = 0$ from (b). If $A \subset B \subset \mathbb{R}$, then $\mu(A) \le \mu(B)$ □

How do we remedy this fact? We cannot loosen (b) because the size of an interval must be its length. We cannot loosen (c) because countable additivity is essential to prove limit theorems. We cannot loosen (d) because that would be very counter-intuitive for what "length" is. This leaves us with (a) as our only option to loosen our definition.

Definition 95 (σ−algebra)**.** Suppose X is a set and $\mathcal{S}$ is a set of subsets of X. Then $\mathcal{S}$ is called a *σ−algebra* on X if the following conditions are satisfied:

- $\emptyset \in \mathcal{S}$,
- If $E \in \mathcal{S}$, then $X \setminus E \in \mathcal{S}$,

- If $E_1, E_2, \ldots$ is a sequence of elements of $\mathcal{S}$, then

$$\bigcup_{k=1}^{\infty} E_k \in \mathcal{S}.$$

As an example, let X be a set. Then $\{\emptyset, X\}$ will be a σ−algebra on X. $\mathcal{P}(X) = 2^X$, the power set of X, will also be a σ−algebra on X. We have some more properties of σ−algebras.

Theorem 94. Let $\mathcal{S}$ be a σ−algebra on X. Then

(a) $X \in \mathcal{S}$,

(b) If $D, E \in \mathcal{S}$, then $D \cup E \in \mathcal{S}$ and $D \cap E \in \mathcal{S}$ and $D \setminus E \in \mathcal{S}$,

(c) If $E_1, E_2, \ldots$ is a sequence of elements of $\mathcal{S}$, then

$$\bigcap_{k=1}^{\infty} E_k \in \mathcal{S}.$$

(c) is what we call closure under countable intersections.

Definition 96 (Measurable Space)**.** A *measurable space* is an ordered pair $(X, \mathcal{S})$, where X is a set and $\mathcal{S}$ is a σ−algebra on X. An element of $\mathcal{S}$ is called an $\mathcal{S}$*−measurable* set, or just simply a *measurable set* if $\mathcal{S}$ is clear from the context.

For example, if $X = \mathbb{R}$ and $\mathcal{S}$ is the set of all subsets of $\mathbb{R}$ that are countable or have a countable component, then the set of rational numbers is $\mathcal{S}$−measurable but the set of positive real numbers is not.

Theorem 95. Suppose X is a set and $\mathcal{A}$ is a set of subsets of X. Then the intersection of all σ−algebras on X that contain $\mathcal{A}$ is a σ−algebra on X.

Consider some examples:

- If X is a set and $\mathcal{A}$ is the set of subsets of X that consist of exactly one element; that is,

  $$\mathcal{A} = \{\{x\} \mid x \in X\},$$

  then the smallest σ−algebra on X containing $\mathcal{A}$ is the set of all subsets E of X such that E is countable or $X \setminus E$ is countable.

- If $\mathcal{A} = \{(0,1), (0,\infty)\}$, then the smallest σ−algebra on $\mathbb{R}$ containing $\mathcal{A}$ is

  $$\{\emptyset, (0,1), (0,\infty), (-\infty,0] \cup [1,\infty), (-\infty,0], [1,\infty), (-\infty,1), \mathbb{R}\}.$$

Definition 97 (Borel set)**.** The smallest $\sigma-$algebra on $\mathbb{R}$ containing all open subsets of $\mathbb{R}$ is called the collection of *Borel subsets* of $\mathbb{R}$. An element of this $\sigma-$algebra is called a *Borel set*.

Some examples:

- Every closed subset of $\mathbb{R}$ is a Borel set because every closed subset of $\mathbb{R}$ is the complement of an open subset of $\mathbb{R}$.
- Every countable subset of $\mathbb{R}$ is a Borel set because if $B = \{x_1, x_2 \dots\}$, then $B = \bigcup_{k=1}^{\infty} \{x_k\}$, which is a Borel set because $\{x_k\}$ is a Borel set.
- Every half open interval $[a, b)$ is a Borel set because $[a, b) = \bigcap_{k=1}^{\infty} (a - \frac{1}{k}, b)$.
- If $f : \mathbb{R} \to \mathbb{R}$ is a function, then the set of points at which f is continuous is the intersection of a sequence of open sets, and thus is a Borel set.

There do exist subsets of $\mathbb{R}$ that are not Borel sets. However, any subset of $\mathbb{R}$ that we can write in a concrete fashion will be a Borel set.

Recall that the *inverse image* (or preimage) of a set A given a function $f : X \to Y$ and $A \subset Y$ is defined to be

$$f^{-1}(A) = \{x \in X \mid f(x) \in A\}.$$

Recall from Math 8 that inverse images have the following nice properties:

- $f^{-1}(Y \setminus A) = X \setminus f^{-1}(A)$, for every $A \subset Y$.
- $f^{-1}\left(\bigcup_{A \in \mathcal{A}} A\right) = \bigcup_{A \in \mathcal{A}} f^{-1}(A)$ for every set $\mathcal{A}$ of subsets of Y.
- $f^{-1}\left(\bigcap_{A \in \mathcal{A}} A\right) = \bigcap_{A \in \mathcal{A}} f^{-1}(A)$ for every set $\mathcal{A}$ of subsets of Y.

We also know that for compositions, where if $f : X \to Y$ and $g : Y \to W$ are functions, then

$$(g \circ f)^{-1}(A) = f^{-1}(g^{-1}(A)),$$

for all $A \subset W$. Using these ideas, we can finally define what it means for a function to be *measurable*.

Definition 98 (Measurable Function)**.** Suppose $(X, \mathcal{S})$ is a measurable space. A function $f : X \to \mathbb{R}$ is called $\mathcal{S}-$*measurable* (or simply just *measurable* if it is clear from the context) if

$$f^{-1}(B) \in S,$$

for every Borel set $B \subset \mathbb{R}$.

Suppose our σ−algebra is just $\{\emptyset, X\}$. Then the only measurable functions from X to $\mathbb{R}$ are constant functions. If our σ−algebra was $\mathcal{P}(X)$, then every function from X to $\mathbb{R}$ would be measurable.

Definition 99 (Characteristic Function)**.** Suppose $E \subset X$. The *characteristic function* of E is the function $\chi_E : X \to \mathbb{R}$ defined by

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

What would be the inverse image of a characteristic function? Let $(X, \mathcal{S})$ be a measurable space, with $E \subset X$ and $B \subset \mathbb{R}$. Then

$$\chi_E^{-1}(B) = \begin{cases} E, & 0 \notin B \text{ but } 1 \in B \\ X \setminus E, & 0 \in B \text{ but } 1 \notin B \\ X, & 0 \in B \text{ and } 1 \in B \\ \emptyset, & 0 \notin B \text{ and } 1 \notin B \end{cases}$$

From this, by definition of measurable, χ_E is $\mathcal{S}$−measurable if and only if $E \in \mathcal{S}$.

Theorem 96 (Measurability Criterion)**.** Suppose $(X, \mathcal{S})$ is a measurable space and $f : X \to \mathbb{R}$ is a function such that

$$f^{-1}((a, \infty)) = \{x \in X \mid f(x) > a\} \in \mathcal{S},$$

for all $a \in \mathbb{R}$. Then f is an $\mathcal{S}$−measurable function.

This condition is equivalent to the following:

- $f^{-1}((a, \infty)) = \{x \in X \mid f(x) \leq a\} \in \mathcal{S}$,
- $f^{-1}((a, \infty)) = \{x \in X \mid f(x) < a\} \in \mathcal{S}$,
- $f^{-1}((a, \infty)) = \{x \in X \mid f(x) \geq a\} \in \mathcal{S}$.

The first point follows because it is the complement to the Measurability criterion. The other two follow because the complement in $\mathcal{S}$ of a measurable subset of $\mathcal{S}$ is also measurable.

Definition 100 (Borel-Measurable)**.** Let $X \subset \mathbb{R}$. A function $f : X \to \mathbb{R}$ is called *Borel-measurable* if $f^{-1}(B)$ is a Borel set for every Borel set $B \subset \mathbb{R}$.

How is measurability affected by functions or algebraic operations?

Theorem 97. Every continuous real-valued function defined on a Borel subset of $\mathbb{R}$ is a Borel measurable function.

Proof. Let $X \subset \mathbb{R}$ be a Borel set and $f : X \to \mathbb{R}$ continuous. To prove that f is Borel-mesaurable, fix $a \in \mathbb{R}$. If $x \in X$ and $f(x) > a$, then there exists $\delta_x > 0$ such that $f(y) > a$ for all $y \in (x - \delta_x, x + \delta_x) \cap X$. Thus

$$f^{-1}((a,\infty)) = \left(\bigcup_{x \in f^{-1}((a,\infty))} (x - \delta_x, x + \delta_x) \right) \cap X.$$

The union is an open subset of $\mathbb{R}$, and so its intersection with X is a Borel set. Thus $f^{-1}((a,\infty))$ is a Borel set. □

Theorem 98. Every increasing function defined on a Borel subset of $\mathbb{R}$ is a Borel measurable function.

Proof. Let $X \subset \mathbb{R}$ be a Borel set and $f : X \to \mathbb{R}$ be increasing. To prove that f is Borel meausrable, fix $a \in \mathbb{R}$. Let $b = \inf f^{-1}((a,\infty))$. Then obviously,

$$f^{-1}((a,\infty)) = (b,\infty) \cap X, \quad \text{or} \quad f^{-1}((a,\infty)) = [b,\infty) \cap X.$$

Either way, $f^{-1}((a,\infty))$ is a Borel set. □

Measurability also interacts well with composition.

Theorem 99. Supppose $(X, \mathcal{S})$ is a measurable space and $f : X \to \mathbb{R}$ is an $\mathcal{S}$–measurable function. Suppose g is a real-valued Borel measurable function defined on a subset of $\mathbb{R}$ that includes the range of f. Then $g \circ f : X \to \mathbb{R}$ is an $\mathcal{S}$–measurable function.

Proof. Let $B \subset \mathbb{R}$ be a Borel set. By properties of inverse images,

$$(g \circ f)^{-1}(B) = f^{-1}(g^{-1}(B)).$$

Because g is Borel-measurable, $g^{-1}(B)$ is a Borel subset of $\mathbb{R}$. Because f is an $\mathcal{S}$–measurable function, $f^{-1}(g^{-1}(B)) \in \mathcal{S}$. Thus the equation above implies that $(g \circ f)^{-1}(B) \in \mathcal{S}$, and so it is $\mathcal{S}$–measurable by definition. □

This has a wide variety of implications. For once, given a measurable function $f : X \to \mathbb{R}$, we know that, for example, $-f, \frac{1}{2}f, |f|$, and f^2 will all be measurable functions, as we can take compositions of $g(x) = -x$, $g(x) = \frac{1}{2}x$, and so on. As for algebraic operations, measurability also acts like we expect it to.

Theorem 100. Suppose $(X, \mathcal{S})$ is a measurable space and $f, g : X \to \mathbb{R}$ are $\mathcal{S}$−measurable. Then the following are all $\mathcal{S}$−measurable functions:

- $f + g$,
- $f - g$,
- fg,
- $\frac{f}{g}$, provided $g(x) =\neq 0$ for all $x \in X$.

Proof.

- We want to show that for $a \in \mathbb{R}$,

$$(f+g)^{-1}((a,\infty)) = \bigcup_{r\in\mathbb{Q}} \Big(f^{-1}((r,\infty)) \cap g^{-1}((a-r,\infty)) \Big),$$

  and so $(f+g)^{-1}((a,\infty)) \in \mathcal{S}$. Let $x \in (f+g)^{-1}((a,\infty))$. Then $a < f(x) + g(x)$, which means that the open interval $(a - g(x), f(x))$ is nonempty, and so it contains some rational number r. It follows that $r < f(x)$, and so $x \in f^{-1}((r,\infty))$, and $a - g(x) < r$. Hence $x \in g^{-1}((a-r,\infty))$. This proves the forward direction. For the backward direction, let $x \in f^{-1}((r,\infty)) \cap g^{-1}((a-r,\infty))$, for some $r \in \mathbb{Q}$. Then $r < f(x)$ and $a - r < g(x)$. Adding these, we find $a < f(x) + g(x)$, and we're done.

- Since $-g$ is an $\mathcal{S}$−measurable function, $f + (-g)$ is also an $\mathcal{S}$−measurable function.

- Notice that we can rewrite

$$fg = \frac{(f+g)^2 - f^2 - g^2}{2},$$

  and since each component is individually $\mathcal{S}$−measurable, their sum will also be $\mathcal{S}$−measurable.

- Suppose $g(x) \neq 0$ for all $x \in X$. The function defined on $\mathbb{R} \setminus \{0\}$ that maps $x \mapsto \frac{1}{x}$ is continuous, and thus Borel-measurable. We know that $\frac{1}{g}$ is $\mathcal{S}$−measurable, and so $\frac{f}{g}$ will be $\mathcal{S}$−measurable as well.

□

What happens when we take a pointwise limit of measurable functions? For Riemann integration, this does not guarantee that the limit be Riemann integrable. Thus it would be nice to have this property available to us.

Theorem 101. Let $(X, \mathcal{S})$ be a measurable space and $f_1, f_2, \ldots$ a sequence of measurable functions from $X \to \mathbb{R}$. Suppose $\lim_{k\to\infty} f_k(x)$ exists for each $x \in X$. Define $f : X \to \mathbb{R}$ by

$$f(x) = \lim_{k\to\infty} f_k(x).$$

Then f is $\mathcal{S}$−measurable.

Proof. We want to show that for $a \in \mathbb{R}$,

$$f^{-1}((a,\infty)) = \bigcup_{j=1}^{\infty} \bigcup_{m=1}^{\infty} \bigcap_{k=m}^{\infty} f_k^{-1}\left(\left(a + \frac{1}{j}, \infty\right)\right).$$

Let $x \in f^{-1}((a,\infty))$. This means there exists $j \in \mathbb{Z}^+$ such that $f(x) > a + \frac{1}{j}$. By limit definition, there exists some $m \in \mathbb{Z}^+$ such that $f_k(x) > a + \frac{1}{j}$ for $k \geq m$. Thus x is in the right side of the equation.

Now let x be in the right side. This means there exists $j, m \in \mathbb{Z}^+$ such that $f_k(x) > a + \frac{1}{j}$, for all $k \geq m$. Taking the limit as $k \to \infty$, this means $f(x) \geq a + \frac{1}{j} > a$. Thus x is in the left side. □

All of our previous definitions, theorems, etc. can be stated for the *extended real line*, or $\overline{R} = \mathbb{R} \cup \{\pm\infty\}$. Usually we will denote this as $[-\infty, \infty]$, to show that infinity is included in our interval. Some theorems can only be stated for the extended real numbers, however, such as the following theorem.

Theorem 102. Suppose $(X, \mathcal{S})$ is a measurable space and $f_1, f_2, \ldots$ is a sequence of $\mathcal{S}$−measurable functions from X to $[-\infty, \infty]$. Define $g, h : X \to [-\infty, \infty]$ by

$$g(x) = \inf\{f_k(x) \mid k \in \mathbb{Z}^+\}, \quad h(x) = \sup\{f_k(x) \mid k \in \mathbb{Z}^+\}.$$

Then g and h are $\mathcal{S}$−measurable.

Proof. Let $a \in \mathbb{R}$. Then by definition of supremum,

$$h^{-1}((a,\infty]) = \bigcup_{k=1}^{\infty} f_k^{-1}((a,\infty]).$$

This, combined with the measurability criterion, tells us that h is indeed $\mathcal{S}$−measurable. Likewise, we can write

$$g(x) = -\sup\{-f_k(x) \mid k \in \mathbb{Z}^+\},$$

for all $x \in X$. Hence g is also $\mathcal{S}$−measurable. □

Having discussed measurability, the natural next step would be to discuss *nonmeasurable* sets. Surely these must exist, otherwise what's the point of defining a measurable set? We can examine this under the context of the outer measure.

Theorem 103. *Let $X \subset \mathbb{R}$ be a bounded measurable set. Suppose there is a bounded, countable set of real numbers $\mathcal{T}$ for which the collection of translates of X, $(t+E)_{t\in\mathcal{T}}$, is disjoint. Then $m^*(E) = 0$.*

Proof. Since the translate of a measurable set is measurable, by the countable additivity of measure over countable disjoint unions of measurable sets,

$$m^*\left(\bigcup_{t\in\mathcal{T}}(t+X)\right) = \sum_{t\in\mathcal{T}} m^*(t+X).$$

Since both X and $\mathcal{T}$ are bounded, the union is also bounded, and hence has finite measure. Thus the left hand side is finite. However, since measure is translation invariant, $m^*(t+X) = m(X) > 0$ for each $t \in \mathcal{T}$. Thus the set $\mathcal{T}$ is countable and the right hand side is finite, with $m^*(X) = 0$. □

Theorem 104 (Vitali Theorem)**.** Any set $E \subset \mathbb{R}$ with $m^*(E) < \infty$ contains a subset that fails to be measurable.

Any set that fills this requirement is called a *Vitali set*.

13.3 General Measures

The word *measure* allows us to use a single word instead of repeating theorems for length, area, and volume. Measure is simply a generalization of the notion of size. We can now formally define what a measure is in general.

Definition 101 (Measure)**.** Let X be a set and $\mathcal{S}$ a $\sigma-$algebra on X. A *measure* on $(X,\mathcal{S})$ is a function $\mu : \mathcal{S} \to [0,\infty]$ such that $\mu(\emptyset) = 0$ and

$$\mu\left(\bigcup_{k=1}^{\infty} E_k\right) = \sum_{k=1}^{\infty} \mu(E_k),$$

for every disjoint sequence $E_1, E_2, \ldots$ of sets in $\mathcal{S}$.

What are some tangible examples of measures?

- If X is a set, then the *counting measure* is the measure μ defined on the $\sigma-$algebra of X, $\mathcal{P}(X)$, by setting $\mu(E) = n$ if E is a finite set containing exactly n elements and $\mu(E) = \infty$ if E is not a finite set.
- If X is a set and $\mathcal{S}$ is a $\sigma-$algebra on X, and $c \in X$, then the *Dirac measure* δ_c on $(X,\mathcal{S})$ is defined as

$$\delta_c(E) = \begin{cases} 1, & c \in E \\ 0, & c \notin E \end{cases}$$

- Let X be a set and $\mathcal{S}$ a σ−algebra on X. Let $w : X \to [0,\infty]$ be a function. Then we can define a measure μ on $(X,\mathcal{S})$ by

$$\mu(E) = \sum_{x\in E} w(x),$$

for $E \in \mathcal{S}$.

However, if we consider the measurable space $(X,\mathcal{P}(X))$, and define $\mu(E) = m^*(E)$, where $E \subset \mathbb{R}$, then this will not be a measure because it is not finitely additive.

Definition 102 (Measure Space)**.** A *measure space* is an ordered triple $(X,\mathcal{S},\mu)$ where X is a set, $\mathcal{S}$ is a σ−algebra on X, and μ is a measure on $(X,\mathcal{S})$.

We study five important properties of measures:

(1) Measure preserves order

(2) Countable subadditivity

(3) Measure of an increasing union

(4) Measure of a decreasing intersection

(5) Measure of a union

Theorem 105. Let $(X,\mathcal{S},\mu)$ be a measure space and $D,E \in \mathcal{S}$ are such that $D \subset E$. Then:

(a) $\mu(D) \le \mu(E)$,

(b) $\mu(E \setminus D) = \mu(E) - \mu(D)$ provided $\mu(D) < \infty$.

Proof. Because $E = D \cup (E \setminus D)$, and this is a disjoint union, we have

$$\mu(E) = \mu(D) + \mu(E \setminus D) \ge \mu(D),$$

proving (a). If $\mu(D) < \infty$, then subtracting $\mu(D)$ from both sides gives us (b). □

Theorem 106. Let $(X,\mathcal{S},\mu)$ be a measure space and $E_1,E_2,\dots \in \mathcal{S}$. Then

$$\mu\left(\bigcup_{k=1}^{\infty} E_k\right) \le \sum_{k=1}^{\infty} \mu(E_k).$$

Proof. Let $D_1 = \emptyset$ and $D_k = E_1 \cup \cdots \cup E_{k-1}$ for $k \geq 2$. Then

$$E_1 \setminus D_1,\, E_2 \setminus D_2, \ldots$$

is a disjoint sequence of subsets of X whose union equals $\bigcup_{k=1}^{\infty} E_k$. Thus

$$\begin{aligned}\mu\left(\bigcup_{k=1}^{\infty} E_k\right) &= \mu\left(\bigcup_{k=1}^{\infty}(E_k \setminus D_k)\right)\\ &= \sum_{k=1}^{\infty} \mu(E_k \setminus D_k)\\ &\leq \sum_{k=1}^{\infty} \mu(E_k).\end{aligned}$$

□

Theorem 107. Let $(X, \mathcal{S}, \mu)$ be a measure space and $E_1 \subset E_2 \subset \cdots$ is an increasing sequence of sets in $\mathcal{S}$. Then

$$\mu\left(\bigcup_{k=1}^{\infty} E_k\right) = \lim_{k\to\infty} \mu(E_k).$$

Proof. Let $E_0 = \emptyset$. Then

$$\bigcup_{k=1}^{\infty} E_k = \bigcup_{j=1}^{\infty}(E_j \setminus E_{j-1}),$$

where the union on the right is a disjoint union. Thus

$$\begin{aligned}\mu\left(\bigcup_{k=1}^{\infty} E_k\right) &= \sum_{j=1}^{\infty} \mu(E_j \setminus E_{j-1})\\ &= \lim_{k\to\infty} \sum_{j=1}^{k} \mu(E_j \setminus E_{j-1})\\ &= \lim_{k\to\infty} \sum_{j=1}^{k} (\mu(E_j) - \mu(E_{j-1}))\\ &\lim_{k\to\infty} \mu(E_k).\end{aligned}$$

□

Theorem 108. Let $(X, \mathcal{S}, \mu)$ be a measure space and $E_1 \supset E_2 \supset \cdots$ is a decreasing sequence of sets in $\mathcal{S}$, and $\mu(E_1) < \infty$. Then

$$\mu\left(\bigcap_{k=1}^{\infty} E_k\right) = \lim_{k\to\infty} \mu(E_k).$$

Proof. By DeMorgan's Laws, we have

$$E_1 \setminus \bigcap_{k=1}^{\infty} E_k = \bigcup_{k=1}^{\infty} (E_1 \setminus E_k).$$

Then $(E_1 \setminus E_1 \subset E_2 \setminus E_1 \subset \cdots$ is an increasing sequence of sets in $\mathcal{S}$. Take μ on both sides

$$\mu\left(E_1 \setminus \bigcap_{k=1}^{\infty} E_k\right) = \lim_{x\to\infty} \mu(E_1 \setminus E_k).$$

We then use the disjoint union property to get

$$\mu(E_1) - \mu\left(\bigcap_{k=1}^{\infty} E_k\right) = \mu(E_1) - \lim_{k\to\infty} \mu(E_1 \setminus E_k),$$

which gives us our desired result. □

Theorem 109. Let $(X, \mathcal{S}, \mu)$ be a measure space and $D, E \in \mathcal{S}$, with $\mu(D \cap E) < \infty$. Then

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

Proof. We have that

$$D \cup E = (D \setminus (D \cap E)) \cup (E \setminus (D \cap E)) \cup (D \cap E).$$

The right side is a disjoint union. Thus

$$\begin{aligned}\mu(D \cup E) &= \mu(D \setminus (D \cap E)) + \mu(E \setminus (D \cap E)) + \mu(D \cap E)\\ &= (\mu(D) - \mu(D \cap E)) + (\mu(E) - \mu(D \cap E)) + \mu(D \cap E)\\ &= \mu(D) + \mu(E) - \mu(D \cap E),\end{aligned}$$

as desired. □

13.4 Lebesgue Measure

While the outer measure does not have all the properties we desire if we consider all subsets of $\mathbb{R}$, if we loosen this requirement to some other types of sets.

Theorem 110. The following are properties of the outer measure restricted to certain subsets.

(a) Suppose A and G are disjoint subsets of $\mathbb{R}$, and G is open. Then

$$m^*(A \cup G) = m^*(A) + m^*(G).$$

(b) Suppose A and F are disjoint subsets of $\mathbb{R}$, and F is closed. Then

$$m^*(A \cup F) = m^*(A) + m^*(F).$$

(c) Suppose A and B are disjoint subsets of $\mathbb{R}$, and B is a Borel set. Then

$$m^*(A \cup B) = m^*(A) + m^*(B).$$

Proof. (a) □

As a direct corollary, we can show that not every subset of $\mathbb{R}$ is a Borel set.

Theorem 111. There exists a set $B \subset \mathbb{R}$ such that $m^*(B) < \infty$ and B is not a Borel set.

Proof. We know that there exists disjoint sets $A, B \subset \mathbb{R}$ such that $m^*(A \cup B) \neq m^*(A) + m^*(B)$. For any such sets, we must have $m^*(B) < \infty$ because otherwise both $m^*(A \cup B)$ and $m^*(A) + m^*(B)$ equal ∞. Then (c) from above tells us that B is not a Borel set. □

Borel sets can be approximated by closed sets. This is useful in applications, as Borel sets are often quite abstract.

Theorem 112. Suppose $B \subset \mathbb{R}$ is a Borel set. Then for every $\epsilon > 0$, there exists a closed set $F \subset B$ such that $m^*(B \setminus F) < \epsilon$.

So is the outer measure actually a measure when restricted to the Borel sets? That is, is $(\mathbb{R}, \mathcal{B}, m^*)$ a measure space, where $\mathcal{B}$ is the σ−algebra of Borel subsets of $\mathbb{R}$?

Proof. Let $B_1, B_2, \ldots$ be a disjoint sequence of Borel subsets of $\mathbb{R}$. Then for $n \in \mathbb{Z}^+$, we have

$$m^*\left(\bigcup_{k=1}^{\infty} B_k\right) \geq m^*\left(\bigcup_{k=1}^{n} B_k\right) = \sum_{k=1}^{n} m^*(B_k),$$

Taking the limit as $n \to \infty$, we have $m^*(\bigcup_{k=1}^{\infty} B_k) \geq \sum_{k=1}^{\infty} m^*(B_k)$. The inequality in the other direction follows from countable subadditivity of outer measure. Hence

$$m^*\left(\bigcup_{k=1}^{\infty} B_k\right) = \sum_{k=1}^{\infty} m^*(B_k).$$

Thus the outer measure is a measure on the $\sigma-$algebra of Borel subsets of $\mathbb{R}$. □

Definition 103 (Lebesgue Measure)**.** The *Lebesgue measure* is the measure on $(\mathbb{R}, \mathcal{B})$, where $\mathcal{B}$ is the $\sigma-$algebra of Borel subsets of $\mathbb{R}$, that assigns to each Borel set its outer measure.

We denote this as, for $E \subset \mathcal{B}$,

$$m(E) = m^*(E).$$

Definition 104 (Lebesgue Measurable Set)**.** A set $A \subset \mathbb{R}$ is called *Lebesgue measurable* if there exists a Borel set $B \subset A$ such that $m(A \setminus B) = 0$.

We denote the set of Lebesgue measurable subsets of $\mathbb{R}$ as $\mathcal{L}$. Every Borel set is Lebesgue measurable because if $A \subset \mathbb{R}$ is a Borel set, then we can take $B = A$ in the definition above. We have several equivalent conditions for being Lebesgue measurable.

Theorem 113. Let $A \subset \mathbb{R}$. Then the following are equivalent:

(a) A is Lebesgue measurable.

(b) For each $\epsilon > 0$, there exists a closed set $F \subset A$ with $m(A \setminus F) < \epsilon$.

(c) There exist closed sets $F_1, F_2, \ldots$ contained in A such that

$$m\left(A \setminus \bigcup_{k=1}^{\infty} F_k\right) = 0.$$

(d) There exists a Borel set $B \subset A$ such that $m(A \setminus B) = 0$.

(e) For each $\epsilon > 0$, there exists an open set $G \supset A$ such that $m(G \setminus A) < \epsilon$.

(f) There exists open sets $G_1, G_2, \ldots$ containing A such that

$$m\left(\bigcap_{k=1}^{\infty} G_k \setminus A\right) = 0.$$

(g) There exists a Borel set $B \supset A$ such that $m(B \setminus A) = 0$.

Now we want to answer the following questions:

- Is $\mathcal{L}$ a σ−algebra on $\mathbb{R}$?
- Is the outer measure a measure on $(\mathbb{R}, \mathcal{L})$? In other words, is $(\mathbb{R}, \mathcal{L}, m^*)$ a measure space?

As it turns out, these are both true statements.

Proof.

- Because Lebesgue measurable implies there exists a closed set $F \subset A$ with $m(A \setminus F) < \epsilon$, $\mathcal{L}$ is the collection of sets satisfying the second condition. This set is a σ−algebra on $\mathbb{R}$, proving the first statement.
- Let $A_1, A_2, \ldots$ be a disjoint sequence f Lebesgue measurable sets. By definition, for each $k \in \mathbb{Z}^+$, there exists a Borel set $B_k \subset A_k$ such that $m(A_k \setminus B_k) = 0$. Now

$$\begin{aligned} m\left(\bigcup_{k=1}^{\infty} A_k\right) &\geq m\left(\bigcup_{k=1}^{\infty} B_k\right) \\ &= \sum_{k=1}^{\infty} m(B_k) \\ &= \sum_{k=1}^{\infty} m(A_k). \end{aligned}$$

This combined with the countable subadditivity of outer measures, this gives us that $m(\bigcup_{k=1}^{\infty} A_k) = \sum_{k=1}^{\infty} m(A_k)$, which shows that m is indeed a measure. □

Sometimes we use a slightly modified definition for what the Lebesgue measure really is.

Definition 105 (Lebesgue Measure v2)**.** The *Lebesgue measure* is the measure on $(\mathcal{L}, \mathbb{R})$, where $\mathcal{L}$ is the $\sigma-$algebra of Lebesgue measurable subsets of $\mathbb{R}$, that assigns to each Lebesgue measurable set its outer measure.

Theorem 114 (Borel-Cantelli Lemma)**.** Let $E_1, E_2, \ldots$ be a countable collection of measurable sets for which $\sum_{k=1}^{\infty} m(E_k) < \infty$. Then almost all $x \in \mathbb{R}$ belong to at most finitely many of the E_k's.

Proof. For each n, by countable subadditivity of m,

$$m\left(\bigcup_{k=n}^{\infty} E_k\right) \leq \sum_{k=n}^{\infty} m(E_k) < \infty.$$

Hence, by the continuity of measure,

$$m\left(\bigcap_{n=1}^{\infty}\left[\bigcup_{k=n}^{\infty} E_k\right]\right) = \lim_{n\to\infty} m\left(\bigcup_{k=n}^{\infty} E_k\right) \leq \lim_{n\to\infty} \sum_{k=n}^{\infty} m(E_k) = 0.$$

Therefore almost all $x \in \mathbb{R}$ fail to belong to $\bigcap_{n=1}^{\infty}\left[\bigcup_{k=n}^{\infty} E_k\right]$ and therefore belong to at most finitely many E_k's. □

13.5 Convergence of Measurable Functions

Theorem 115 (Littlewood's Three Principles)**.** Roughly speaking, the following principles hold:

- Every measurable set is nearly a finite union of intervals.
- Every measurable function is nearly continuous.
- Every pointwise convergent sequence of measurable functions is nearly uniformly convergent.

For continuous functions, if $f_n : E \to \mathbb{R}$ and $n = 1, 2, \ldots$ is a sequence of continuous functions, and if $f_n \to f$ pointwise on E, we know that f will be continuous. Does this hold true if instead of continuous, f_n were measurable instead? That is, if $f_n : E \to \mathbb{R}$ are measurable, then is f measurable?

Theorem 116. Let $f_n : E \to \mathbb{R}$, and $n = 1, 2, \ldots$ be a sequence of measurable functions on E that converges pointwise on E to the function f. Then f is measurable.

Proof. Let E_0 be a subset of E for which $m(E_0) = 0$ and $f_n \to f$ pointwise on $E \setminus E_0$. Since $m(E_0) = 0$, we know f is measurable if and only if its restriction to $E \setminus E_0$ is measurable. Therefore by replacing E by $E \setminus E_0$, we may assume the sequence converges pointwise on all of E. Fix any number c. We must show that $\{x \in E \mid f(x) < c\}$ is measurable. Observe that for $x \in E$, since $\lim_{n\to\infty} f_n(x) = f(x)$, that

$$f(x) < c \Longleftrightarrow \exists n, k \in \mathbb{Z}^+ \text{ for which } f_j(x) < c - \frac{1}{n}, \quad \forall j \geq k.$$

But for any natural numbers n and j, since f_j is measurable, the set $\{x \in E \mid f_j(x) < c - \frac{1}{n}\}$ is measurable. Therefore for any k, the intersection of the countable collection of measurable sets

$$\bigcap_{j=k}^{\infty} \left\{x \in E \mid f_j(x) < c - \frac{1}{n}\right\}$$

is also measurable. Consequently, since the union of a countable collection of measurable sets is measurable,

$$\{x \in E \mid f(x) < c\} = \bigcup_{1 \leq k, n < \infty} \left[\bigcap_{j=k}^{\infty} \left\{f_j(x) < c - \frac{1}{n}\right\}\right]$$

is measurable as well, completing our proof. □

Recall that pointwise convergence does not necessarily imply uniform convergence. However, a pointwise convergent sequence of functions on a measure space with finite total measure almost converges uniformly, in the sense that it converges uniformly except on a set that can have arbitrarily small measure.

Theorem 117 (Egorov's Theorem)**.** Let $(X, \mathcal{S}, \mu)$ be a measure space such that $\mu(X) < \infty$. Let $f_1, f_2, \ldots$ be a sequence of $\mathcal{S}$−measurable functions from X to $\mathbb{R}$ that converges pointwise on X to a function $f : X \to \mathbb{R}$. Then for every $\epsilon > 0$, there exists a set $E \in \mathcal{S}$ such that $\mu(X \setminus E) < \epsilon$ and $f_1, f_2, \ldots$ converges uniformly to f on E.

Proof. Suppose $\epsilon > 0$. Fix some $n \in \mathbb{Z}^+$. By definition of pointwise convergence,

$$\bigcup_{m=1}^{\infty} \bigcap_{k=m}^{\infty} \left\{x \in X \mid |f_k(x) - f(x)| < \frac{1}{n}\right\} = X.$$

Then for some $m \in \mathbb{Z}^+$, let

$$A_{m,n} = \bigcap_{k=m}^{\infty} \left\{ x \in X \mid |f_k(x) - f(x)| < \frac{1}{n} \right\}.$$

By construction, $A_{1,n} \subset A_{2,n} \subset \cdots$, which is an increasing sequence of sets, and so we can rewrite

$$\bigcup_{m=1}^{\infty} A_{m,n} = X.$$

This implies that $\lim_{m\to\infty} \mu(A_{m,n}) = \mu(X)$. Thus there exists $m_n \in \mathbb{Z}^+$ such that

$$\mu(X) - \mu(A_{m,n}) < \frac{\epsilon}{2^n}.$$

If we let $E = \bigcap_{n=1}^{\infty} A_{m_n,n}$, then

$$\begin{aligned} \mu(X \setminus E) &= \mu\left(X \setminus \bigcap_{n=1}^{\infty} A_{m_n,n} \right) \\ &= \mu\left(\bigcup_{n=1}^{\infty} (X \setminus A_{m_n,n}) \right) \\ &\le \sum_{n=1}^{\infty} \mu(X \setminus A_{m_n,n}) \\ &< \epsilon. \end{aligned}$$

Now we must show that $f_1, f_2, \ldots$ does converge uniformly to f on E. Let $\epsilon > 0$ again. Let $n \in \mathbb{Z}^+$ be such that $\frac{1}{n} < \epsilon$. Then $E \subset A_{m_n,n}$, and so

$$|f_k(x) - f(x)| < \frac{1}{n} < \epsilon,$$

for all $k \ge m_n$ and $x \in E$. Thus $f_n \to f$ uniformly on E. □

Definition 106 (Simple Function)**.** We call a function *simple* if it takes on only finitely many values.

If $(X, \mathcal{S})$ is a measurable space and $f : X \to \mathbb{R}$ is simple, and $c_1, \ldots, c_n$ are distinct nonzero values of f, then

$$f = c_1 \chi_{E_1} + \cdots + c_n \chi_{E_n},$$

where $E_k = f^{-1}(\{c_k\})$. This tells us that f is $\mathcal{S}$–measurable iff $E_1, \ldots, E_n \in \mathcal{S}$.

Theorem 118 (Simple Approximation Theorem)**.** Let $(X, \mathcal{S})$ be a measurable space and $f : X \to [-\infty, \infty]$ be $\mathcal{S}$–measurable. Then there exists $f_n : X \to \mathbb{R}$ for $n = 1, 2, \ldots$ such that

(a) Each f_k is a simple $\mathcal{S}$–measurable function.

(b) $|f_k(x)| \le |f_{k+1}(x)| \le |f(x)|$, for all $k \in \mathbb{Z}^+$ and all $x \in X$.

(c) $\lim_{k\to\infty} f_k(x) = f(x)$ for every $x \in X$.

(d) If f is bounded, then $f_n \to f$ uniformly on f.

Theorem 119 (Luzin's Theorem)**.** Let $g : \mathbb{R} \to \mathbb{R}$ be a Borel measurable function. Then for every $\epsilon > 0$, there exists a closed set $F \subset \mathbb{R}$ such that $m(\mathbb{R} \setminus F) < \epsilon$ and $g|_F$ is a continuous function on F.

Definition 107 (Lebesgue Measurable Function)**.** A function $f : A \to \mathbb{R}$ with $A \subset \mathbb{R}$ is called *Lebesgue measurable* if $f^{-1}(B)$ is a Lebesgue measurable set for every Borel set $B \subset \mathbb{R}$.

At first glance, the concepts of Lebesgue measurable sets and Borel sets seem very similar. Rest assured, there do exist Lebesgue measurable sets that are not Borel sets, but that is beyond the scope of this course. Likewise, there also exists Lebesgue measurable functions that are not Borel measurable, but again, these are unlikely to arise in our study of measure theory, and are not really *that* useful.

Theorem 120. Let $f : \mathbb{R} \to \mathbb{R}$ be a Lebesgue measurable function. Then there exists a Borel measurable function $g : \mathbb{R} \to \mathbb{R}$ such that

$$m(\{x \in \mathbb{R} \mid g(x) \neq f(x)\}) = 0.$$

Proof. We know that there exists $f_n : \mathbb{R} \to \mathbb{R}$ for $n = 1, 2, \ldots$ such that $f_n \to f$ pointwisely on $\mathbb{R}$, and f_n are Lebesgue measurable simple functions. Let $k \in \mathbb{Z}^+$. Then there exists $c_1, \ldots, c_n \in \mathbb{R}$ and disjoint Lebesgue measurable sets $A_1, \ldots, A_n \subset \mathbb{R}$ such that

$$f_k = c_1 \chi_{A_1} + \cdots + c_n \chi_{A_n}.$$

Then for each $j \in [1, n]$, there exists a Borel set $B_j \subset A_j$ such that $m(A_j \setminus B_j) = 0$. Now let

$$g_k = c_1 \chi_{B_1} + \cdots + c_n \chi_{B_n}.$$

Then g_k is a Borel measurable function and $m(\{x \in \mathbb{R} \mid g_k(x) \neq f_k(x)\}) = 0$. If $x \notin \bigcup_{k=1}^{\infty}\{x \in \mathbb{R} \mid g_k(x) \neq f_k(x)\}$, then $g_k(x) = f_k(x)$ for all $k \in \mathbb{Z}^+$ and thus $\lim_{k\to\infty} g_k(x) = f(x)$. Let

$$E = \left\{x \in \mathbb{R} \mid \lim_{k\to\infty} g_k(x) \text{ exists in } \mathbb{R}\right\}.$$

E is a Borel subset of $\mathbb{R}$. Also,

$$\mathbb{R} \setminus E \subset \bigcup_{k=1}^{\infty}\{x \in \mathbb{R} \mid g_k(x) \neq f_k(x)\},$$

and so $m(\mathbb{R} \setminus E) = 0$. For $x \in \mathbb{R}$, let

$$g(x) = \lim_{k\to\infty} (\chi_E g_k)(x).$$

If $x \in E$, then the limit exists; if $x \in \mathbb{R} \setminus E$, then the limit also exists because $(\chi_E g_k)(x) = 0$ for all $k \in \mathbb{Z}^+$. For each k, the function $\chi_E g_k$ is Borel measurable. Hence g is Borel measurable, and since

$$\{x \in \mathbb{R} \mid g(x) \neq f(x)\} \subset \bigcup_{k=1}^{\infty}\{x \in \mathbb{R} \mid g_k(x) \neq f_k(x)\},$$

we know that

$$m(\{x \in \mathbb{R} \mid g(x) \neq f(x)\}) = 0.$$

□

14 Lebesgue Integration

We have shown before that the Riemann integral, while easy to understand, cannot handle many functions of interest. We can use our robust theory of measures to develop a theory of integration that fixes many of the problems with Riemann integration.

14.1 Integration with Respect to a Measure

In this section, we only consider nonnegative functions.

Definition 108 ($\mathcal{S}$−partition)**.** Let $\mathcal{S}$ be a σ−algebra on a set X. An *$\mathcal{S}$−partition* of X is a finite collection $A_1, \ldots, A_m$ of disjoint sets in $\mathcal{S}$ such that $A_1 \cup \cdots \cup A_m = X$.

As an analogue to Riemann integration, using our partition we need to be able to define lower/upper sums. Since we are working with an arbitrary measure, if X is a closed interval $[a,b]$ in $\mathbb{R}$ and μ is the Lebesgue measure on the Borel subsets of $[a,b]$, the sets $A_1, \ldots, A_m$ do not need to be subintervals of $[a,b]$ like in Riemann integration; they just need to be Borel sets.

Definition 109 (Lower/Upper Lebesgue Sum)**.** Let $(X, \mathcal{S}, \mu)$ be a measure space. Let $f : X \to [0,\infty]$ be an $\mathcal{S}$−measurable function, and P an $\mathcal{S}$−partition $A_1, \ldots, A_m$ of X. The *lower Lebesgue sum* $\mathcal{L}(f,P)$ is defined by

$$\mathcal{L}(f,P) = \sum_{j=1}^{m} \mu(A_j) \inf_{A_j} f.$$

Similarly, the *upper Lebesgue sum* $\mathcal{U}(f,P)$ is defined by

$$\mathcal{U}(f,P) = \sum_{j=1}^{m} \mu(A_j) \sup_{A_j} f.$$

We denote the integral of an $\mathcal{S}$−measurable function f with respect to μ by

$$\int f\,d\mu.$$

Definition 110 (Lebesgue Integral)**.** Let $(X,\mathcal{S},\mu)$ be a measure space, and $f : X \to [0,\infty]$ an $\mathcal{S}$−measurable function. The *Lebesgue integral* of f with respect to μ, denoted $\int f\,d\mu$, is defined by

$$\int f\,d\mu = \sup\{\mathcal{L}(f,P) \mid P \text{ is an } \mathcal{S}\text{−partition of } X\}.$$

We expect that each $\mathcal{S}$−partition $A_1,\ldots,A_m$ of X will lead to an approximation of f from below by the $\mathcal{S}$−measurable simple function $\sum_{j=1}^m (\inf_{A_j} f)\chi_{A_j}$. Thus

$$\sum_{j=1}^m \mu(A_j)\inf_{A_j} f \approx \int f\,d\mu.$$

The first property of our integral should be fairly simple: the Lebesgue integral of a characteristic function should just be the measure of a set.

Theorem 121. Let $(X,\mathcal{S},\mu)$ be a measure space, and let $E \in \mathcal{S}$. Then

$$\int \chi_E\,d\mu = \mu(E).$$

Proof. Let P be the $\mathcal{S}$−partition of X consisting of E and its complement $X \setminus E$. Clearly, $\mathcal{L}(\chi_E,P) = \mu(E)$. Thus

$$\int \chi_E\,d\mu \geq \mu(E).$$

To prove the other direction, let P be a $\mathcal{S}$−partition $A_1,\ldots,A_m$ of X. Then $\mu(A_j)\inf_{A_j}\chi_E$ equals $\mu(A_j)$ if $A_j \subset E$ and is 0 otherwise. Thus

$$\begin{aligned}
\mathcal{L}(\chi_E,P) &= \sum_{\{j \mid A_j \subset E\}} \mu(A_j) \\
&= \mu\left(\bigcup_{\{j \mid A_j \subset E\}} A_j\right) \\
&\leq \mu(E),
\end{aligned}$$

as desired. This completes our proof. □

Let us return to our old friend, the Dirichlet function. Notice that we can rewrite $f_D(x) = \chi_{\mathbb{Q}}$. Consider the Lebesgue measure on $\mathbb{R}$. We have that

$$\int \chi_{\mathbb{Q}}\, dm = 0,$$

because $m(\mathbb{Q}) = 0$. Hence from the very start, we can already integrate this function, when with the Riemann integral, we immediately encountered some problems. We had to generalize the Riemann integral through a complicated process just to be able to integrate $f_D(x)$. Hence Lebesgue integration is much more powerful than Riemann integration is. As a direct consequence, because $m([0,1] \setminus \mathbb{Q}) = 1$, we can say that

$$\int \chi_{[0,1]\setminus\mathbb{Q}}\, dm = 1,$$

but the lower Riemann integral tells us that that would be 0, which is obviously incorrect.

Theorem 122. Let $(X, \mathcal{S}, \mu)$ be a measure space, and $E_1, \ldots, E_n$ be disjoint sets in $\mathcal{S}$. Also let $c_1, \ldots, c_n \in [0, \infty]$. Then

$$\int \left(\sum_{k=1}^{n} c_k \chi_{E_k} \right) d\mu = \sum_{k=1}^{n} c_k \mu(E_k).$$

Proof. Without loss of generality, assume $E_1, \ldots, E_n$ is an $\mathcal{S}$−partition of X by replacing $n+1$ and setting $E_{n+1} = X \setminus (E_1 \cup \cdots \cup E_n)$ and $c_{n+1} = 0$. Let P be this $\mathcal{S}$−partition, then $\mathcal{L}(\sum_{k=1}^{n} c_k \chi_{E_k}, P) = \sum_{k=1}^{n} c_k \mu(E_k)$. Thus

$$\int \left(\sum_{k=1}^{n} c_k \chi_{E_k} \right) d\mu \geq \sum_{k=1}^{n} c_k \mu(E_k).$$

To prove the other direction, let P be an $\mathcal{S}$–partition $A_1,\ldots,A_m$ of X. Then

$$\begin{aligned}
\mathcal{L}\left(\sum_{k=1}^{n} c_k \chi_{E_k}, P\right) &= \sum_{j=1}^{m} \mu(A_j) \min_{\{i \mid A_j \cap E_i \neq \emptyset\}} c_i \\
&= \sum_{j=1}^{m} \sum_{k=1}^{n} \mu(A_j \cap E_k) \min_{\{i \mid A_j \cap E_i \neq \emptyset\}} c_i \\
&\leq \sum_{j=1}^{m} \sum_{k=1}^{n} \mu(A_j \cap E_k) c_k \\
&= \sum_{k=1}^{n} c_k \sum_{j=1}^{m} \mu(A_j \cap E_k) \\
&= \sum_{k=1}^{n} c_k \mu(E_k).
\end{aligned}$$

□

Theorem 123. Let $(X,\mathcal{S},\mu)$ be a measure space and $f,g : X \to [0,\infty]$ be $\mathcal{S}$–measurable functions such that $f(x) \leq g(x)$ for all $x \in X$. Then

$$\int f \, d\mu \leq \int g \, d\mu.$$

Proof. Let P be an $\mathcal{S}$–partition $A_1,\ldots,A_m$ of X. Then

$$\inf_{A_j} f \leq \inf_{A_j} g,$$

for each $j \in [1,m]$. Thus $\mathcal{L}(f,P) \leq \mathcal{L}(g,P)$, and so $\int f \, d\mu \leq \int g \, d\mu$. □

Theorem 124. Let $(X,\mathcal{S},\mu)$ be a measure space and $f : X \to [0,\infty]$ be $\mathcal{S}$–measurable. Then

$$\begin{aligned}
\int f \, d\mu = \sup\Bigg\{ \sum_{j=1}^{m} c_j \mu(A_j) \,\Big|\, & A_1,\ldots,A_m \text{ are disjoint sets in } \mathcal{S}, \\
& c_1,\ldots,c_m \in [0,\infty), \text{ and} \\
& f(x) \geq \sum_{j=1}^{m} c_j \chi_{A_j}(x), \quad \forall x \in X \Bigg\}.
\end{aligned}$$

We use this theorem to prove a central theorem which allows us to interchange limits and integrals in certain circumstances.

Theorem 125 (Monotone Convergence Theorem)**.** Let $(X, \mathcal{S}, \mu)$ be a measure space and $0 \le f_1 \le f_2 \le \cdots$ be an increasing sequence of measurable functions. Define $f : X \to [0, \infty]$ by

$$f(x) = \lim_{k\to\infty} f_k(x).$$

Then

$$\lim_{k\to\infty} \int f_k \, d\mu = \int f \, d\mu.$$

Proof. We know f is measurable. Because $f_k(x) \le f(x)$ for all $x \in X$, $\int f_k \, d\mu \le \int f \, d\mu$ for each $k \in \mathbb{Z}^+$. Thus

$$\lim_{k\to\infty} \int f_k \, d\mu \le \int f \, d\mu.$$

To prove the other direction, let $A_1, \ldots, A_m$ be disjoint sets in $\mathcal{S}$ and $c_1, \ldots, c_m \in [0, \infty)$ are such that

$$f(x) \ge \sum_{j=1}^{m} c_j \chi_{A_j}(x), \quad \forall x \in X.$$

Let $t \in (0, 1)$. Then for $k \in \mathbb{Z}^+$, let

$$E_k = \left\{ x \in X \mid f_k(x) \ge t \sum_{j=1}^{m} c_j \chi_{A_j}(x) \right\}.$$

Then $E_1 \subset E_2 \subset \cdots$ is an increasing sequence of sets in $\mathcal{S}$ whose union equals X. It follows that

$$\lim_{k\to\infty} \mu(A_j \cap E_k) = \mu(A_j),$$

for each $j \in [1, m]$, and

$$f_k(x) \ge \sum_{j=1}^{m} t c_j \chi_{A_j \cap E_k}(x), \quad \forall x \in X.$$

Thus

$$\int f_k \, d\mu \ge t \sum_{j=1}^{m} c_j \mu(A_j \cap E_k).$$

We take the limit as $k \to \infty$ of both sides to get

$$\lim_{k\to\infty} \int f_k \, d\mu \ge t \sum_{j=1}^{m} c_j \mu(A_j).$$

Finally we take $t \to 1$, and then we take the supremum of the resulting inequality over all $\mathcal{S}$–partitions and $c_1, \ldots, c_m \in [0, \infty)$, completing our proof. □

We conclude this section with a consequence of the Monotone Convergence Theorem.

Theorem 126 (Fatou's Lemma)**.** Let $(X, \mathcal{S}, \mu)$ is a measure space, and $f_1, f_2, \ldots$ is a sequence of nonnegative $\mathcal{S}$–measurable functions on X. Define a function $f : X \to [0, \infty]$ by $f(x) = \liminf_{k\to\infty} f_k(x)$. Then

$$\int f\, d\mu \leq \liminf_{k\to\infty} \int f_k\, d\mu.$$

14.2 Properties of Lebesgue Integral

In the previous section, we only considered nonnegative functions. How do we integrate real-valued functions?

Definition 111 (f–plus; f–minus)**.** Let $f : X \to [-\infty, \infty]$ be a function. We define f^+ and f^- from $X \to [0, \infty]$ by

$$f^+(x) = \begin{cases} f(x), & f(x) \geq 0, \\ 0, & f(x) < 0 \end{cases}, \quad f^-(x) = \begin{cases} 0, & f(x) \geq 0, \\ -f(x), & f(x) < 0 \end{cases}$$

Notice that using this notation,

$$f = f^+ - f^-, \quad \text{and} \quad |f| = f^+ + f^-.$$

Definition 112 (Lebesgue Integral of Real-Valued Function)**.** Let $(X, \mathcal{S}, \mu)$ be a measure space, and $f : X \to [-\infty, \infty]$ an $\mathcal{S}$–measurable function such that at least one of $\int f^+\, d\mu$ and $\int f^-\, d\mu$ is finite. The *Lebesgue integral* of f with respect to μ, denoted $\int f\, d\mu$, is defined by

$$\int f\, d\mu = \int f^+\, d\mu - \int f^-\, d\mu.$$

Of course, not every function is Lebesgue integrable. If we consider the Lebesgue measure on $\mathbb{R}$ and $f : \mathbb{R} \to \mathbb{R}$ defined by

$$f(x) = \begin{cases} 1, & x \geq 0 \\ -1, & x < 0 \end{cases},$$

then $\int f\, dm$ is not defined because $\int f^+\, d\mu = \infty$, and $\int f^-\, d\mu = \infty$ as well. The Lebesgue integral should have the same properties as Riemann integration does; otherwise it would be useless. And indeed, it does.

Theorem 127 (Properties of Lebesgue Integral)**.** Let $(X, \mathcal{S}, \mu)$ be a measure space.

(1) (*Homogeneity*): If $f : X \to [-\infty, \infty]$ is a function such that $\int f\, d\mu$ is defined, and if $c \in \mathbb{R}$, then

$$\int cf\, d\mu = c \int f\, d\mu.$$

(2) (*Additivity*): If $f, g : X \to \mathbb{R}$ are $\mathcal{S}$–measurable functions such that $\int |f|\, d\mu < \infty$ and $\int |g|\, d\mu < \infty$, then

$$\int (f+g)\, d\mu = \int f\, d\mu + \int g\, d\mu.$$

(3) (*Order-Preservation*): If $f, g : X \to \mathbb{R}$ are $\mathcal{S}$–measurable functions such that $\int f\, d\mu$ and $\int g\, d\mu$ are defined, and $f(x) \le g(x)$ for all $x \in X$, then

$$\int f\, d\mu \le \int g\, d\mu.$$

(4) If $f : X \to [-\infty, \infty]$ is a function such that $\int f\, d\mu$ is defined, then

$$\left| \int f\, d\mu \right| \le \int |f|\, d\mu.$$

Proof. (1) Suppose f is nonnegative and $c \ge 0$. If P is an $\mathcal{S}$–partition of X, then $\mathcal{L}(cf, P) = c\mathcal{L}(f, P)$. Thus $\int cf\, d\mu = c \int f\, d\mu$. If f takes on values in $[-\infty, \infty]$, then suppose $c \ge 0$. Then

$$\begin{aligned}
\int cf\, d\mu &= \int (cf)^+\, d\mu - \int (cf)^-\, d\mu \\
&= \int cf^+\, d\mu - \int cf^-\, d\mu \\
&= c\left(\int f^+\, d\mu - \int f^-\, d\mu \right) \\
&= c \int f\, d\mu.
\end{aligned}$$

If $c < 0$, then $-c > 0$, and so

$$\begin{aligned}\int cf\,d\mu &= \int (cf)^+\,d\mu - \int (cf)^-\,d\mu \\ &= \int (-c)f^-\,d\mu - \int (-c)f^+\,d\mu \\ &= (-c)\left(\int f^-\,d\mu - \int f^+\,d\mu\right) \\ &= c\int f\,d\mu.\end{aligned}$$

(2) Since

$$(f+g)^+ - (f+g)^- = f+g = f^+ - f^- + g^+ - g^-,$$

thus

$$(f+g)^+ + f^- + g^- = (f+g)^- + f^+ + g^+.$$

Since both sides of the equation are sums of nonnegative functions, we integrate both sides with respect to μ and get

$$\int (f+g)^+\,d\mu + \int f^-\,d\mu + \int g^-\,d\mu = \int (f+g)^-\,d\mu + \int f^+\,d\mu + \int f^+\,d\mu + \int g^+\,d\mu.$$

We rearrange and get

$$\int (f+g)^+\,d\mu - \int (f+g)^-\,d\mu = \int f^+\,d\mu - \int f^-\,d\mu + \int g^+\,d\mu - \int g^-\,d\mu,$$

where the left side is not of the form $\infty - \infty$ because $(f+g)^+ \leq f^+ + g^+$, and $(f+g)^- \leq f^- + g^-$. Hence we rewrite our equation as

$$\int (f+g)\,d\mu = \int f\,d\mu + \int g\,d\mu.$$

(3) Suppose $\int |f|\,d\mu < \infty$ and $\int |g|\,d\mu < \infty$. Additivity and homogeneity with $c = -1$ imply that

$$\int g\,d\mu - \int f\,d\mu = \int (g-f)\,d\mu.$$

The last integral is nonnegative because $g(x) - f(x) \geq 0$ for all $x \in X$.

(4) Because $\int f\,d\mu$ is defined, f is an $\mathcal{S}$–measurable function and at least one of $\int f^+\,d\mu$ and $\int f^-\,d\mu$ is finite. Then

$$\begin{aligned}\left|\int f\,d\mu\right| &\leq \left|\int f^+\,d\mu - \int f^-\,d\mu\right| \\ &\leq \int f^+\,d\mu + \int f^-\,d\mu \\ &= \int (f^+ + f^-)\,d\mu \\ &= \int |f|\,d\mu.\end{aligned}$$

□

14.3 Limit Integral Theorems

Definition 113 (Lebesgue Integral over Subset)**.** Let $(X,\mathcal{S},\mu)$ be a measure space and $E \in \mathcal{S}$. If $f : X \to [-\infty,\infty]$ is an $\mathcal{S}$–measurable function, then $\int_E f\,d\mu$ is defined by

$$\int_E f\,d\mu = \int \chi_E f\,d\mu,$$

if the right side exists. Otherwise, $\int_E f\,d\mu$ is undefined.

Alternatively, we can think of this as

$$\int f|_E\,d\mu_E,$$

where μ_E is the measure restricting μ to the elements of $\mathcal{S}$ that are contained in E. We can bound integrals over subsets.

Theorem 128. Let $(X,\mathcal{S},\mu)$ be a measure space, and $E \in \mathcal{S}$. Let $f : X \to [-\infty,\infty]$ be a function such that $\int_E f\,d\mu$ is defined. Then

$$\left|\int_E f\,d\mu\right| \leq \mu(E)\sup_E |f|.$$

Proof. Let $c = \sup_E |f|$. Then

$$\begin{aligned}\left|\int_E f\,d\mu\right| &= \left|\int \chi_E f\,d\mu\right| \\ &\le \int \chi_E |f|\,d\mu \\ &\le \int c\chi_E\,d\mu \\ &= c\mu(E).\end{aligned}$$

□

Theorem 129 (Bounded Convergence Theorem)**.** Let $(X, \mathcal{S}, \mu)$ be a measure space with $\mu(X) < \infty$. Suppose $f_n : X \to \mathbb{R}$ is a sequence of $\mathcal{S}$–measurable functions for $n = 1, 2, \ldots$ that converges pointwise on X to some $f : X \to \mathbb{R}$. If there exists $c \in (0, \infty)$ such that for all $k \in \mathbb{N}$ and $x \in X$,

$$|f_k(x)| \le c,$$

then

$$\lim_{k\to\infty} f_k\,d\mu = \int f\,d\mu.$$

Proof. We know f is measurable. Suppose c satisfies the hypothesis of this theorem. Take $\epsilon > 0$. Then by Egorov's Theorem, there exists $E \in \mathcal{S}$ such that $\mu(X \setminus E) < \frac{\epsilon}{4c}$, and $f_n : X \to \mathbb{R}$ converges uniformly to f on E. Now

$$\begin{aligned}\left|\int f_k\,d\mu - \int f\,d\mu\right| &= \left|\int_{X\setminus E} f_k\,d\mu - \int_{X\setminus E} f\,d\mu + \int_E (f_k - f)\,d\mu\right| \\ &\le \int_{X\setminus E} |f_k|\,d\mu + \int_{X\setminus E} |f|\,d\mu + \int_E |f_k - f|\,d\mu \\ &< \frac{\epsilon}{2} + \mu(E)\sup_E |f_k - f|.\end{aligned}$$

Since $f_n \to f$ uniformly on E, and $\mu(E) < \infty$, the right side of the inequality above is less than ϵ for k sufficiently large. □

Consider $f, g : X \to [-\infty, \infty]$, measurable functions on our measure space $(X, \mathcal{S}, \mu)$. If

$$\mu(\{x \in X \mid f(x) \ne g(x)\}) = 0,$$

then $\int f\,d\mu = \int g\,d\mu$ by definition of Lebesgue integral. This tells us that whatever happens on a set of measure zero, it does not matter.

Definition 114 (Almost Every)**.** Let $(X,\mathcal{S},\mu)$ be a measure space. A set $E \in \mathcal{S}$ is said to contain *almost every* element of X if $\mu(X \setminus E) = 0$.

As an example, almost every real number is irrational, because $m(\mathbb{Q}) = 0$. Many theorems about integrals can be relaxed so that the hypotheses apply almost everywhere as opposed to everywhere.

Theorem 130. Let $(X,\mathcal{S},\mu)$ be a measure space. Let $g : X \to [0,\infty]$ be $\mathcal{S}$–measurable, and $\int g\,d\mu < \infty$. Then for every $\epsilon > 0$, there exists $\delta > 0$ such that

$$\int_B g\,d\mu < \epsilon$$

for every set $B \in \mathcal{S}$ such that $\mu(B) < \epsilon$.

This theorem tells us that if a nonnegative function has a finite integral, then its integral over all small sets (in the sense of measure) is small.

Proof. Let $\epsilon > 0$. Let $h : X \to [0,\infty)$ be a simple $\mathcal{S}$–measurable function such that $0 \le h \le g$ and

$$\int g\,d\mu - \int h\,d\mu < \frac{\epsilon}{2}.$$

Define

$$H = \max\{h(x) \mid x \in X\},$$

and pick $\delta > 0$ such that $H\delta < \frac{\epsilon}{2}$. Suppose $B \in \mathcal{S}$ and $\mu(B) < \delta$. Then

$$\begin{aligned}\int_B g\,d\mu &= \int_B (g-h)\,d\mu + \int_B h\,d\mu \\ &\le \int (g-h)\,d\mu + H\mu(B) \\ &< \frac{\epsilon}{2} + H\delta \\ &< \epsilon.\end{aligned}$$

□

We can also get around the requirement that the measure of the entire space must be finite, like for example in Egorov's Theorem, by restriction attention to a key set of finite measure.

Theorem 131. Let $(X,\mathcal{S},\mu)$ be a measure space, and let $g : X \to [0,\infty]$ be $\mathcal{S}$–measurable such that $\int g\,d\mu < \infty$. Then for every $\epsilon > 0$, there exists $E \in \mathcal{S}$ such that $\mu(E) < \infty$ and

$$\int_{X \setminus E} g\,d\mu < \epsilon.`$$

Proof. Let $\epsilon > 0$. Let P be an $\mathcal{S}$−partition $A_1, \ldots, A_m$ of X such that

$$\int g \, d\mu < \epsilon + \mathcal{L}(g, P).$$

Let E be the union of those A_j such that $\inf_{A_j} f > 0$. Then $\mu(E) < \infty$ because otherwise $\mathcal{L}(g, P) = \infty$, contradicting our hypothesis. Then

$$\begin{aligned}\int_{X \setminus E} g \, d\mu &= \int g \, d\mu - \int \chi_E g \, d\mu \\ &< (\epsilon + \mathcal{L}(g, P)) - \mathcal{L}(\chi_E g, P) \\ &= \epsilon.\end{aligned}$$

□

Now we want to generalize both the Monotone Convergence Theorem and the Bounded Convergence Theorem, as their requirements are too stringent. We use the idea of almost every to be able to do this.

Theorem 132 (Dominated Convergence Theorem)**.** Let $(X, \mathcal{S}, \mu)$ be a measure space and let $f : X \to [-\infty, \infty]$ be $\mathcal{S}$−measurable, and $f_n : X \to \mathbb{R}$ for $n = 1, 2, \ldots$ are $\mathcal{S}$−measurable functions such that

$$\lim_{k \to \infty} f_k(x) = f(x),$$

for almost every $x \in X$. If there exists an $\mathcal{S}$−measurable function $g : X \to [0, \infty]$ such that

$$\int g \, d\mu < \infty, \quad \text{and} \quad |f_k(x)| \le g(x),$$

for all $k \in \mathbb{Z}^+$, and almost every $x \in X$, then we say g *dominates* f, and

$$\lim_{k \to \infty} \int f_k \, d\mu = \int f \, d\mu.$$

Proof. Let $E \in \mathcal{S}$. Then

$$\begin{aligned}\left| \int f_k \, d\mu - \int f \, d\mu \right| &= \left| \int_{X \setminus E} f_k \, d\mu - \int_{X \setminus E} f \, d\mu - \int_E f \, d\mu \right| \\ &\le \left| \int_{X \setminus E} f_k \, d\mu \right| - \left| \int_{X \setminus E} f \, d\mu \right| - \left| \int_E f \, d\mu \right| \\ &\le 2 \int_{X \setminus E} g \, d\mu + \left| \int_E \, d\mu - \int_E f \, d\mu \right|.\end{aligned}$$

We have to consider two cases: First, if $\mu(X) < \infty$, then let $\epsilon > 0$. We know that there exists $\delta > 0$ such that $\int_B g \, d\mu < \frac{\epsilon}{4}$ for every set $B \in \mathcal{S}$ such that $\mu(B) < \delta$.

By Egorov's Theorem, there exists a set $E \in \mathcal{S}$ such that $\mu(X \setminus E) < \delta$ and $f_n \to f$ uniformly on E. Now

$$\left|\int f_k \, d\mu - \int f \, d\mu\right| < \frac{\epsilon}{2} + \left|\int_E (f_k - f) \, d\mu\right|.$$

Because $f_n \to f$ uniformly on E and $\mu(E) < \infty$, the last term on the right is less than $\frac{\epsilon}{2}$ for sufficiently large k. Thus

$$\lim_{k\to\infty} f_k \, d\mu = \int f \, d\mu.$$

Next, if $\mu(X) = \infty$. Let $\epsilon > 0$. Again we know that there exists $E \in \mathcal{S}$ such that $\mu(E) < \infty$ and

$$\int_{X \setminus E} g \, d\mu < \frac{\epsilon}{4}.$$

This gives us that

$$\left|\int f_k \, d\mu - \int f \, d\mu\right| < \frac{\epsilon}{2} + \left|\int_E f_k \, d\mu - \int_E f \, \mu\right|.$$

Using the first case applied to the sequence $f_1|_E, f_2|_E, \ldots$, the last term on the right is less than $\frac{\epsilon}{2}$ for all sufficiently large k. Thus

$$\lim_{k\to\infty} f_k(x) \, d\mu = \int f \, d\mu.$$

□

Let's return to the Riemann integral for a second. How does it interact with our newly-developed tools?

Theorem 133. Suppose $a < b$ and that $f : [a, b] \to \mathbb{R}$ is bounded. Then f is Riemann integrable iff

$$m^*(\{x \in [a, b] \mid f \text{ is continuous at } x) = 0.$$

Furthermore, if f is Riemann integrable, then f is Lebesgue integrable and

$$\int_a^b f = \int_{[a,b]} f \, dm.$$

Definition 115 ($\int_a^b f$)**.** Suppose $-\infty \leq a < b \leq \infty$, and $f : (a, b) \to \mathbb{R}$ is

Lebesgue measurable. Then

$$\int_a^b f \text{ or } \int_a^b f(x)\,dx \text{ are equivalent to } \int_{(a,b)} f\,dm,$$

and

$$\int_b^a f = -\int_a^b f.$$

Now recall our old friend the L^1–norm. We can redefine it in terms of a measure, and when we do so we also slightly change the notation.

Definition 116 ($\mathcal{L}^1$–norm; Lebesgue Space)**.** Let $(X, \mathcal{S}, \mu)$ be a measure space and $f : X \to [-\infty, \infty]$ be $\mathcal{S}$–measurable. Then the $\mathcal{L}^1$*–norm* of f, denoted $\|f\|_1$, is defined by

$$\|f\|_1 = \int |f|\,d\mu.$$

The *Lebesgue space* $\mathcal{L}^1(\mu)$ is defined by

$$\mathcal{L}^1(\mu) = \{f \mid f \text{ is } \mathcal{S}\text{–measurable from } X \text{ to } \mathbb{R} \text{ and } \|f\|_1 < \infty\}.$$

We can approximate functions in the Lebesgue space in three ways:

- Simple functions
- Step functions
- Continuous functions

Explicitly, what is an example of a function in $\mathcal{L}^1(\mu)$? Suppose $E_1, \ldots, E_n$ are disjoint subsets of X in our measure space, and that $a_1, \ldots, a_n$ are distinct nonzero real numbers. Then

$$a_1\chi_{E_1} + \cdots + a_n\chi_{E_n} \in \mathcal{L}^1(\mu),$$

if and only if $E_k \in \mathcal{S}$ and $\mu(E_k) < \infty$ for all $k \in [1, n]$. Furthermore, the $\mathcal{L}^1$–norm will be

$$\|a_1\chi_{E_1} + \cdots + a_n\chi_{E_n}\|_1 = |a_1|\mu(E_1) + \cdots + |a_n|\mu(E_n).$$

Theorem 134. Let μ be a measure and $f \in \mathcal{L}^1(\mu)$. Then for every $\epsilon > 0$, there exists a simple function $g \in \mathcal{L}^1(\mu)$ such that

$$\|f - g\|_1 < \epsilon.$$

Proof. Let $\epsilon > 0$. Then there exists simple functions $g_1, g_2 \in \mathcal{L}^1(\mu)$ such that $0 \le g_1 \le f^+$ and $0 \le g_2 \le f^-$, and

$$\int (f^+ - g_1)\, d\mu < \frac{\epsilon}{2}, \quad \int (f^- - g_2)\, d\mu < \frac{\epsilon}{2}.$$

Let $g = g_1 - g_2$. Then g is another simple function in $\mathcal{L}^1(\mu)$. Hence

$$\begin{aligned}\|f - g\|_1 &= \|(f^+ - g_1) - (f^- - g_2)\|_1 \\ &= \int (f^+ - g_1)\, d\mu + \int (f^- - g_2)\, d\mu \\ &< \epsilon.\end{aligned}$$

□

When we say $\mathcal{L}^1(\mathbb{R})$, this means $\mathcal{L}^1(m)$, where m is the Lebesgue measure on either the Borel subsets of $\mathbb{R}$ or the Lebesgue measurable subsets of $\mathbb{R}$. When in $\mathcal{L}^1(\mathbb{R})$, the notation $\|f\|_1$ denotes the integral of the absolute value of f with respect to the Lebesgue measure on $\mathbb{R}$, as expected.

Definition 117 (Step Function)**.** A *step function* is a function $g : \mathbb{R} \to \mathbb{R}$ of the form

$$g = a_1 \chi_{I_1} + \cdots + a_n \chi_{I_n},$$

where $I_1, \ldots, I_n$ are intervals of $\mathbb{R}$, and $a_1, \ldots, a_n \in \mathbb{R}_{\ge 0}$.

If g is a step function, and $I_1, \ldots, I_n$ are disjoint, then

$$\|g\|_1 = |a_1||I_1| + \cdots + |a_n||I_n|.$$

In particular,

$$g \in \mathcal{L}^1(\mu) \Longleftrightarrow I_1, \ldots, I_n \text{ are bounded}$$

Theorem 135. Let $f \in \mathcal{L}^1(\mathbb{R})$. Then for every $\epsilon > 0$, there exists a step function $g \in \mathcal{L}^1(\mathbb{R})$ such that

$$\|f - g\|_1 < \epsilon.$$

Finally, Luzin's Theorem gives us a way to approximate Borel measurable functions by continuous functions.

Theorem 136. Let $f \in \mathcal{L}^1(\mathbb{R})$. Then for every $\epsilon > 0$, there exists a continuous function $g : \mathbb{R} \to \mathbb{R}$ such that

$$\|f - g\|_1 < \epsilon,$$

and $\{x \in \mathbb{R} \mid g(x) \neq\}$ is a bounded set.

14.4 Uniform Integrability

We conclude this chapter on Lebesgue integration by establishing, for functions that are integrable over a set of finite measure, a more general criterion for justifying passage of the limit under the integral sign.

Theorem 137. Let $E \subset \mathbb{R}$ be such that $\mu(E) < \infty$, and let $\delta > 0$. Then E is the disjoint union of a finite collection of sets , each of which has measure less than δ.

Proof. By the continuity of measure,

$$\lim_{n\to\infty} \mu(E \setminus [-n,n]) = \mu(\emptyset) = 0.$$

Choose some $k \in \mathbb{Z}^+$ such that $m(E \setminus [-k,k]) < \delta$. By choosing a fine enough partition of $[-k,k]$, express $E \cap [-k,k]$ as the disjoint union of a finite collection of sets, each of which has measure less than δ. □

Theorem 138. Let $(X, \mathcal{S}, \mu)$ be a measure space and let $f \in \mathcal{L}^1(\mu)$. If f is integrable over X, then for each $\epsilon > 0$, there exists some $\delta > 0$ such that

$$\text{if } E \subset X \text{ is measurable and } \mu(E) < \delta, \text{ then } \int_E |f|\, d\mu < \epsilon.$$

Definition 118 (Uniform Integrability)**.** A family $\mathcal{F}$ of measurable functions on X is said to be *uniformly integrable* over $E \subset X$ if for each $\epsilon > 0$, there exists $\delta > 0$ such that for each $f \in \mathcal{F}$,

$$\text{if } E \subset X \text{ is measurable and } \mu(E) < \delta, \text{ then } \int_E |f|\, d\mu < \epsilon.$$

As an example, say we take

$$\mathcal{F} = \{f \mid f \text{ is measurable on } E \text{ and } |f| \leq g \text{ on } E\}.$$

Theorem 139. Let $f_n : E \to \mathbb{R}$ for $n = 1, 2, \cdots < \infty$ be a sequence of integrable functions over E. Then f_n is uniformly integrable.

15 Measure-Theoretic Differentiation

To develop a rigorous theory of differentiation, we must first cover a few essential lemmas/theorems that serve as the foundation for an almost-everywhere version of the Fundamental Theorem of Calculus.

15.1 Preliminaries

Theorem 140 (Markov's Inequality)**.** Let $(X, \mathcal{S}, \mu)$ be a measure space and $h \in \mathcal{L}^1(\mu)$. Then

$$\mu(\{x \in X \mid |h(x)| \geq c\}) \leq \frac{1}{c}\|h\|_1,$$

for all $c > 0$.

Proof. Let $c > 0$. Then

$$\begin{aligned}\mu(\{x \in X \mid |h(x)| \geq c\}) &= \frac{1}{c}\int_{\{x \in X \mid |h(x)| \geq c\}} c \, d\mu \\ &\leq \frac{1}{c}\int_{\{x \in X \mid |h(x)| \geq c\}} |h| \, d\mu \\ &\leq \frac{1}{c}\|h\|_1.\end{aligned}$$

□

Now, intuitvely speaking, the notation $n * I$, where I is a bounded nonempty open interval, is exactly what we expect it to be: another interval that has n times the length of I that is centered at the same spot. For example, if $I = (0, 10)$, then $3 * I = (-10, 20)$.

Theorem 141 (Vitali Covering Lemma)**.** Let $I_1, \ldots, I_n$ be a list of bounded nonempty open intervals of $\mathbb{R}$. Then there exists a disjoint sublist $I_{k_1}, \ldots, I_{k_m}$ such that

$$I_1 \cup \cdots \cup I_n \subset (3 * I_{k_1}) \cup \cdots \cup (3 * I_{k_m}).$$

Example 19. Let $n = 4$, and

$$I_1 = (0, 10), \quad I_2 = (9, 15), \quad I_3 = (14, 22), \quad I_4 = (21, 31).$$

Then

$$3 * I_1 = (-10, 20), \quad 3 * I_2 = (3, 21), \quad 3 * I_3 = (6, 30), \quad 3 * I_4 = (11, 41).$$

Thus

$$I_1 \cup I_2 \cup I_3 \cup I_4 \subset (3 * I_1) \cup (3 * I_4).$$

Proof. Let k_1 be such that

$$|I_{k_1}| = \max\{|I_1|, \ldots, |I_n|\}.$$

□

Definition 119 (Hardy-Littlewood Maximal Function)**.** Let $h : \mathbb{R} \to \mathbb{R}$ be a Lebesgue measurable function. Then the *Hardy-Littlewood maximal function* of h is the function $h^* : \mathbb{R} \to [0, \infty]$ defined by

$$h^*(b) = \sup_{t \geq 0} \frac{1}{2t} \int_{b-t}^{b+t} |h|.$$

To put this somewhat arbitrary construction in words, $h^*(b)$ gives us the supremum over all bounded intervals centered at b of the average of $|h|$ on those intervals. To explicitly calculate this function for something we know, if we let $h(x) = \chi_{[0,1]}(x)$, then

$$h^*(b) = \begin{cases} \frac{1}{2(1-b)}, & b \leq 0 \\ 1, & 0 < b < 1 \\ \frac{1}{2b}, & b \geq 1 \end{cases}.$$

Theorem 142 (Hardy-Littlewood Maximal Inequality)**.** Suppose $h \in \mathcal{L}^1(\mathbb{R})$. Then

$$m(\{b \in \mathbb{R} \mid h^*(b) > c\}) \leq \frac{3}{c} \|h\|_1,$$

for every $c > 0$.

Proof. Let F be a closed, bounded subset of $\{b \in \mathbb{R} \mid h^*(b) > c\}$. We want to show that $|F| \leq \frac{3}{c} \int_{-\infty}^{\infty} |h|$. For each $b \in F$, there exists $t_b > 0$ such that

$$\frac{1}{2t_b} \int_{b-t_b}^{b+t_b} |h| > c.$$

Clearly

$$F \subset \bigcup_{b \in F} (b - t_b, b + t_b).$$

By Heine-Borel, this open cover must have a finite subcover. In other words, there exists $b_1, \dots, b_n \in F$ such that

$$F \subset (b_1 - t_{b_1}, b_1 + t_{b_1}) \cup \cdots \cup (b_n - t_{b_n}, b_n + t_{b_n}).$$

We label the above intervals as $I_1, \dots, I_n$. By the Vitali Covering Lemma, there exists a disjoint sublist $I_{k_1}, \dots, I_{k_m}$ such that

$$I_1 \cup \cdots \cup I_n \subset (3 * I_{k_1}) \cup \cdots \cup (3 * I_{k_m}).$$

Hence

$$\begin{aligned}
|F| &\le |I_1 \cup \cdots \cup I_n| \\
&\le |(3 * I_{k_1}) \cup \cdots \cup (3 * I_{k_m})| \\
&\le |3 * I_{k_1}| + \cdots + |3 * I_{k_m}| \\
&= 3(|I_{k_1}| + \cdots + |I_{k_m}|) \\
&< \frac{3}{c}\left(\int_{I_{k_1}} |h| + \cdots + \int_{I_{k_m}} |h|\right) \\
&\le \frac{3}{c}\int_{-\infty}^{\infty} |h|.
\end{aligned}$$

□

15.2 Derivatives of Integrals

Theorem 143 (Lebesgue Differentiation Theorem; Version 1)**.** Let $f \in \mathcal{L}^1(\mathbb{R})$. Then

$$\lim_{t \to 0} \frac{1}{2t} \int_{b-t}^{b+t} |f - f(b)| = 0$$

for almost every $b \in \mathbb{R}$.

This tells us that the average amount by which a function in $\mathcal{L}^1(\mathbb{R})$ differs from its values is small almost everywhere on small intervals.

Proof. □

We already saw the Fundamental Theorem of Calculus before, but our requirements have now been loosened that f only needs to be Lebesgue measurable, and its absolute value must have a finite Lebesgue integral.

Theorem 144 (Fundamental Theorem of Calculus)**.** Let $f \in \mathcal{L}^1(\mathbb{R})$. Define $g : \mathbb{R} \to \mathbb{R}$ by

$$g(x) = \int_{-\infty}^{x} f.$$

Suppose $b \in \mathbb{R}$ and f is continuous at b. Then g is differentiable at b and

$$g'(b) = f(b).$$

Proof. If $t \neq 0$, then

$$\begin{aligned}
\left| \frac{g(b+t) - g(b)}{t} - f(b) \right| &= \left| \frac{\int_{-\infty}^{b+t} f - \int_{-\infty}^{b} f}{t} - f(b) \right| \\
&= \left| \frac{\int_{b}^{b+t} f}{t} - f(b) \right| \\
&= \left| \frac{\int_{b}^{b+t} (f - f(b))}{t} \right| \\
&\leq \sup_{\{x \in \mathbb{R} \mid |x-b| < |t|\}} |f(x) - f(b)|.
\end{aligned}$$

If $\epsilon > 0$, then by continuity of f at b, the last quantity is less than ϵ for t sufficiently close to 0. Thus g is differentiable at b, and $g'(b) = f(b)$. □

Since a function in $\mathcal{L}^1(\mathbb{R})$ need not be continuous anywhere, the Fundamental Theorem of Calculus might provide no information about differentiating the integral of such a function. This means we need another theorem.

Theorem 145 (Lebesgue Differentiation Theorem; Version 2)**.** Let $f \in \mathcal{L}^1(\mathbb{R})$. Define $g : \mathbb{R} \to \mathbb{R}$ by

$$g(x) = \int_{-\infty}^{x} f.$$

Then $g'(b) = f(b)$ for almost every $b \in \mathbb{R}$.

Proof. Let $t \neq 0$. Then

$$\begin{aligned}
\left| \frac{g(b+t) - g(b)}{t} - f(b) \right| &= \left| \frac{\int_{b}^{b+t} (f - f(b))}{t} \right| \\
&\leq \frac{1}{t} \int_{b}^{b+t} |f - f(b)| \\
&\leq \frac{1}{t} \int_{b-t}^{b+t} |f - f(b)|,
\end{aligned}$$

for all $b \in \mathbb{R}$. By the first version of the Lebesgue Differentiation Theorem, □

Theorem 146. There does not exist a Lebesgue measurable set $E \subset [0,1]$ such that

$$m(E \cap [0,b]) = \frac{b}{2},$$

for all $b \in [0,1]$.

Proof. Suppose to the contrary that there did exist a Lebesgue measurable set. Define $g : \mathbb{R} \to \mathbb{R}$ by

$$g(b) = \int_{-\infty}^{b} \chi_E.$$

Thus $g(b) = \frac{b}{2}$ for all $b \in [0,1]$. Hence $g'(b) = \frac{1}{2}$ for all $b \in [0,1]$. The Lebesgue Differentiation Theorem implies that $g'(b) = \chi_E(b)$ for almost every $b \in \mathbb{R}$. However, χ_E never takes on the value $\frac{1}{2}$, which contradicts the conclusion before. □

Theorem 147. Let $f \in \mathcal{L}^1(\mathbb{R})$. Then for almost every $b \in \mathbb{R}$,

$$f(b) = \lim_{t \to 0} \frac{1}{2t} \int_{b-t}^{b+t} f.$$

This theorem tells us that a function in $\mathcal{L}^1(\mathbb{R})$ is equal almost everywhere to the limit of its average over small intervals. It holds at every number b at which f is continuous. Remarkably, even if f is discontinuous everywhere, the conclusion still holds for almost every real number b.

Proof. Let $t > 0$. Then

$$\begin{aligned} \left| \left(\frac{1}{2t} \int_{b-t}^{b+t} f \right) - f(b) \right| &= \left| \frac{1}{2t} \int_{b-t}^{b+t} (f - f(b)) \right| \\ &\leq \frac{1}{2t} \int_{b-t}^{b+t} |f - f(b)|. \end{aligned}$$

We conclude by applying the Lebesgue Differentiation Theorem. □

Definition 120 (Density)**.** Let $E \subset \mathbb{R}$. The *density* of E at a number $b \in \mathbb{R}$ is

$$\lim_{t \to 0} \frac{m(E \cap (b-t, b+t))}{2t},$$

if this exists, and otherwise is undefined.

The density of $[0,1]$ at b will be

$$\begin{cases} 1, & b \in (0,1) \\ \frac{1}{2}, & b = 0 \text{ or } b = 1 \\ 0, & \text{otherwise} \end{cases}$$

Theorem 148 (Lebesgue Density Theorem)**.** Let $E \subset \mathbb{R}$ be a Lebesgue measurable set. Then the density of E is 1 at almost every element of E and is 0 at almost every element of $\mathbb{R} \setminus E$.

16 Intro to Functional Analysis

We take a survey of select functional analysis topics in this chapter. This chapter assumes knowledge of complex analysis (122A) and linear algebra (108AB).

16.1 Banach Spaces

Definition 121 (Sequence Space; ℓ^p)**.**

Definition 122 (Complete Metric Space)**.** A metric space (M,d) is called *complete* if every Cauchy sequence in M converges to some element of M.

Not every metric that we think of is complete. The metric space $(\mathbb{Q},d)$, where $d(x,y) = |x-y|$, is not complete. To see this, for $k \in \mathbb{Z}^+$, let

$$x_k = \frac{1}{10^{1!}} + \frac{1}{10^{2!}} + \cdots + \frac{1}{10^{k!}}.$$

If $j < k$, then

$$|x_k - x_j| = \frac{1}{10^{(j+1)!}} + \cdots + \frac{1}{10^{k!}} < \frac{2}{10^{(j+1)!}}.$$

Thus $x_1, x_2, \ldots$ is a Cauchy sequence in $\mathbb{Q}$. However, $x_1, x_2, \ldots$ does not converge to an element of $\mathbb{Q}$ because the limit of this sequence would have decimal expansion $0.110001000000000000000001\ldots$ that is neither a terminating nor repeating decimal.

Definition 123 (Banach Space)**.** A complete normed vector space is called a *Banach space*.

Some examples and nonexamples:

- The vector space $C([0,1])$ with the norm defined by

$$\|f\| = \sup_{[0,1]} |f|,$$

the C–norm, is a Banach space. In other words, $(C([0,1]), \|\cdot\|_{C([0,1])})$ is a Banach space.

- The vector space ℓ^1 with the norm

$$\|(a_1, a_2, \ldots)\|_1 = \sum_{k=1}^{\infty} |a_k|$$

is a Banach space.

- The normed space $(C([0,1]), \|\cdot\|_1)$ is not a Banach space.

Theorem 149. Let V be a normed space. Then V is a Banach space if and only if $\sum_{k=1}^{\infty} g_k$ converges for every sequence $g_1, g_2, \ldots$ in V such that $\sum_{k=1}^{\infty} \|g_k\| < \infty$.

Proof. $(\Longrightarrow)$: Let V be a Banach space. Let $g_1, g_2, \ldots$ be a sequence in V such that $\sum_{k=1}^{\infty} \|g_k\| < \infty$. Suppose $\epsilon > 0$. Let $n \in \mathbb{Z}^+$ such that $\sum_{m=n}^{\infty} \|g_m\| < \epsilon$. For $j \in \mathbb{Z}^+$, let f_j denote the partial sum defined by

$$f_j = g_1 + \cdots + g_j.$$

If $k > j \geq n$, then

$$\begin{aligned} \|f_k - f_j\| &= \|g_{j+1} + \cdots + g_k\| \\ &\leq \|g_{j+1}\| + \cdots + \|g_k\| \\ &\leq \sum_{m=n}^{\infty} \|g_m\| \\ &< \epsilon. \end{aligned}$$

Thus $f_1, f_2, \ldots$ is a Cauchy sequence in V. Because V is Banach, this sequence converges to some element of V, which is precisely what it means for $\sum_{k=1}^{\infty} g_k$ to converge.
$(\Longleftarrow)$: Suppose $\sum_{k=1}^{\infty} g_k$ converges for every sequence $g_1, g_2, \ldots$ in V such that $\sum_{k=1}^{\infty} \|g_k\| < \infty$. Suppose $f_1, f_2, \ldots$ is a Cauchy sequence in V. It suffices to show that some subsequence of this converges. Dropping to a subsequence and setting $f_0 = 0$, we can assume that

$$\sum_{k=1}^{\infty} \|f_k - f_{k-1}\| < \infty.$$

Hence $\sum_{k=1}^{\infty} (f_k - f_{k-1})$ converges. The partial sum of this series after n terms is f_n. Thus $\lim_{n\to\infty} f_n$ exists, completing the proof. □

Recall from linear algebra what a *linear map*, or *linear transformation* is. If instead of two vector spaces, we consider normed vector spaces, then we form a new type of linear map.

Definition 124 (Bounded Linear Map)**.** Suppose V and W are normed vector spaces and $T : V \to W$ is a linear map. The *norm* of T, denoted $\|T\|$, is defined by

$$\|T\| = \sup\{\|T(f)\| \mid f \in V \text{ and } \|f\| \leq 1\}.$$

We call T *bounded* if $\|T\| < \infty$, and we denote the set of bounded linear maps from V to W as $\mathcal{B}(V, W)$.

Consider the Banach space $(C([0,3]), \|\cdot\|_{C([0,3])})$. We define $T : C([0,3]) \to C([0,3])$ by

$$T(f(x)) = x^2 f(x).$$

Then T is a bounded linear map, and $\|T\| = 9$.

Theorem 150. Let V and W be normed spaces. Then $\|S + T\| \leq \|S\| + \|T\|$ and $\|\alpha T\| = |\alpha|\|T\|$ for all $S, T \in \mathcal{B}(V, W)$ and all $\alpha \in \mathbb{F}$. Furthermore, the function $\|\cdot\|$ is a norm on $\mathcal{B}(V, W)$.

Proof. Let $S, T \in \mathcal{B}(V, W)$. Then

$$\begin{aligned} \|S + T\| &= \sup\{\|(S + T)(f)\| \mid f \in V \text{ and } \|f\| \leq 1\} \\ &\leq \sup\{\|S(f)\| + \|T(f)\| \mid f \in V \text{ and } \|f\| \leq 1\} \\ &\leq \sup\{\|S(f)\| \mid f \in V \text{ and } \|f\| \leq 1\} \\ &+ \sup\{\|T(f)\| \mid f \in V \text{ and } \|f\| \leq 1\} \\ &= \|S\| + \|T\|. \end{aligned}$$

□

Theorem 151. Let V be a normed space and W a Banach space. Then $\mathcal{B}(V, W)$ is a Banach space.

Proof. Let $T_1, T_2, \ldots$ be a Cauchy sequence in $\mathcal{B}(V, W)$. If $f \in V$, then

$$\|T_i(f) - T_k(f)\| \leq \|T_j - T_k\|\|f\|,$$

which implies that $T_1(f), T_2(f), \ldots$ is a Cauchy sequence in W. Because W is Banach, this sequence has a limit in W, which we call $T(f)$. We have now defined a function $T : V \to W$, which is a linear map. Clearly for each $f \in V$,

$$\begin{aligned} \|T(f)\| &\leq \sup\{\|T_k(f) \mid k \in \mathbb{Z}^+\} \\ &\leq (\sup\{\|T_k\| \mid k \in \mathbb{Z}^+)\|f\|. \end{aligned}$$

Thus $T \in \mathcal{B}(V, W)$. However, we still need to show that $\lim_{k\to\infty} \|T_k - T\| = 0$. Let $\epsilon > 0$, and let $n \in \mathbb{Z}^+$ be such that $\|T_i - T_k\| < \epsilon$ for all $j \geq n$ and $k \geq n$. Suppose $j \geq n$ and $f \in V$. Then

$$\begin{aligned}\|(T_j - T)(f)\| &= \lim_{k\to\infty} \|T_i(f) - T_k(f)\| \\ &\leq \epsilon\|f\|.\end{aligned}$$

Thus $\|T_j - T\| \leq \epsilon$, completing our proof. □

Theorem 152. A linear map from one normed vector space to another normed vector space is continuous if and only if it is bounded.

Proof. Let V and W be normed vector spaces and let $T : V \to W$ be a linear map.
$(\Longrightarrow)$: Suppose T is continuous but not bounded. Then there exists a sequence $f_1, f_2, \ldots$ in V such that $\|f_k\| \leq 1$ for all $k \in \mathbb{Z}^+$ and $\|T(f_k)\| \to \infty$ as $k \to \infty$. Hence

$$\lim_{k\to\infty} \frac{f_k}{\|T(f_k)\|} = 0, \quad T\left(\frac{f_k}{\|T(f_k)\|}\right) = \frac{T(f_k)}{\|T(f_k)\|} \not\to 0.$$

However, this implies that T is not continuous, a contradiction.
$(\Longleftarrow)$: Suppose T is bounded. Let $f \in V$ and $f_1, f_2, \ldots$ is a sequence in V such that $\lim_{k\to\infty} f_k = f$. Then

$$\begin{aligned}\|T(f_k) - T(f)\| &= \|T(f_k - f)\| \\ &\leq \|T\|\|f_k - f\|.\end{aligned}$$

Thus $\lim_{k\to\infty} T(f_k) = T(f)$, and so T is continuous. □

Recall that a *linear functional* on a vector space V is a linear map from V to $\mathbb{F}$. We introduce the following notation: if V is a real vector space, U is a subspace of V, and $h \in V$, then

$$U + \mathbb{R}h = \{f + \alpha h \mid f \in U \text{ and } \alpha \in \mathbb{R}\}.$$

Theorem 153 (Extension Lemma)**.** Let V be a real normed vector space, and let U be a subspace of V. Let $\psi : U \to \mathbb{R}$ is a bounded linear functional. Suppose $h \in V \setminus U$. Then ψ can be extended to a bounded linear functional $\varphi : U + \mathbb{R}h \to \mathbb{R}$ such that $\|\varphi\| = \|\psi\|$.