Real Analysis Notes

Contents

Abstract

These notes cover both standard real analysis, measure theory and integration (with a little bit of functional analysis), and analysis on manifolds. Measure theory starts at Lebesgue integration, and analysis on manifolds starts at multivariable differential calculus.

1. Sequences

Definition (convergence): A sequence (a_n) converges to $a \in \mathbb{R}$ if for all $\varepsilon > 0$, there exists some N such that $n > N \Rightarrow |a_n - a| < \varepsilon$.

Definition (divergence): A sequence can either diverge to positive infinity (for all $M > 0$, there exists an N such that $n > N \Rightarrow a_n > M$), negative infinity (for all $M < 0$, there exists an N such that $n > N \Rightarrow a_n < M$), or neither, in which case the limit does not exist.

Proposition: If a sequence converges, then the limit is unique.

Proof: Suppose $a_n \to x, y$, where $x \neq y$. We know that $|a_n - x|, |a_n - y| < \frac{\varepsilon}{2}$ for arbitrarily large n. Thus we have

$$|x - y| \leq |x - a_n| + |a_n - y| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

However, this holds for every $\varepsilon > 0$, which implies $x = y$, a contradiction. ■

Proposition: Convergent sequences are bounded.

Proof: Note that eventually $|a_n - a| < 1$, where $\lim_{n\to\infty} a_n = a$. Thus $1 - a < a_n < 1 + a$. Now just take the max and min of the finitely many terms that occur before this happens to get bounds on a_n. ■

Proposition:
a) $(c \cdot a_n) \to c \cdot a$
b) $(a_n + b_n) \to a + b$
c) $(a_n - b_n) \to a - b$
d) $(a_n \cdot b_n) \to a \cdot b$
e) $\left(\frac{a_n}{b_n}\right) \to \frac{a}{b}$

Proof:
a) Suppose $\varepsilon > 0$. Then there exists N such that for all $n \geq N$, we have

$$|a_n - a| < \frac{\varepsilon}{|c|} \Rightarrow |c \cdot a_n - c \cdot a| < \varepsilon.$$

Thus $\lim_{n\to\infty} c \cdot a_n = c \cdot a$.
b) Suppose $\varepsilon > 0$. Then there exists N_1, N_2 such that for all $n_1 \geq N_1, n_2 \geq N_2$, we have

$$\left|a_{n_1} - a\right|, \left|b_{n_2} - b\right| < \frac{\varepsilon}{2}.$$

Set $N = \max\{N_1, N_2\}$. Then, for all $n \geq N$, we have

$$|(a_n + b_n) - (a + b)| \leq |a_n - a| + |b_n - b| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

Thus $\lim_{n \to \infty}(a_n + b_n) = a + b$.

c) Negate (b_n) and use the last two bullets.

d) Since (a_n) converges, we have $|a_n| \leq C$ for some C for all n. There exists some N_1 such that for all $n \geq N_1$, we have $|a_n - a| < \frac{\varepsilon}{2|b|+1}$ (note that $2|b| + 1 > 0$). Similarly, there exists some N_2 such that for all $n \geq N_2$, we have $|b_n - b| < \frac{\varepsilon}{2C+1}$ (note that $2C + 1 > 0$). Let $N = \max\{N_1, N_2\}$. Then, for all $n \geq N$, we have

$$|a_n b_n - ab| = |a_n b_n - a_n b + a_n b - ab| \leq |a_n||b_n - b| + |a_n - a||b| < C \cdot \frac{\varepsilon}{2C + 1} + |b| \cdot \frac{\varepsilon}{2|b| + 1} < \varepsilon.$$

Thus $\lim_{n \to \infty} a_n b_n = ab$.

e) Reciprocate (b_n) (assuming only finitely many terms are 0), and apply the last bullet.

■

Proposition: Suppose (a_n) and (b_n) convergent series and $a_n \leq b_n$ for all n. Then

$$\lim_{n \to \infty} a_n \leq \lim_{n \to \infty} b_n.$$

Proof: Let $a_n \to A$ and $b_n \to B$, and suppose for the sake of contradiction that $A > B$. Then for sufficiently large n we have

$$|a_n - A| < \frac{A - B}{2} \quad \text{and} \quad |b_n - B| < \frac{A - B}{2}.$$

Expanding the absolute values yields

$$\frac{3B - A}{2} < b_n < \frac{A + B}{2} < a_n < \frac{3A - B}{2},$$

which is a contradiction. ■

Theorem (squeeze theorem): Suppose $a_n \leq x_n \leq b_n$ for arbitrarily large n and $a_n, b_n \to L$. Then $x_n \to L$.

Proof: We have

$$L - \varepsilon < a_n \leq x_n \leq b_n < L + \varepsilon$$

for arbitrarily large n, which implies $|x_n - L| < \varepsilon$. ■

Theorem (monotone convergence theorem): A monotone sequence converges if and only if it is bounded. Further, if the sequence is increasing and bounded, then it converges to the supremum of the set of elements of the sequence. If it's decreasing and bounded, then it converges to the infinum of the set of the elements of the sequence. If a monotone sequence diverges, then it diverges to ∞ or $-\infty$, depending on if it's increasing or decreasing.

Proof: If the sequence converges, then clearly it's bounded. Now suppose the sequence is monotone increasing and bounded. Let (a_n) be the sequence and let $S = \{a_n \mid n \geq 1\}$. Since the sequence is bounded, S is bounded, so $\sup(S)$ exists. We claim that $\lim_{n\to\infty} a_n = \sup(S)$. By definition of supremum, for all $\varepsilon > 0$, there exists some N such that $\sup(S) - \varepsilon \leq a_N \leq \sup(S)$. Since the sequence is increasing, this implies $\sup(S) - \varepsilon \leq a_N \leq a_n \leq \sup(S)$ for all $n \geq N$. This implies that $|\sup(S) - a_n| < \varepsilon$ for all $n \geq N$, which means a_n converges to $\sup(S)$ as desired. Negating the sequence proves the infinum case.

The divergence part of the theorem just means that the sequence doesn't bounce around, which is obvious from monotonicity. ■

Proposition: Suppose $S \subseteq \mathbb{R}$ is bounded above. Then there exists a sequence (a_n) where $a_n \in S$ for each n and

$$\lim_{n\to\infty} a_n = \sup(S).$$

Similarly, if S is bounded below, then there exists a sequence (b_n) where $b_n \in S$ for each n and

$$\lim_{n\to\infty} b_n = \inf(S).$$

Proof: We prove the infinum case, as the supremum case follows upon negation.

Note by definition, for each $n \geq 1$, there exist some $x \in S$ such that $\inf(S) \leq x \leq \inf(S) + \frac{1}{n}$. Let such an x be a_n. Then we have

$$\inf(S) \leq a_n \leq \inf(S) + \frac{1}{n}.$$

Note that both the left and the right converge to $\inf(S)$, and thus by the squeeze theorem, (a_n) must also converge to $\inf(S)$. ■

Proposition: A sequence converges to a if and only if every subsequence converges to a.

Proof: Since the oringinal sequence is a subsequence, if all subsequences converge, then so does the original.

Now suppose the original sequence $(a_n) \to a$, and consider some arbitrary subsequence $\left(a_{n_k}\right)$. For all $\varepsilon > 0$, there exists N such that for all $n \geq N$, we have $|a_n - a| < \varepsilon$. Now note that for all $k \geq K$ for some K, we have $n_k \geq N$. Thus, for all $k \geq K$, we have $\left|a_{n_k} - a\right| < \varepsilon$, which means $\left(a_{n_k}\right) \to a$. ■

Proposition: If a monotone sequence (a_n) has a convergent subsequence, then (a_n) converges to the same limit.

Proof: Suppose the sequence is monotone increasing (decreasing is proved the exact same). Then clearly the subsequence is increasing as well. We know by the monotone convergence theorem that

$$\lim_{n\to\infty} a_{n_k} = \sup\left(\left\{a_{n_k} : k \geq 1\right\}\right).$$

Then since $k \leq n_k$ for all k (n_k is a subsequence of $\mathbb{N}$), we have

$$a_k \leq a_{n_k} \leq \sup\left(\left\{a_{n_k} : k \geq 1\right\}\right).$$

Thus (a_n) is bounded, so by monotone convergence, it converges. Thus, since every subsequence converges to the main series' limit, $(a_n) \to \sup\left(\left\{a_{n_k} : k \geq 1\right\}\right)$. ■

Lemma: Every sequence has a monotone subsequence.

Proof: Let (a_n) be the sequence. Define a peak to be an element of the sequence that's bigger than every later element. First suppose the sequence has finitely many peaks. To start the subsequence, pick the next element after the last peak. Then, since there are no more peaks, there must be an element bigger than the one chosen. We can keep doing this and get an increasing subsequence.

Now suppose there are infinitely many peaks. Then each peak must be less than the previous one by definition, so the subsequence of peaks is monotone decreasing. ■

Theorem (Bolzano-Weierstrass theorem): Every bounded sequence has a convergent subsequence.

Proof: Every sequence has a monotone subsequence, and since the original sequence is bounded, this subsequence is bounded. Thus it converges by monotone convergence. ■

Definition (Cauchy): A sequence (a_n) is *Cauchy* if for all $\varepsilon > 0$ there exists some N such that $|a_m - a_n| < \varepsilon$ for all $m, n \geq N$.

Proposition: Every Cauchy sequence is bounded.

Proof: There exists N such that for all $m, n \geq N$, we have

$$|a_m - a_n| < 1.$$

Thus, for all $m \geq N$, we have

$$|a_m - a_N| < 1.$$

This bounds a_m with $m \geq N$ between $a_N - 1$ and $a_N + 1$. Then, simply take the maximum and minimum of all the previous terms to see that the sequence is indeed bounded. ■

Theorem: A sequence converges if and only if it is Cauchy.

Proof: First suppose $(a_n) \to a$. Then, for all $\varepsilon > 0$, there exists N such that for all $n \geq N$, we have $|a_n - a| < \frac{\varepsilon}{2} \Rightarrow -\frac{\varepsilon}{2} < a_n - a < \frac{\varepsilon}{2}$. For any $m \geq N$, we also have $|a_m - a| < \frac{\varepsilon}{2} \Rightarrow -\frac{\varepsilon}{2} < a - a_m < \frac{\varepsilon}{2}$. Adding the two yields

$$-\varepsilon < a_n - a_m < \varepsilon \Rightarrow |a_n - a_m| < \varepsilon$$

for all $n, m \geq N$. Thus (a_n) is Cauchy.

Now suppose (a_n) is Cauchy. Thus, (a_n) is bounded, and so by Bolzano Weierstrass, there is some convergent subsequence. Let this subsequence be $\left(a_{n_k}\right) \to a$. Thus for all $\varepsilon > 0$, there exists K such that for all $n_k \geq K$, we have

$$\left|a_{n_k} - a\right| < \frac{\varepsilon}{2}.$$

Since (a_n) is Cauchy, for all $\varepsilon > 0$, there exists M such that for all $m, n_k \geq M$, we have

$$\left|a_m - a_{n_k}\right| < \frac{\varepsilon}{2}.$$

Let $N = \max\{K, M\}$. Let $m, n_k \geq N$. Then both inequalities are true. Adding the two and using the triangle inequality yields

$$|a_m - a| \leq \left|a_m - a_{n_k}\right| + \left|a_{n_k} - a\right| < \varepsilon.$$

Thus $(a_n) \to a$. ■

Definition (limsup and liminf): Let (x_n) be a sequence. Then define

$$\limsup_{n \to \infty} x_n = \lim_{n \to \infty} \left(\sup_{m \geq n} x_m \right)$$

and

$$\liminf_{n \to \infty} x_n = \lim_{n \to \infty} \left(\inf_{m \geq n} x_m \right).$$

Remark: Most of the time we will write $\limsup x_n$ to signify the limit superior, and similarly for the limit inferior.

Proposition: Let $\limsup_{n\to\infty} x_n = L$ and $\liminf_{n\to\infty} x_n = M$. Then, for all $\varepsilon > 0$, there exists N_1 such that for all $n \geq N_1$ we have

$$L + \varepsilon > x_n.$$

Similarly, there exists some N_2 such that for all $n \geq N_2$ we have

$$x_n > M - \varepsilon.$$

Proof: We prove the infinum case, as the superior case follows similarly.

We proceed by contradiction. Suppose there exists some $\varepsilon > 0$ such that for all N, there exists some $n \geq N$ such that $x_n \leq M - \varepsilon$. Thus $\inf_{m\geq n} x_m \leq M - \varepsilon$. Thus we have

$$\varepsilon \leq M - \inf_{m\geq n} x_m = \left|M - \inf_{m\geq n} x_m\right|.$$

However, this is a contradiction, since $\liminf x_n = M$. ■

Proposition: Let $L = \limsup_{n\to\infty} x_n$ and $M = \liminf_{n\to\infty} x_n$. Then, for all $\varepsilon > 0$, there exist infinitely many N such that

$$L \geq x_N \geq L - \varepsilon$$

and

$$M \leq x_N \leq M + \varepsilon.$$

Proof: We do the supremum case, as the infinum case follows similarly.

Let $\varepsilon > 0$. Suppose for the sake of contradiction that there are only finitely many N such that $L \geq x_N \geq L - \varepsilon$. Let N' be the last of these. Then, for all $n > N'$, we have

$$L - \varepsilon > x_n.$$

This implies that for all $n > N'$, we have

$$\sup_{m\geq n} x_m \leq L - \varepsilon < L \Rightarrow \sup_{m\geq n} x_n - L \leq -\varepsilon < 0 \Rightarrow \left|\sup_{m\geq n} x_n - L\right| \geq \varepsilon.$$

However, this contradicts the fact that $L = \limsup_{n\to\infty} x_n$. Thus we have a contradiction. ■

Proposition: Suppose (x_n) is a bounded sequence. Then there is a subsequence that converges to $\limsup_{n\to\infty} x_n$ and a subsequence that converges to $\liminf_{n\to\infty} x_n$.

Proof: We prove the supremum case, as the infinum case follows similarly. Let $\limsup_{n\to\infty} x_n = L \in \mathbb{R}$, which exists because (x_n) is bounded.. Let N_1 be the smallest integer such that $L - 1 \leq a_{N_1} \leq L$. Then let N_2 be the smallest integer greater than N_1 such that $L - \frac{1}{2} \leq a_{N_2} \leq L$. We know this must exist since by the previous proposition, there are infinitely such N_2 that satisfy the inequality. We can then inductively build the sequence, taking the smallest interger N_k greater than N_{k-1} such that $L - \frac{1}{k} \leq a_{N_k} \leq L$. Then by the squeeze theorem we have that $\left(x_{N_k}\right)$ converges to L, as desired.

Remark: This also proves Bolzano-Weierstrass.

■

Proposition: A sequence converges if and only if $\limsup_{n\to\infty} x_n = \liminf_{n\to\infty} x_n$.

Proof: Note that

$$\inf_{m\geq n} x_m \leq x_n \leq \sup_{m\geq n} x_m$$

by definition. Thus, by the squeeze theorem, we can conclude (x_n) converges to $\limsup x_n = \liminf x_n$.

Now suppose $L = \limsup x_n \neq \liminf x_n = M$. By the previous proposition, there are two subsequences that converge to L and M. Since they converge to different numbers, we must have that $\lim_{n\to\infty} x_n$ does not exist. ■

1.1. Problems

Problem: Suppose (a_n) is a sequence and $f : \mathbb{N} \to \mathbb{N}$ is a bijection. Prove the following:
a) if (a_n) diverges to ∞, then $\left(a_{f(n)}\right)$ diverges to ∞.
b) if (a_n) converges to L, then $\left(a_{f(n)}\right)$ converges to L.

Solution:
a) We have that for every M, there exists N such that $\forall n \geq N$, we have $a_n > M$. Since f is a bijection, there exists some N' such that for all $n' \geq N'$, we have $f(n') \geq N$ (this is because eventually every number less than N, it will be an output of some input to f). Thus $\left(a_{f(n)}\right)$ dose diverge to infinity.
b) Basically the same as before, except we have the convergence condition.

Problem: Suppose (a_n) is a sequence for which $a_n \to a$. Define

$$b_n = \frac{a_1 + \cdots + a_n}{n}.$$

Prove that $b_n \to a$.

Solution: Suppose $\forall n \geq N$, we have $|a_n - a| < \frac{\varepsilon}{2}$ for some $\varepsilon > 0$. Let $M = \max\{|a_k - a| : k < N\}$. For all $n \geq \frac{2M(N-1)}{\varepsilon}$, we have

$$\begin{aligned}\left|\frac{(a_1 - a) + \cdots + (a_n - a)}{n}\right| &\leq \frac{1}{n}(|a_1 - a| + \cdots + |a_n - a|) \\ &< \frac{1}{n}\left(M(N-1) + \frac{\varepsilon}{2}(n - N)\right) \\ &= \frac{M(N-1)}{n} + \frac{\varepsilon}{2}\left(1 - \frac{N}{n}\right) \\ &< \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.\end{aligned}$$

Thus $b_n \to a$.

Problem: Let a_1, a_2 be real numbrers, and define

$$a_n := \frac{a_{n-1} + a_{n-2}}{2}.$$

Does (a_n) converge?

Solution: Using the characteristic equation, we have

$$a_n = \frac{2a_2 + a_1}{3} + \frac{4}{3}(a_2 - a_1)\left(-\frac{1}{2}\right)^n.$$

Letting $n \to \infty$ yields $a_n \to \frac{2a_2+a_1}{3}$.

2. Series

Definition (series convergence): A series converges if the sequence of its partial sums converges.

Proposition: Suppose $\sum_{i=1}^{\infty} a_i = A$ and $\sum_{i=1}^{\infty} b_i = B$.
a) $\sum_{i=1}^{\infty} (a_i + b_i) = A + B$.
b) For any $c \in \mathbb{R}$, we have $\sum_{i=1}^{\infty} c \cdot a_i = c \cdot A$.

Proof:
a) Let (s_n) be the sequence of partial sums for (a_n), and define (t_n) similarly. We have $\sum_{i=1}^{n} (a_i + b_i) = s_n + t_n$. Thus the sequence of partial sums for the sum of the series is $(s_n + t_n)$. Then limit laws imply that the partial sums converge to $A + B$.
b) Follows from the same argument as the previous bullet.

■

Proposition (divergence test): If $a_k \nrightarrow 0$, then $\sum_{i=1}^{\infty} a_i$ diverges.

Proof: We prove the contrapositive. If the sum converges, then the sequence of partial sums is Cauchy. Thus, if $\varepsilon > 0$, there exists N such that $\forall n \geq m \geq N$, we have

$$|a_m + a_{m+1} + \cdots + a_n| < \varepsilon.$$

Letting $n = m$ yields

$$|a_n| < \varepsilon,$$

which implies $a_n \to 0$. ■

Proposition (root test): Suppose $\sum_{n=1}^{\infty} a_n$ is a series. If $\rho = \limsup_{n \to \infty} |a_n|^{\frac{1}{n}}$ is less than 1, then the series converges absolutely. If it's greater than 1, then the series diverges.

Proof: Suppose $\rho < 1$. Then there exists ε such that $\rho + \varepsilon < 1$, and since ρ is the limsup of $|a_n|^{\frac{1}{n}}$, there exists some N such that every term with $n \geq N$ is less than $\rho + \varepsilon$. Thus we have

$$\sum_{n=1}^{\infty} |a_n| < \sum_{n=1}^{N-1} |a_n| + \sum_{n=N}^{\infty} (\rho + \varepsilon)^n.$$

The first sum is finite, and the second sum is a geometric series with $r < 1$, and so the sum converges. Thus the initial series absolutely converges.

If $\rho > 1$, then there exists ε such that $\rho - \varepsilon > 1$. Note that by limsup properties, there exists infinitely $|a_n|^{\frac{1}{n}}$ for which the terms are greater than $\rho - \varepsilon > 1$. Thus there's a smaller series where there are infinitely many terms with $(\rho - \varepsilon)^n$, and since $\rho - \varepsilon > 1$, the terms are unbounded, and so the initial series diverges. ■

2.1. Useful Lemmas

Here's a whole section dedicated to lemmas that can be used to bound series/help show convergence, etc.

Lemma (summation by parts):

$$\sum_{k=0}^{N} (a_{k+1} - a_k) b_k = a_{N+1} b_{N+1} - a_0 b_0 - \sum_{k=0}^{N} a_{k+1} (b_{k+1} - b_k).$$

Proof: Combine sums, cancel terms, telescope. ■

Lemma (Abel's lemma): Suppose (b_n) is a positive monotone decreasing sequence, and suppose the partial sums of (a_n) are bounded by A. Then

$$\left| \sum_{k=1}^{n} a_n b_n \right| \leq A b_1.$$

Proof: Let s_n be the partial sums of a_n. Then by summation by parts, we have

$$\begin{aligned} \left| \sum_{k=1}^{n} a_k b_k \right| &= \left| s_n b_{n+1} + \sum_{k=1}^{n} s_k (b_{k+1} - b_k) \right| \\ &\leq A b_{n+1} + A \sum_{k=1}^{n} (b_{k+1} - b_k) \\ &= A b_{n+1} + (A b_1 - A b_{n+1}) = A b_1. \end{aligned}$$

■

2.2. Riemann Rearrangement Theorem (INCOMPLETE)

Lemma: Let $\sum_{n=1}^{\infty} a_n = A$ and $\sum_{n=1}^{\infty} b_n = B$. Define an interlacing of the two sequences as the combination of the two sequences such that if a_k appears in the new sequence, the next term in this sequence that appears from (a_n) is a_{k+1}, and similarly for B. Then the sum of this interlacing series converges to $A + B$.

Proof: Let (c_n) be the interlacing. Pick N_1 such that $\forall n \geq N_1$, we have

$$\left|\sum_{k=1}^{n} a_k - A\right| < \frac{\varepsilon}{2}.$$

Define N_2 similarly for b_n. Define M_1 such that $a_{N_1} = c_{M_1}$, define M_2 similarly for b_n, and let $M = \max\{M_1, M_2\}$. Thus we have

$$\left|\sum_{k=1}^{M} c_k - A - B\right| = \left|\sum_{k=1}^{n_1} a_k - A + \sum_{k=1}^{n_2} b_k - B\right| < \left|\sum_{k=1}^{n_1} a_k - A\right| + \left|\sum_{k=1}^{n_2} b_k - B\right|.$$

Since $n_1 \geq N_1$ and $n_2 \geq N_2$ (because of our choice of M), we have

$$\left|\sum_{k=1}^{n_1} a_k - A\right| + \left|\sum_{k=1}^{n_2} b_k - B\right| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

Thus the interlaced sequence does converge to $A + B$. ■

Lemma: Suppose $\sum_{n=1}^{\infty} a_i$ is a conditionally convergent sequence. Let a_n^+ be nth positive term in the series, and define a_n^- similarly. Then we have

$$\sum_{i=1}^{\infty} a_n^+ = \infty \quad \text{and} + \sum_{i=1}^{\infty} a_n^- = -\infty.$$

Proof: If the two series were to converge to real numbers, we could take the absolute value of the negative series, and then by the previous lemma, for any interlacing, we'll get a convergent series. Since the absolute value of our initial series is an interlacing of the two, that would imply the series is absolutely convergent, which is a contradiction.

Now suppose the positive series diverges and the negative series converges to $-L$ with $L > 0$ (the opposite case is shown to be impossible similarly). For all M, there exists N such that $\forall n \geq N$, we have

$$\sum_{i=1}^{n} a_i^+ > M + L.$$

Pick N' such that a_N^+ shows up in $(a_n)_{1 \leq n \leq N'}$. Then $\forall n \geq N'$, we have

$$\sum_{i=1}^{n} a_i = \sum_{i=1}^{n_1} a_i^+ + \sum_{i=1}^{n_2} a_i^-.$$

Here we must have $n_1 \geq N$. Thus we have

$$\sum_{i=1}^{n_1} a_i^+ + \sum_{i=1}^{n_2} a_i^- > (M + L) - L = M.$$

Thus the initial series diverges, which is a contradiction. ∎

Theorem (Riemann rearrangement theorem): Suppose $\sum_{i=1}^{\infty} a_i$ is a conditionally convergent series. We can find a bijection $f : \mathbb{N} \to \mathbb{N}$ such that for any $\alpha, \beta \in \mathbb{R} \cup \{-\infty, \infty\}$ with $\alpha \leq \beta$, we have

$$\limsup_{n\to\infty} \left(\sum_{i=1}^{n} a_{f(i)} \right) = \beta \quad \text{and} \quad \liminf_{n\to\infty} \left(\sum_{i=1}^{n} a_{f(i)} \right) = \alpha.$$

Proof: Let a_n^+ be the nth positive term in the series, and define a_n^- similarly. By the previous lemma, we have $\sum_{i=1}^{\infty} a_n^+ = \infty$ and $\sum_{i=1}^{\infty} a_n^- = -\infty$. Note that $(a_n) \to 0$ since the series converges, and since (a_n^+) and (a_n^-) are subsequences, they both must also converge to 0.

We break off into cases:
a) $-\infty < \alpha \leq \beta < \infty$
b) $\beta = \infty$ and α is finite or $\alpha = -\infty$ and β is finite.
c) $\beta = \infty, \alpha = -\infty$

Part a)

Without loss of generality, supose $\beta \geq 0$ (if it wasn't, we'd just start the process of creating the rearrangement with negative terms). Let P_1 be the smallest number such that

$$\sum_{i=1}^{P_1} a_n^+ \geq \beta.$$

Let N_1 be the smallest number such that

$$\sum_{i=1}^{P_1} a_n^+ + \sum_{i=1}^{N_1} a_n^- \leq \alpha.$$

Now inductively define P_k as the smallest number such that

$$\sum_{i=1}^{P_1} a_n^+ + \sum_{i=1}^{N_1} a_n^- + \cdots + \sum_{i=P_{k-1}+1}^{P_k} a_i^+ \geq \beta$$

and define N_k to be the smallest number such that

$$\sum_{i=1}^{P_1} a_n^+ + \sum_{i=1}^{N_1} a_n^- + \cdots + \sum_{i=P_{k-1}+1}^{P_k} a_i^+ + \sum_{i=N_{k-1}+1}^{N_k} a_n^- \leq \alpha.$$

This is alawys possible, since we know the positive series and the negative series both diverge to infinities, so starting from any point through the series and adding further terms will still diverge to infinity.

Let (b_n) be the partial sums of the rearranged series, where the rearranged series is

$$a_1^+, a_2^+, ..., a_{P_1}^+, a_1^-, a_2^-, ..., a_{N_1}^-, a_{P_1+1}^+, ...$$

We prove the limsup of the series converges to β. The liminf follows similarly.

Pick $\varepsilon > 0$. There exists some M such that $\forall n \geq M$, we have $|a_n^+| < \varepsilon$. Thus this holds $\forall P_k \geq M$. By construction, we have $b_{P_1+N_1+\dots+N_{k-1}+P_k-1} \leq \beta \leq b_{P_1+N_1+\dots+N_{k-1}+P_k}$. Thus we must have $\beta \leq b_{P_1+N_1+\dots+N_{k-1}+P_k} \leq \beta + \varepsilon$ work any $P_k \geq M$. Note that again by construction, the supremum of a tail of the partial sums sequence will be $b_{P_1+N_1+\dots+N_{k-1}+P_k}$ for some k. Thus, we have for any $m \geq P_k$ (where $P_k \geq M$ for some working k), we have

$$\left|\sup_{n\geq m} b_n - \beta\right| < \varepsilon.$$

Thus we have $\limsup_{n\to\infty} b_n = \beta$. ■

2.3. Double Sums

Lemma: Supose the iterated series

$$\sum_{i=1}^{\infty}\sum_{j=1}^{\infty}|a_{ij}|$$

converges. Then

$$\sum_{j=1}^{\infty}\sum_{i=1}^{\infty}|a_{ij}|$$

converges to the same value.

Proof: First we show that the second series converges. For the sake of contradiction, suppose it doesn't. Since all the terms are positive, there are two cases in which the double doesn't converge: for some j, the single sum in i doesn't converge, or the sum over j of the single sums doesn't converge.

Suppose for some j, the single sum $\sum_{i=1}^{\infty}|a_{ij}|$ doesn't converge. Then note

$$+\infty = \sum_{i=1}^{\infty}|a_{ij}| \leq \sum_{i=1}^{\infty}\left(\sum_{j=1}^{\infty}|a_{ij}|\right),$$

but this contradicts the first double sum converging.

For the second case, let $p_j := \sum_{i=1}^{\infty}|a_{ij}|$. Then $\sum_{j=1}^{\infty} p_j$ doesn't converge. Fix large $M \geq 0$. Then by definition, there exists N_M for which

$$\sum_{j=1}^{N_M} p_j \geq M.$$

Since there are finitely many p_j in this sum, there exists J for which

$$\sum_{i=1}^{J} |a_{ij}| > p_j - \frac{\varepsilon}{N_M}$$

for all $1 \leq j \leq N_M$. Thus we have

$$M \leq \sum_{j=1}^{N_M} \sum_{i=1}^{J} |a_{ij}| + \varepsilon = \sum_{i=1}^{J} \sum_{j=1}^{N_M} |a_{ij}| + \varepsilon.$$

Since the first iterated series converges, the inner sums are bounded by their infinite sum value, so the right side is at most $\sum_{i=1}^{J} \sum_{j=1}^{\infty} |a_{ij}| + \varepsilon$. This implies that the first double sum gets arbitrarily large (we can pick $\varepsilon = \frac{1}{2}$ for concreteness), since M was arbitrary, so the first double sum cannot converge, contradiction.

Now we prove they converge to the same value. Define $b_i := \sum_{j=1}^{\infty} |a_{ij}|$ and $c_j := \sum_{i=1}^{\infty} |a_{ij}|$. Then

$$\sum_{i=1}^{\infty} \sum_{j=1}^{\infty} |a_{ij}| = \sum_{i=1}^{\infty} b_i = S_1 \quad \text{and} \quad \sum_{j=1}^{\infty} \sum_{i=1}^{\infty} |a_{ij}| = \sum_{i=1}^{\infty} c_j = S_2.$$

Pick $\varepsilon > 0$. Then there exists N for which $S_1 + \varepsilon > \sum_{i=1}^{N} b_i > S_1 - \varepsilon$ and $S_2 + \varepsilon > \sum_{j=1}^{N} c_j > S_2 - \varepsilon$. Then, since we're only dealing with a finite amount of infinite sums (namely b_i, c_j for $i, j \leq N$), there exists M for which $b_i + \frac{\varepsilon}{N} > \sum_{i=1}^{M} |a_{ij}| > b_i - \frac{\varepsilon}{N}$ and $c_j + \frac{\varepsilon}{N} > \sum_{j=1}^{M} |a_{ij}| > c_j - \frac{\varepsilon}{N}$. Plugging these in to the first inequalities yields

$$S_1 + 2\varepsilon > \sum_{i=1}^{N} \sum_{j=1}^{M} |a_{ij}| > S_1 - 2\varepsilon \quad \text{and} \quad S_2 + 2\varepsilon > \sum_{j=1}^{N} \sum_{i=1}^{M} |a_{ij}| > S_2 - 2\varepsilon.$$

Now let $P = \max\{M, N\}$. Note that each double sum is bounded above by their corresponding value, so increasing both upper indices to P will still keep both double sums bounded, yielding

$$S_1 \geq \sum_{i=1}^{P} \sum_{j=1}^{P} |a_{ij}| > S_1 - 2\varepsilon \quad \text{and} \quad S_2 \geq \sum_{j=1}^{P} \sum_{i=1}^{P} |a_{ij}| > S_2 - 2\varepsilon.$$

Now suppose for the sake of contradiction that $S_1 \neq S_2$. Then letting $\varepsilon = |S_1 - S_2|/2$ would yield a contradiction, since it would implie $\sum_{i=1}^{P} \sum_{j=1}^{P} |a_{ij}| > \sum_{j=1}^{P} \sum_{i=1}^{P} |a_{ij}|$ or vice versa.■

Theorem (Fubini's theorem for sums): Suppose

$$\sum_{i=1}^{\infty} \sum_{j=1}^{\infty} |a_{ij}|$$

converges. Then both $\sum_{i=1}^{\infty} \sum_{j=1}^{\infty} a_{ij}$ and $\sum_{j=1}^{\infty} \sum_{i=1}^{\infty} a_{ij}$ converge, and

$$\lim_{n \to \infty} s_{nn} = \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} a_{ij} = \sum_{j=1}^{\infty} \sum_{i=1}^{\infty} a_{ij},$$

where $s_{nn} = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij}$.

Proof: From the previous lemma, we know that both double sums converge, so we just need to prove the second equation. First we show the limit exists.

Let $t_{mn} = \sum_{i=1}^{m} \sum_{j=1}^{n} |a_{ij}|$. Note that (t_{nn}) is increasing and bounded by $\sum_{i=1}^{\infty} \sum_{j=1}^{\infty} |a_{ij}|$, and so converges. Thus the sequence is Cauchy. Then for $n \geq m$, we have

$$|s_{nn} - s_{mm}| \leq \sum_{i=m+1}^{n} \sum_{j=1}^{n} |a_{ij}| + \sum_{i=1}^{n} \sum_{j=m+1}^{n} |a_{ij}| + \sum_{i=m+1}^{n} \sum_{j=m+1}^{n} |a_{ij}| = |t_{nn} - t_{mm}|.$$

Since (t_{nn}) is Cauchy, we can make the right arbitrarily small for large enough n, m, so (s_{nn}) is also Cauchy.

Now let $\lim_{n\to\infty} s_{nn} = S$. We need to show that S equals the double sums. We only show it's equal to the first, as the second follows similarly. We have

$$|s_{mn} - S| \leq |s_{mn} - s_{nn}| + |s_{nn} - S|.$$

For the first term, assuming without loss of generality that $n \geq m$, we have

$$\begin{aligned} |s_{mn} - s_{nn}| \leq \sum_{i=m+1}^{n} \sum_{j=1}^{n} |a_{ij}| &\leq \sum_{i=m+1}^{n} \sum_{j=1}^{n} |a_{ij}| + \sum_{i=1}^{n} \sum_{j=m+1}^{n} |a_{ij}| + \sum_{i=m+1}^{n} \sum_{j=m+1}^{n} |a_{ij}| \\ &= |t_{nn} - t_{mm}|. \end{aligned}$$

Thus

$$|s_{mn} - S| \leq |t_{nn} - t_{mm}| + |s_{nn} - S|.$$

Since (t_{nn}) is Cauchy, and since $s_{nn} \to S$, there exists N for which $n, m \geq N$ implies both terms are less than $\frac{\varepsilon}{2}$. Thus

$$|s_{mn} - S| < \varepsilon$$

for all $n, m \geq N$. Letting $n \to \infty$, then $m \to \infty$ (which we can do since we know the iterated series converges) yields

$$\left| \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} a_{ij} - S \right| < \varepsilon.$$

Since ε is arbitrary, the two are equal. ■

2.4. Problems

Problem (Abel's test): Let $\{x_n\}$ and $\{w_n\}$ be two sequences of reals such that

- the sequence of partial sums $\{s_n\}$ of $\sum x_n$ is bounded
- $\lim_{n\to\infty} w_n = 0$
- $\sum |w_{n+1} - w_n|$ converges

Prove that $\sum w_n x_n$ converges.

Solution: Suppose the partial sums are bounded by M. Pick $\varepsilon > 0$. Applying summation by parts, we obtain

$$\sum_{k=m}^{n} w_k x_k = s_n w_n - s_{m-1} w_m + \sum_{k=m}^{n-1} s_k (w_k - w_{k+1}) \leq M\left(|w_n| + |w_m| + \sum_{k=m}^{n-1} |w_k - w_{k+1}| \right).$$

The second and third bullet point guarantee there exists some N such that $m, n \geq N \Rightarrow |w_n|, |w_m| < \frac{\varepsilon}{3M}$ and $\sum_{k=m}^{n-1} |w_k - w_{k+1}| < \frac{\varepsilon}{3M}$. Thus, there exists some N such that $n, m \geq N$ implies

$$\sum_{k=m}^{n} w_k x_k < \varepsilon.$$

Thus $\sum w_k x_k$ is Cauchy and converges.

Problem: Suppose $\sum x_n$ is an absolutely convergent series. Show that any rearrangement of x_n is also absolutely convergent, and the rearranged series converges to the same value as $\sum x_n$.

Solution: Let $f : \mathbb{N} \to \mathbb{N}$ be a bijection. We want to show that $\sum_{k=1}^{\infty} |x_{f(k)}|$ converges. By absolute convergence, there exists N such that $n \geq m \geq N \Rightarrow \sum_{k=m}^{n} |x_k| < \varepsilon$. We can also see that there exists N' such that $n \geq N' \Rightarrow f(n) \geq N$ (in particular, $N' = \max\{f^{-1}(k) : 1 \leq k \leq N\}$). Then, for all $n \geq m \geq N'$, we have

$$\sum_{k=m}^{n} |x_{f(k)}| < \sum_{k=N}^{\max_{m \leq i \leq n} f(i)} |x_k| < \varepsilon.$$

Thus the series is Cauchy, so it converges.

Now consider $\left|\sum_{k=1}^{\infty} x_k - x_{f(k)}\right|$. Let $M = \max\{N, N'\}$. Then clearly there exists M' such that $\{f(k) : 1 \leq k \leq M'\}$ contains all integers in $[1, M]$. Note also that we clearly have $M' \geq M$. Its clear then that for any $n \geq M'$, the nth partial sum will have the terms x_k for $1 \leq k \leq M$ cancel. Then all that's left is a subset of both sequences that start from at least index M. But then using the triangle inequality, we're left with subsets of both sequences that are Cauchy, since $M \geq N, N'$. Thus

$$\left|\sum_{k=1}^{n} x_k - x_{f(k)}\right| < 2\varepsilon$$

for all $n \geq M'$. This implies that $\left|\sum_{n=1}^{\infty} x_k - \sum_{k=1}^{\infty} x_{f(k)}\right| < 2\varepsilon$. Since ε was arbitrary, the two series must converge to the same value.

3. The Topology of $\mathbb{R}$

Definition (open set): A set $U \subseteq \mathbb{R}$ is *open* if for every $x \in U$, there is a number $\delta > 0$ such that

$$(x - \delta, x + \delta) \subseteq U.$$

This is called the δ neighborhood of x, and is denoted $V_\delta(x)$.

Proposition:
a) If $\{U_\alpha\}$ is a collection of open sets, then $\bigcup_\alpha U_\alpha$ is also an open set.
b) If $\{U_\alpha\}$ is a finite collection of open sets, then $\bigcap_\alpha U_\alpha$ is an open set.

Proof:
a) Consider some $x \in \bigcup_\alpha U_\alpha$. Then $x \in U_i$ for some i. Thus for some δ, $V_\delta(x) \subseteq U_i$. Thus $V_\delta(x) \subseteq \bigcup_\alpha U_\alpha$, so $\bigcup_\alpha U_\alpha$ is open.
b) Consider some $x \in \bigcap_\alpha U_\alpha$. For each U_i in $\{U_\alpha\}$, there exists δ_i such that $V_{\delta_i}(x) \subseteq U_i$. Let $\delta = \min\{\delta_\alpha\}$. Then $V_\delta(x) \subseteq V_{\delta_i}(x) \subseteq U_i$. Thus $V_\delta(x) \subseteq \bigcap_\alpha U_\alpha$.

■

Theorem: Every open set is a countable union of disjoint open intervals.

Proof: Let A be an open set. For $x \in A$, let $I_x = (\alpha, \beta)$, where $\alpha = \inf\{a : (a, x) \subseteq A\}$ and $\beta = \sup\{b : (x, b) \subseteq A\}$. For any x, y, we must have $I_x = I_y$ or $I_x \cap I_y = \emptyset$, because if they overlap but aren't equal, then you could extend one of them, contradicting us choosing the largest possible interval.

We claim that these intervals make up A. Note that for every $x \in A$ we have $x \in I_x \subseteq A$, so the union of all the intervals is A. Further, because $\mathbb{Q}$ is dense if $\mathbb{R}$, every open interval contains a rational number, so there cannot be more intervals than rationals. Thus the number of intervals is countable. ■

Definition (closed set): A set $A \subseteq \mathbb{R}$ is *closed* if A^c is open.

Definition (limit point): A point x is a limit point of a set A if there is a sequence of points $a_1, a_2 \ldots$ from $A \backslash \{x\}$ such that $a_n \to x$.

Theorem: A set is closed if and only if it contains all its limit points.

Proof: First we show that if a set is closed, then it contains all its limit points. We proceed by contradiction. Let x be a limit point not in A. Then we have the following:

- There exists a sequence (a_n) with each term in A such that $\lim_{n\to\infty} a_n = x$.
- $x \in A^c$, which is an open set, so there exists δ such that $V_\delta(x) \subseteq A^c$.

Since the sequence converges to x, we must have $|a_n - x| < \delta \Rightarrow x - \delta < a_n < x + \delta$ for all $n \geq N$ for some N. Thus implies $a_n \in V_\delta(x)$ for all $n \geq N$. However, this is impossible, since $a_n \in A$, while $V_\delta(x) \subseteq A^c$. Thus we have a contradiction.

Now we prove the other by contrapositive, that is we prove that if a set is not closed, then it doesn't contain all its limit points. Suppose A is not closed. Then A^c is not open. Thus, there exists some $x \in A^c$ such that every δ neighborhood of x contains some element not in A^c, which is equivalent to it containing an element in A. Let a_n be an element in A that is contained in the $\frac{1}{n}$ neighborhood of x. We claim $\lim_{n\to\infty} a_n = x$, which proves the claim.

Let $\varepsilon > 0$, and pick some integer k such that $\frac{1}{k} < \varepsilon$. Then we have $|a_k - x| < \frac{1}{k} < \varepsilon$ by definition. Note that this implies $|a_n - x| < \frac{1}{k} < \varepsilon$ for all $n \geq k$, since if a_n is in a $\frac{1}{n}$ neighborhood of x, then it's also in a $\frac{1}{k}$ neighborhood of x, which is further in an ε neighborhood of x. Thus a_n converges to x. ■

Proposition (closure): Let A bet a set, and let L be the set of all the limit points of A. Then *closure* of A is $\overline{A} = A \cup L$.

Example: If $A = (0, 1)$, then $L = [0, 1]$, so $\overline{A} = [0, 1]$. Basically all the closure does it add boundary points not already in a set.

Proposition:

- If $\{U_\alpha\}$ is a finite collection of closed sets, then $\bigcup_\alpha U_\alpha$ is also a closed set.
- If $\{U_\alpha\}$ is a collection of closed sets, then $\bigcap_\alpha U_\alpha$ is also a closed set.

Proof: Follows from the union/intersection proposition of open sets and De Morgan's laws. ■

3.1. Heine-Borel Theorem

Definition (cover): Let A be a set. The collection of sets $\{U_\alpha\}$ are a *cover* of A if

$$A \subseteq \bigcup_\alpha U_\alpha.$$

If each U_α is open, then $\{U_\alpha\}$ is an *open cover* of A. If a finite subset of $\{U_\alpha\}$ is a cover of A, then that subset is a finite subcover of A.

Definition (compact): A set A is *compact* if every open cover of A contains a finite subcover of A.

Theorem (Heine-Borel theorem): A set $S \subseteq \mathbb{R}$ is compact if and only if S is closed and bounded.

Proof: Suppose S is compact. Then for every open cover of S, there exists a finite subcover. Let $I_n = (-n, n)$. Clearly the set $\{I_n\}_{n \geq 1}$ is an open cover of S. Thus, there exists a sequence $n_1, ..., n_k$ such that $\{I_{n_1}, ..., I_{n_k}\}$ is an open cover of S. WLOG $n_1 < n_2 < \cdots < n_k$. We have

$$S \subseteq \bigcup_{j=1}^{k} I_{n_j} = I_{n_k}.$$

Since I_{n_k} is bounded, then clearly S is bounded.

Next we show that S is closed by contradiction. Suppose S is compact and doesn't contain all its limit points. That is, there exists a sequence (a_n) contained in S that converges to some point x not in S. Let $I_n = \left(-\infty, x - \frac{1}{n}\right) \cup \left(x + \frac{1}{n}, \infty\right)$. Clearly

$$S \subseteq \bigcup_{k=1}^{\infty} I_k = \mathbb{R} \backslash \{x\}.$$

Suppose $\{I_{n_1}, ..., I_{n_k}\}$ is a finite subcover, where the indexes are increasing. Then $\cup_{j=1}^{k} I_{n_j} = I_{n_k}$. Thus

$$S \subseteq I_{n_k} = \left(-\infty, x - \frac{1}{n_k}\right) \cup \left(x + \frac{1}{n_k}, \infty\right),$$

which implies

$$S \cap \left(x - \frac{1}{n_k}, x + \frac{1}{n_k}\right) = \emptyset.$$

However this is a contradiction, since the equation implies the sequence cannot converge to x without containing elements outside of S, contradiction.

Now we prove the other direction. Suppose S is closed and bounded. For every $x \in \mathbb{R}$, define $S_x = S \cap (-\infty, x]$. Suppose $\mathcal{F}$ is an open cover of S, and let

$$B = \{x : \mathcal{F} \text{ contains a finite subcover of } S_x\}.$$

Note that since S is bounded, $M = \sup(S)$ and $L = \inf(S)$ exist. We also know that there exist sequences that converge to both, so they're limit points. Thus, since S is closed, we have $L, M \in S$.

We want to show that $M \in B$, since $S_M = S$. Note that we already have $L \in B$, since the cover $\mathcal{F}$ must contain some set that covers L, just take that set as the subcover.

Assume for the sake of contradiction that $M \notin B$. Then clearly we can't have $x \in B$ for any $x \geq M$, since otherwise we would get a finite subcover for $S \cap (-\infty, x]$ as well. Thus $x < M$ for all $x \in B$, which implies B is bounded from above. Since B is nonempty, we can then let

$T = \sup(B)$. Note also that B contains infinitely elements, since for all $x \in B$, any number less than x is also in b, and if $\sup(B) = L$, then we can show by a similar argument as for the first case (next paragraph) that this is impossible. Since B has infinitely many elements, this implies that for all $\varepsilon > 0$, there's some element $b \in B$ such that $T - \varepsilon < b < T$.

We have two cases:

Case 1: $T \in S$

Since $\mathcal{F}$ covers S, some open set in it contains T, call it U. Pick $\delta = \min\{\delta_1, \delta_2\}$, where $T + \delta_1 < M$ and $V_{\delta_2}(x) \subseteq U$. Thus we have $\left(T - \delta, T + \frac{\delta}{2}\right] \subseteq U$.

Note that $T - \delta \in B$, since if not, then we can't have $T \in B$ via the same argument we made to show that $x < M$ for all $x \in B$. Thus, there exists some finite subcover F of $\mathcal{F}$ that covers $S_{T-\delta}$. However, note that this implies $F \cup \left(T - \delta, T + \frac{\delta}{2}\right)$ covers $S_{T+\frac{\delta}{2}}$, which contradicts $T = \sup(B)$. Thus we have a contradiction in this case.

Case 2: $T \notin S$

Since $T \notin S$ and S is closed, $T \in S^c$, which is an open set. Pick δ so that $V_\delta(T) \subseteq S^c$. Thus $\left[T - \frac{\delta}{2}, T + \frac{\delta}{2}\right] \cap S = \emptyset$, which implies $S \cap \left(-\infty, T - \frac{\delta}{2}\right] = S \cap \left(-\infty, T + \frac{\delta}{2}\right]$.

Note we showed that for all $\varepsilon > 0$, there exists $b \in B$ such that $T - \varepsilon < b < T$. Thus, picking $\varepsilon = \frac{\delta}{2}$, we have that $T - \frac{\delta}{2} < a \in B$, which again by an argument made earlier implies that $T - \frac{\delta}{2} \in B$. Thus there's some finite subcover of $S_{T-\frac{\delta}{2}} = S \cap \left(-\infty, T - \frac{\delta}{2}\right]$. However, we also showed that $S_{T-\frac{\delta}{2}} = S_{T+\frac{\delta}{2}}$, so this same subcover works for this set. This implies that $T + \frac{\delta}{2} \in B$, which contradicts $T = \sup(B)$, so we again have a contradiction ■

Theorem (Heine-Borel expanded): Suppose $A \subseteq \mathbb{R}$. The following are equivalent:
a) A is compact.
b) A is closed and bounded.
c) If (a_n) is a sequence of numbers in A, then there is a subsequence $\left(a_{n_k}\right)$ that converges to a point in A.

Proof: The equivalence of a) and b) was the last theorem. Suppose A is closed and bounded. Then any sequence coming from A is bounded, and so has a convergent subsequence by Bolzano-Weierstrass. The limit of this subsequence is clearly a limit point of A, and since A is closed, it must be contained in A.

If A is not closed, then there's some limit point of A not in A. Let (a_n) be a sequence that converges to this limit point. Then every subsequence must also converge to that limit point, which again is not in A.

If A is not bounded, then we can create an unbounded sequence. Just let a_k be some element of A that is greater than k, which must exist since A is unbounded. Clearly every subsequence of (a_n) also diverges. This establishes the equivalence of b) and c). ■

3.2. Problems

Problem: Construct a set whose set of limit points is $\mathbb{Z}$.

Solution: Let $A_k = \left\{k + \frac{1}{2}, k + \frac{1}{3}, k + \frac{1}{4}, ...\right\}$ for all $k \in \mathbb{Z}$. We claim

$$A = \bigcup_{k=-\infty}^{\infty} A_k$$

has $\mathbb{Z}$ as its set of limit points. First note that for each $k \in \mathbb{Z}$, the sequence $a_n = k + \frac{1}{n}$ for $n \geq 2$ converges to k. Thus $\mathbb{Z}$ is a subset of the set of limit points of A.

Now consider some non-integer α. Note that $\{\alpha\} \in \left(\frac{1}{n+1}, \frac{1}{n}\right]$ for some integer n. Then the closest α gets to any element of A is $\min\left\{\left|\{\alpha\} - \frac{1}{n}\right|, \left|\{\alpha\} - \frac{1}{n+1}\right|\right\}$, where each option corresponds to $\lfloor\alpha\rfloor + \frac{1}{n}$ and $\lfloor\alpha\rfloor\frac{1}{n+1}$ respectively. Thus, we can't get arbitrarily close to any non integer α (choose $\varepsilon = \min\left\{\left|\{\alpha\} - \frac{1}{n}\right|, \left|\{\alpha\} - \frac{1}{n+1}\right|\right\}$), so α is not a limit point.

Problem: Prove that the set of limit points of a set is closed.

Solution: We show the complement is open. Let A be our set, and let L be the set of limit points of A. Consider some $x \in L^c$. Since x is not a limit point, this implies that for every sequence $(a_n) \in A \setminus \{x\}$, there exists $\varepsilon > 0$ such that for all N, there exists some $n \geq N$ such that $|a_n - x| \geq \varepsilon$. Note that this inequality implies $\left|a_n - \left(x + \frac{\varepsilon}{2}\right)\right| \geq \frac{\varepsilon}{2}$ and $\left|a_n - \left(x - \frac{\varepsilon}{2}\right)\right| \geq \frac{\varepsilon}{2}$ (we get these by splitting the inequalities into cases based on if the inside if the absolute value is positive or negative). Thus, no sequence in $A \setminus \{x\}$ converges to $x + \frac{\varepsilon}{2}$ or $x - \frac{\varepsilon}{2}$. To show that these are not limit points, we need to show that the sequences can come from $A \setminus \left\{x \pm \frac{\varepsilon}{2}\right\}$. However, this isn't an issue. Since any sequence doesn't converge to those values, the sequences can only contain finitely many terms that are $x \pm \frac{\varepsilon}{2}$, so removing those won't affect the convergence. Similarly, adding in any amount of terms equal to x to the sequences won't change the convergence.

Thus we've showed $x \pm \frac{\varepsilon}{2} \in L^c$. In fact, for any $\delta < \frac{\varepsilon}{2}$, we can show that $x \pm \delta \in L^c$ by a similar method to what we did above. Thus we have $V_{\frac{\varepsilon}{2}}(x) \in L^c$ for any $x \in L^c$, which implies L^c is open, which means L is closed, as desired.

Definition (interior, exterior, boundary):

The *interior* of a set A, denoted $\text{Int}(A)$, is the set of points x such that there is an open neighborhood of x that is a subset of A.

The *exterior* of a set A, denoted $\text{Ext}(A)$, is the set of points x such that there is an open of x that is a subset is a subset of A^c.

The *boundary* of set A, denoted ∂A, is the set of points X such that every neighborhood of x contains points in A and A^c.

Problem: Prove that for any set A, we have $\mathbb{R} = \mathrm{Int}(A) \cup \partial A \cup \mathrm{Ext}(A)$, and is a disjoint union.

Solution: It's clear that $\mathrm{Int}(A) \cap \mathrm{Ext}(A) = \emptyset$. Similarly, $\mathrm{Int}(A) \cap \partial A = \emptyset$ and $\mathrm{Ext}(A) \cap \partial A = \emptyset$. Now consider some $x \in \mathbb{R}$, and suppose it's not in $\mathrm{Int}(A)$ or $\mathrm{Int}(B)$. Then that implies that every open neighborhood of x contains points in both A and A^c, which means $x \in \partial A$. We can similarly show that if x is not in two of the sets, then it must be in the other one. Thus, we have $\mathbb{R} = \mathrm{Int}(A) \cup \partial A \cup \mathrm{Ext}(A)$.

4. Continuity

4.1. Functional Limits

Definition (functional limit): Let $f : A \to \mathbb{R}$ and let c be a limit point of A. Then

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

if for all $\varepsilon > 0$ there exists some $\delta > 0$ such that for every $x \in A$ for which $0 < |x - c| < \delta$, we have

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

For the one sided limit $\lim_{x \to c^+}$, the condition on x is relaxed to $c < x < c + \delta$. Similarly, for the one sided limit $\lim_{x \to c^-}$, the condition on x is relaxed to $c - \delta < x < c$.

Proposition: A limit $\lim_{x \to c} f(x)$ can converge to at most one value.

Proof: Suppose $\varepsilon > 0$. Then there exists δ_1 such that when $0 < |x - c| < \delta_1$, we have $|f(x) - L_1| < \frac{\varepsilon}{2}$. There also exists δ_2 such that $|L_2 - f(x)| < \frac{\varepsilon}{2}$. Let $\delta = \min\{\delta_1, \delta_2\}$. Then we have, for all $0 < |x - c| < \delta$, we have

$$|L_2 - L_1| = |(L_2 - f(x)) + (f(x) - L_1)| \leq |L_2 - f(x)| + |f(x) - L_1| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

Since this holds for all $\varepsilon > 0$, this implies $L_2 - L_1 = 0$, as desired. ■

Proposition: $\lim_{x \to x} f(x) = L$ if and only if $\lim_{x \to c^+} f(x) = \lim_{x \to c^-} f(x) = L$.

Proof: Suppose $\lim_{x \to c^+} f(x) = \lim_{x \to c^-} f(x) = L$. Thus, for all $\varepsilon > 0$, there exist $\delta_1, \delta_2 > 0$ such that

$$|f(x) - L| < \varepsilon \quad \text{when} \quad c < x < c + \delta_1$$

and

$$|f(x) - L| < \varepsilon \quad \text{when} \quad c - \delta_2 < x < c.$$

Let $\delta = \min\{\delta_1, \delta_2\}$. Then we have

$$|f(x) - L| < \varepsilon \quad \text{when} \quad 0 < |x - c| < \delta,$$

which implies $\lim_{x \to c} f(x) = L$.

Now suppose $\lim_{x \to c} f(x) = L$. Then for all $\varepsilon > 0$, there exists $\delta > 0$ such that

$$|f(x) - L| < \varepsilon \quad \text{when} \quad 0 < |x - c| < \delta.$$

This implies

$$|f(x) - L| < \varepsilon \quad \text{when} \quad c < x < c + \delta$$

and

$$|f(x) - L| < \varepsilon \quad \text{when} \quad c - \delta < x < c,$$

which implies that both one-sided limits are equal to L. ∎

Theorem: Assume $A \subseteq \mathbb{R}$, $f : A \to \mathbb{R}$, and c is a limit point of A. Then $\lim_{x \to c} f(x) = L$ if and only if, for every sequence a_n from A for which $a_n \neq c$ and $a_n \to c$, we have $f(a_n) \to L$.

Proof: We assume that $a_n, x \neq c$.

First suppose $\lim_{x \to c} f(x) = L$. Then for all $\varepsilon > 0$, there exists δ such that

$$0 < |x - c| < \delta \Rightarrow |f(x) - L| < \varepsilon$$

for $x \in A$. Let (a_n) be an arbitrary sequence in A converging to c. Then, there exists N such that for all $n \geq N$, we have $|a_n - c| < \delta$. This implies $|f(a_n) - L| < \varepsilon$ for all $n \geq N$, which shows that $f(a_n) \to L$.

Now supose $\lim_{x \to c} f(x) \neq L$. That is, there exists $\varepsilon > 0$ such that for all $\delta > 0$ there exists $x \in A$ such that $0 < |x - c| < \delta \Rightarrow |f(x) - L| \geq \varepsilon$. In particular, setting $\delta_n = \frac{1}{n}$, there always exists x_n within $0 < |x - c| < \delta_n$ such that $|f(x) - L| \geq \varepsilon$. Clearly $x_n \to c$, while $f(x_n) \nrightarrow L$, so we're done. ∎

Proposition (limit laws): Let f and g be functions from $A \subseteq \mathbb{R}$ to $\mathbb{R}$ and let c be a limit point of A. Assume $\lim_{x \to c} f(x) = L$ and $\lim_{x \to c} g(x) = M$. Then

- $\lim_{x \to c} [k \cdot f(x)] = k \cdot L$
- $\lim_{x \to c} [f(x) + g(x)] = L + M$
- $\lim_{x \to c} [f(x) \cdot g(x)] = L \cdot M$
- $\lim_{x \to c} \frac{f(x)}{g(x)} = \frac{L}{M}$ if $M \neq 0$ and $g(x) \neq 0$ for any $x \in A$.

Proof: The limits holds for all sequences converging to c, and these laws apply to sequences, so in turn these laws hold for limits. ∎

Theorem (squeeze theorem): Let f, g, h be functions from $A \subseteq \mathbb{R}$ to $\mathbb{R}$, let c be the limit point of A, suppose

$$f(x) \leq g(x) \leq h(x)$$

for all $x \in A$, and suppose

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

Then

$$\lim_{x \to c} g(x) = L.$$

Proof: Same reasoning as last proposition. ■

4.2. Continuity

Definition (continuity): A function $f : A \to \mathbb{R}$ is *continuous at a point* $c \in A$ if for all $\varepsilon > 0$, there exists some $\delta > 0$ such that for all $x \in A$ where $|x - c| < \delta$ we have

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

If f is continuous at every point in its domain, then f is called *continuous.*

Remark: Note that if $c \in A$ is not a limit point of A, then it's automatically continuous, since we can pick δ so that $|x - c| < \delta$ contains no values in $A \setminus \{x\}$. Thus the condition $|f(x) - f(c)| < \varepsilon$ is vacuosly true.

Proposition: Let $f : A \to \mathbb{R}$ and $c \in A$. Then the follwing are equivalent:

a) f is continuous at c.
b) For all $\varepsilon > 0$ there exists some $\delta > 0$ such that for all $x \in A$ where $|x - c| < \delta$ we have $|f(x) - f(c)| < \varepsilon$.
c) For any ε-neighborhood of $f(c)$, denoted $V_\varepsilon(f(c))$, there exists some δ neighborhood of c, denoted $V_\delta(c)$, with the property that for any $x \in A$ for which $x \in V_\delta(c)$, we have $f(x) \in V_\varepsilon(f(c))$.
d) For all sequences $(a_n) \in A$ converging to c, we have $f(a_n) \to f(c)$.
e) $\lim_{x \to c} f(x) = f(c)$ if c is a limit point of A.

Proof: a) is equivalent to b) by definition. b) is equivalent to c), just rephrased in term of neighborhoods. The proof that a) is equivalent to d) is basically identical to the proof of sequences converging to c converge to $\lim_{x \to} f(x)$ under f. d) is equivalent to e) using that same theorem.
■

Proposition (continuity laws): Let $f : A \to \mathbb{R}$ and $g : A \to \mathbb{R}$ be continuous at c, and let $c \in A$. Then the follwing are true:

a) $k \cdot f(x)$ is continuous at c for all $k \in \mathbb{R}$.
b) $f(x) + g(x)$ is continuous at c.
c) $f(x) \cdot g(x)$ is continuous at c.
d) $\frac{f(x)}{g(x)}$ is continuous at c, provided $g(x) \neq 0$ for all $x \in A$.

Proof: By the previous proposition, we can rephrase these as limits, and then we apply our limit laws. ■

Problem (continuous compositions): Suppose $A, B \subseteq \mathbb{R}$, $g : A \to B$ and $f : B \to \mathbb{R}$. If g is continuous at $c \in A$, and f is continuous at $g(c) \in B$, then $f \circ g : A \to \mathbb{R}$ is continuous at c.

Proof: Consider an arbitrary sequence (a_n) from A converging to c. Then by continuity we have that $g(a_n) \to g(c)$. Note that $(g(a_n))$ is a sequence in B converging to $f(c)$, so again by continuity we have $f(g(a_n)) \to f(g(c))$. Since (a_n) was arbitrary, this holds for any sequence converging to c. Thus, $f \circ g$ is continuous at c. ■

4.3. Topological Continuity

Definition (pre-image): Let $X, Y \subseteq \mathbb{R}$ and $f : X \to Y$. For $B \subseteq Y$, define the *pre-image* (or *inverse*)

$$f^{-1}(B) = \{x \in X : f(x) \in B\}.$$

Example: Define $f : [2, 10] \to \mathbb{R}$ as $f(x) = 5x$. Letting $B = (1, 20)$, we have

$$f^{-1}(B) = [2, 4) = \left(\frac{1}{5}, 4\right) \cup [2, 10].$$

Theorem: Let $f : X \to \mathbb{R}$. Then f is continuous if and only if for every open set B, we have $f^{-1}(B) = A \cap X$ for some open set A.

Proof: Let f be continuous and let B be an open set in $\mathbb{R}$. We want to show that for any $x_0 \in f^{-1}(B)$, there exists δ such that if $x \in X \cap V_\delta(x)$, then $x \in f^{-1}(B)$. To that end, suppose $x_0 \in f^{-1}(B)$. This implies $f(x_0) \in B$. Thus, there exists ε such that $V_\varepsilon(f(x_0)) \subseteq B$ (since B is open). By continuity, this implies that there exists δ such that if $x \in X$ and $x \in V_\delta(x_0)$, then $f(x) \in V_\varepsilon(f(x_0))$. In particular, this implies that $x \in f^{-1}(V_\varepsilon(f(x_0))) \subseteq f^{-1}(B)$, as desired.

Now suppose for every open set B, $f^{-1}(B) = A \cap X$ for some open set A. Pick $x_0 \in X$ and $\varepsilon > 0$. Note that $V_\varepsilon(f(x_0))$ is open, so we have that $f^{-1}(V_\varepsilon(f(x_0))) = A \cap X$ for some open set

A. Clearly $x_0 \in A \cap X = f^{-1}(V_\varepsilon(f(x_0)))$, and since A is open, there exists some δ such that $V_\delta(x_0) \subseteq A$. In particular, if $x \in X \cap V_\delta(x_0)$, then

$$x \in X \cap V_\delta(x_0) \subseteq X \cap A = f^{-1}(V_\varepsilon(f(x_0))) \Rightarrow f(x) \in V_\varepsilon(f(x_0)).$$

Thus f is continuous, as desired. ■

Remark: Note that this condition guarantees continuity at every point in the domain.

4.4. The Extreme Value Theorem

Proposition: Suppose $f : X \to \mathbb{R}$ is continuous. If $A \subseteq X$ is compact, then $f(A)$ is compact.

Proof: Suppose $\{U_\alpha\}$ is an open cover of $f(A)$. Consider $\{f^{-1}(U_\alpha)\}$. We have $f^{-1}(U_\alpha) = X \cap V_\alpha$ for some open set V_α by the previous proposition. We show that $\{V_\alpha\}$ is an open cover of A. Consider $x_0 \in A$. Then $f(x_0) \in f(A)$, which means $f(x_0) \in U_i$ for some i, which implies $x_0 \in f^{-1}(U_i) = X \cap V_i \Rightarrow x_0 \in V_i$. Thus $\{V_\alpha\}$ is indeed an open cover of A.

Since A is compact, there exists some finite subcover $\{V_1, V_2, ..., V_k\}$. We claim that $\{U_1, U_2, ..., U_k\}$ is a finite subcover of $f(A)$, where V_i corresponds to U_i through $f^{-1}(U_i) = X \cap V_i$. Consider $y_0 \in f(A)$. Then $f(x_0) = y_0$ for some x_0. Thus $x_0 \in V_i$ for some i, since $\{V_\alpha\}$ is an finite subcover of A. However, $x_0 \in X$, which implies $x_0 \in X \cap V_i = f^{-1}(U_i)$. This implies $y_0 = f(x_0) \in U_i$, as desired. ■

Corollary: A continuous function on a compact set is bounded.

Proof: Suppose $f : A \to \mathbb{R}$ is continuous and A is compact. By the previous proposition, $f(A)$ is compact, which means $f(A)$ is bounded. ■

Theorem (extreme value theorem): A continuous function on a compact set attains a maximum and a minimum.

Proof: Suppose $f : A \to \mathbb{R}$ is continuous and A is compact, and let $M = \sup\{f(x) : x \in A\}$ and $L = \inf\{f(x) : x \in A\}$. These exist since $f(A)$ is bounded by the corollary. Note that since $f(A)$ is compact, it's closed and thus contains all its limit points. Since M and L are limit points of $f(A)$, they must both be in $f(A)$. Thus, there exists x_1 and x_2 such that $f(x_1) = M$, $f(x_2) = L$, as desired. ■

4.5. The Intermediate Value Theorem

Lemma: If f is continuous and $f(c) > 0$, then there exists a δ such that $x \in (c-\delta, c+\delta) \Rightarrow f(x) > 0$. Likewise, if f is continuous and $f(c) < 0$, then there exists a δ such that $x \in (c-\delta, c+\delta) \Rightarrow f(x) < 0$.

Proof: Without loss of generality, suppose $f(c) > 0$. Let $\varepsilon = \frac{f(c)}{2}$. Note that by continuity, there exists δ such that

$$|x-c| < \delta \Rightarrow |f(x) - f(c)| < \frac{f(c)}{2}.$$

Unraveling the second inequality yields

$$0 < \frac{f(c)}{2} < f(x) < \frac{3f(c)}{2}$$

for all $x \in (c-\delta, c+\delta)$, as desired. ■

Proposition: If f is continuous on $[a, b]$ and $f(a)$ and $f(b)$ have different signs, then there is some $c \in (a, b)$ for which $f(c) = 0$.

Proof: Without loss of generality, assume $f(a) > 0 > f(b)$. Let

$$A = \{t : f(x) > 0, \forall x \in [a, t]\}.$$

Note that $a \in A$ and $b \notin A$. Thus A is nonempty and bounded above, so let $c = \sup(A)$. If $f(c) = 0$, then we're done.

Otherwise for the sake contradiction, assume $f(c) > 0$. Then by the previous proposition, we know that there exists δ such that $x \in (c-\delta, c+\delta) \Rightarrow f(x) > 0$, which also implies $x \in \left(c - \delta, c + \frac{\delta}{2}\right] \Rightarrow f(x) > 0$. Note that $c - \delta \in A$, since otherwise it would be an upper bound on A. But this implies that $f(x) > 0$ for all $x \in \left[a, c + \frac{\delta}{2}\right]$, which implies $c + \frac{\delta}{2} \in A$, which implies c is not an upper bound of A.

We can similarly show that $f(c) < 0$ implies a contradiction. ■

Theorem (intermediate value theorem): If f is continuous on $[a, b]$ and α is any number between $f(a)$ and $f(b)$, then there exists $c \in (a, b)$ such that $f(c) = \alpha$.

Proof: If $f(a) = f(b)$, then there's nothing to show, so suppose without loss of generality that $f(a) < \alpha < f(b)$. Now let $g(x) = f(x) - \alpha$. Clearly g is continuous on $[a, b]$, and note that $g(a) = f(a) - \alpha < 0$ and $g(b) = f(b) - \alpha > 0$. Thus by the previous proposition, there exists some $c \in (a, b)$ such that $g(c) = f(c) - \alpha = 0 \Rightarrow f(c) = \alpha$, as desired. ■

4.6. Uniform Continuity

Definition (uniform continuity): Let $f : A \to \mathbb{R}$. f is *uniformly continuous* if for all $\varepsilon > 0$ there exists some $\delta > 0$ such that for all $x, y \in A$, $|x - y| < \delta$ implies $|f(x) - f(y)| < \varepsilon$.

Proposition (sequential formulation): A function $f : X \to \mathbb{R}$ is uniformly continuous if and only if for every pair of sequences $(x_n), (y_n) \in X$ such that if $\lim_{n\to\infty}(x_n - y_n) = 0$, then $\lim_{n\to\infty}(f(x_n) - f(y_n)) = 0$.

Proof: First suppose f is uniformly continuous. Let $\varepsilon > 0$. Then there exists δ such that $|x - y| < \delta \Rightarrow |f(x) - f(y)| < \varepsilon$. Consider sequences $(x_n), (y_n)$ such that $\lim_{n\to\infty}(x_n - y_n) = 0$. This implies that there exists some N such that for all $n \geq N$, we have $|x_n - y_n| < \delta \Rightarrow |f(x) - f(y)| < \varepsilon$. Thus $\lim_{n\to\infty}(f(x_n) - f(y_n)) = 0$.

Now suppose f is not uniformly continuous. Then there exists $\varepsilon > 0$ such that for all $\delta > 0$, there exists $x, y \in X$ such that $|x - y| < \delta \Rightarrow |f(x) - f(y)| \geq \varepsilon$. Thus, for $\delta_n = \frac{1}{n}$, there exists $x_n, y_n \in A$ such that $|x_n - y_n| < \delta_n$, which implies $|f(x_n) - f(y_n)| \geq \varepsilon$. Note that $x_n - y_n < \frac{1}{n}$ converges to 0 via the squeeze theorem. However, $|f(x_n) - f(y_n)| \geq \varepsilon$ for all n, which implies $\lim_{n\to\infty}(f(x_n) - f(y_n)) \neq 0$, as desired. ∎

Proposition: If $f : A \to \mathbb{R}$ is continuous and A is compact, then f is uniformly continuous on A.

Proof: Let $\varepsilon > 0$. For each $c \in A$, let $\delta_c > 0$ be the number such that $|x - c| < \delta_c \Rightarrow |f(x) - f(c)| < \frac{\varepsilon}{2}$. Note that $\left\{\left(c - \frac{\delta_c}{2}, c + \frac{\delta_c}{2}\right)\right\}$ over all c forms an open cover of A. Since A is compact, there exists a finite subcover of these open sets,

$$\left\{\left(c_1 - \frac{\delta_{c_1}}{2}, c_1 + \frac{\delta_{c_1}}{2}\right), ..., \left(c_n - \frac{\delta_{c_n}}{2}, c_n + \frac{\delta_{c_n}}{2}\right)\right\}.$$

Let δ_{c_k} be the minimum over all δ_{c_i}.

Suppose $x, y \in A$ such that $|x - y| < \frac{\delta_{c_k}}{2}$. We have $x \in \left(c_i - \frac{\delta_{c_i}}{2}, c_i + \frac{\delta_{c_i}}{2}\right)$ for some c_i (since the intervals are a finite subcover). Then by the triangle inequality, we have

$$|y - c_i| \leq |y - x| + |x - c_i| < \frac{\delta_{c_k}}{2} + \frac{\delta_{c_i}}{2} \leq \delta_{c_i}.$$

Thus we have $|x - c_i| < \delta_{c_i}$ and $|y - c_i| < \delta_{c_i}$. This implies that $|f(x) - f(c_i)| < \frac{\varepsilon}{2}$ and $|f(y) - f(c_i)| < \frac{\varepsilon}{2}$. Then by the triangle inequality we have

$$|f(x) - f(y)| \leq |f(x) - f(c_i)| + |f(c_i) - f(y)| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

Thus f is uniformly continuous. ∎

4.7. Problems

Problem: Let $f : \mathbb{R} \to \mathbb{R}$. Prove that $\lim_{x\to 0^+} f\left(\frac{1}{x}\right) = \lim_{x\to\infty} f(x)$ if one of converges to L.

Solution: We show that if $\lim_{x\to\infty} f(x) = L$, then we also have $\lim_{x\to 0^+} f\left(\frac{1}{x}\right) = L$. Going the other way is similar.

The hypothesis implies that for all $\varepsilon > 0$, there exists N_ε such that $x > N_\varepsilon \Rightarrow |f(x) - L| < \varepsilon$. Let $\delta = \frac{1}{N_\varepsilon}$. Note that $0 < x < \delta = \frac{1}{N_\varepsilon} \Rightarrow \frac{1}{x} > N_\varepsilon \Rightarrow \left|f\left(\frac{1}{x}\right) - L\right| < \varepsilon$. This implies that $\lim_{x\to 0^+} f\left(\frac{1}{x}\right) = L$, as desired.

Problem: Let $f : [0, 1] \to \mathbb{R}$ be continuous with $f(0) = f(1)$. Show that there exist $x, y \in [0, 1]$ which are a distance $\frac{1}{2}$ apart for which $f(x) = f(y)$.

Solution: Define $g : \left[0, \frac{1}{2}\right] \to \mathbb{R}$ as $g(x) = f\left(x + \frac{1}{2}\right) - f(x)$. We need to prove that $g(c) = 0$ for some $c \in \left[0, \frac{1}{2}\right]$. Clearly g is continuous. Note that $g(0) = f\left(\frac{1}{2}\right) - f(0)$ and $g\left(\frac{1}{2}\right) = f(1) - f\left(\frac{1}{2}\right)$. Adding the equations yields $g(0) + g\left(\frac{1}{2}\right) = f(1) - f(0) = 0 \Rightarrow g(0) = -g\left(\frac{1}{2}\right)$. If $g(0) = 0$, then we're done. Otherwise, $g(0)$ and $g\left(\frac{1}{2}\right)$ have different signs, and by the IVT, $f(c) = 0$ for some $c \in \left[0, \frac{1}{2}\right]$.

Problem: Let S be a dense subset of $\mathbb{R}$, and assume that f and g are continuous functions on $\mathbb{R}$. Prove that if $f(x) = g(x)$ for all $x \in S$, then $f(x) = g(x)$ for all $x \in \mathbb{R}$.

Solution: Consider $x_0 \notin S$. Since S is dense in $\mathbb{R}$, there exists $a_n \in S$ such that $x_0 - \frac{1}{n} < a_n < x_0$. Thus $(a_n) \to x_0$, and note that $f(a_n) = g(a_n)$ for all n. Thus the limits of these functions are the same, and since both are continuous, we have $f(x_0) = g(x_0)$, as desired.

Remark: This shows that if a solution to the Cauchy functional is given to be continuous, it must be linear, since on $\mathbb{Q}$ the function must be linear.

Problem: Prove that if $f : \mathbb{R} \to \mathbb{R}$ is continuous and A is connected, then $f(A)$ is connected.

Solution: We prove the contrapositive. Suppose $f(A)$ is not connected. Thus there exist open sets U, V such that $U \cap V = \emptyset$, they both intersect $f(A)$, and $(U \cap f(A)) \cup (V \cap f(A)) = A$.

Now consider $U' = f^{-1}(U)$ and $V' = f^{-1}(V)$. Note that both are open (since f is continuous), and that $U' \cap V' = \emptyset$, since otherwise this would imply $U \cap V \neq \emptyset$. Now suppose $y_0 \in U \cap f(A)$. Then $f(x_0) = y_0$ for some $x_0 \in A$. Note that x_0 will also be in U'. This $U' \cap A \neq 0$, and similarly, $V' \cap A \neq 0$.

Now we show $(U' \cap A) \cup (V' \cap A) = A$. Suppose $x_0 \in A$. Then $f(x_0) \in f(A)$, which implies $f(x_0)$ is in either U or V, WLOG U. Then $x_0 \in U'$, which implies $x_0 \in U' \cap A \Rightarrow x_0 \in (U' \cap A) \cup (V' \cap A)$. Thus $A \subseteq (U' \cap A) \cup (V' \cap A)$.

Now suppose $x_0 \in (U' \cap A) \cup (V' \cap A)$. WLOG x_0 comes from the first term (the two terms are disjoin by $U' \cap V' = \emptyset$). Then clearly $x_0 \in A$, which implies $(U' \cap A) \cup (V' \cap A) \subseteq A$, so we're done.

Problem: Suppose $f : \mathbb{R} \to \mathbb{R}$ and $g : \mathbb{R} \to \mathbb{R}$ are uniformly continuous.

a) Prove $f + g$ is uniformly continuous.
b) If f and g are bounded, prove that fg is uniformly continuous.
c) Prove that $g \circ f$ is uniformly continuous.

Solution:

a) Let $\varepsilon > 0$. Then there exists δ_1, δ_2 such that $|x - y| < \delta_1 \Rightarrow |f(x) - f(y)| < \frac{\varepsilon}{2}$ and $|x - y| < \delta_2 \Rightarrow |g(x) - g(y)| < \frac{\varepsilon}{2}$. Let $\delta = \min\{\delta_1, \delta_2\}$. If $|x - y| < \delta$, then we have
$$|x - y| < \delta \Rightarrow |f(x) + g(x) - f(y) - g(y)| \leq |f(x) - f(y)| + |g(x) - g(y)| < \varepsilon.$$
Thus $f + g$ is uniformly continuous.

b) Let $\varepsilon > 0$. Let $M = \max\{M_1, M_2\}$, where M_1 bounds f and M_2 bounds g. There exists δ_1, δ_2 such that $|x - y| < \delta_1 \Rightarrow |f(x) - f(y)| < \frac{\varepsilon}{2M}$ and $|x - y| < \delta_2 \Rightarrow |g(x) - g(y)| < \frac{\varepsilon}{2M}$. Let $\delta = \min\{\delta_1, \delta_2\}$. Then we have
$$\begin{aligned} |x - y| < \delta \Rightarrow & |f(x)g(x) - f(y)g(y)| \\ & \leq |f(x)g(x) - f(x)g(y)| + |f(x)g(y) - f(y)g(y)| \\ & \leq M(|f(x) - f(y)| + |g(x) - g(y)|) < \varepsilon. \end{aligned}$$
Thus fg is uniformly continuous.

c) Let $\varepsilon > 0$. Then there exists δ such that $|x - y| < \delta \Rightarrow |g(x) - g(y)| < \varepsilon$. There also exists δ' such that $|x - y| < \delta' \Rightarrow |f(x) - f(y)| < \delta \Rightarrow |g(f(x)) - g(f(y))| < \varepsilon$. This $g \circ f$ is uniformly continuous.

Problem: Let $h : [0, 1)$ be a uniformly continuous function. Prove that there is a unique continuous map $g : [0, 1] \to \mathbb{R}$ such that $g(x) = h(x)$ for all $x \in [0, 1)$.

Solution: Let $(x_n) \in [0, 1)$ such that $x_n \to 1$. Pick $\varepsilon > 0$. By uniform continuity, there exists $\delta > 0$ such that $|x - y| < \delta \Rightarrow |f(x) - f(y)| < \varepsilon$ for $x, y \in [0, 1)$. Since x_n converges, it's Cauchy, so there exists N such that $n, m \geq N$ implies $|x_n - x_m| < \delta \Rightarrow |f(x_n) - f(x_m)| < \varepsilon$. Thus $(f(x_n))$ is Cauchy, so it converges to some L_1.

Now let (y_n) be another sequence in $(0, 1]$ that converges to 1. By the same process as above, $f(y_n)$ converges to L_2. Then, there exists N such that $n \geq N$ implies

$$|x_n - 1| < \frac{\delta}{2}, \quad |y_n - 1| < \frac{\delta}{2}, \quad |f(x_n) - L_1| < \varepsilon, \quad |f(y_n) - L_2| < \varepsilon.$$

We can do this by finding individual N for each inequality and taking the max. From the triangle inequality applied to the first two, we obtain $|x_n - y_n| < \delta$. Thus from uniform continuity, $|f(x_n) - f(y_n)| < \varepsilon$. Now we have

$$|L_2 - L_1| \leq |L_2 - f(y_n)| + |f(y_n) - f(x_n)| + |f(x_n) - L_1| < 3\varepsilon.$$

Since ε is arbitrary, we must have $L_2 = L_1$. Thus for any sequence that converges to 1, we have that $\lim_{n\to\infty} f(c_n) = L$ for a unique L.

We define $g(x)$ as being equal to $h(x)$ on $(0, 1]$ and equal to L at 1. Then g is continuous at 1, since every for every $x_n \to 1$, we have $g(x_n) \to g(1) = L$ by the above. Since L is unique, g is also unique.

5. Differentiation

Definition (derivative): Let A be an open set (this will often be an interval), $f : A \to \mathbb{R}$, and $c \in A$. We say f is *differnetiable* at c is

$$\lim_{x \to c} \frac{f(x) - f(c)}{x - c}$$

exists. If C is the collection of points at which f is differentiable, then the *derivative* of f is a function $f' : C \to \mathbb{R}$ where

$$f'(c) = \lim_{x \to c} \frac{f(x) - f(c)}{x - c}.$$

Remark: This definition is equivalent to

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

Remark: I won't be super picky about the kind of set that functions are defined on in this chapter for basic derivative results. I'll just declare that they're differentiable at some point, or take for granted that sequences exist that converge to limit points, since most of the time the set that functions are defined on in practice are intervals.

Proposition: Suppose $f : A \to \mathbb{R}$ is differentiable ar $c \in A$. Then f is continuous at c.

Proof: We have

$$\begin{aligned} \lim_{x \to c}[f(x) - f(c)] &= \lim_{x \to x}\left[\frac{f(x) - f(c)}{x - c} \cdot (x - c)\right] \\ &= \left(\lim_{x \to c} \frac{f(x) - f(c)}{x - c}\right)\left(\lim_{x \to c} x - c\right) \\ &= f'(c) \cdot 0 = 0 \\ &\Rightarrow \lim_{x \to c} f(x) = f(c). \end{aligned}$$

∎

Proposition (derivative rules): Let $f, g : A \to \mathbb{R}$ be differentiable at $c \in A$. Then we have the following:

a) $(f+g)'(c) = f'(c) + g'(c)$
b) $(kf)'(c) = kf'(c)$
c) $(fg)'(c) = f'(c)g(c) + f(c)g'(c)$
d) $\left(\frac{f}{g}\right)'(c) = \frac{f'(c)g(c) - f(c)g'(c)}{g(c)^2}$

Proof:

a)

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

b)

$$\lim_{x \to c} \frac{(kf)(x) - (kf)(c)}{x - c} = k \cdot \lim_{x \to c} \frac{f(x) - f(c)}{x - c} = k \cdot f'(c)$$

c)

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

d) The quotient rule follows much easier using the chain rule and product rule, which we prove next. ■

Proposition (chain rule): Let $g : A \to B$ and $f : B \to \mathbb{R}$. If g is differentiable at $c \in A$ and f is differentiable at $g(c) \in B$, then

$$(f \circ g)'(c) = f'(g(c))g'(c).$$

Proof: Consider the following function:

$$Q(y) = \begin{cases} \frac{f(y) - f(g(c))}{y - g(c)} & \text{if } y \neq g(c) \\ f'(g(c)) & \text{if } y = g(c) \end{cases}$$

This function takes place of the difference quotient in the limit and ensure that the quotient doesn't have divide by 0 problems (that's what the second case is for).

Note that Q is continuous at $g(c)$ since f is differentiable at $g(c)$ (and approaching Q from above or below $g(c)$ will alawys result in case 1).

Next we show that

$$\frac{f(g(x)) - f(g(c))}{x - c} = Q(g(x)) \cdot \frac{g(x) - g(c)}{x - c}.$$

If $g(x) \neq g(c)$, then we plug in the quotient in case 1. Then the denominator of $Q(g(x))$ cancels with $g(x) - g(c)$ and we're done. If $g(x) = g(c)$, then we want to show that

$$\frac{f(g(x)) - f(g(c))}{x - c} = f'(g(c)) \cdot \frac{g(x) - g(c)}{x - c}.$$

Then applying $g(x) = g(c)$ yields 0 on both sides.

Now we have

$$\begin{aligned}(f \circ g)'(c) &= \lim_{x \to c} \frac{f(g(x)) - f(g(c))}{x - c} \\ &= \lim_{x \to c} Q(g(x)) \cdot \frac{g(x) - g(c)}{x - c} \\ &= f'(g(c))g'(c).\end{aligned}$$

■

5.1. Min and Max

Definition (local min/max): Let $f : A \to \mathbb{R}$. Then f has a *local maximum* at $c \in A$ if there exists some $\delta > 0$ such that for all $x \in A$ for which $|x - c| < \delta$, we have

$$f(x) \leq f(c).$$

Similarly, f has a *local minimum* at $c \in A$ if there exists some $\delta > 0$ such that for all $x \in A$ for which $|x - c| < \delta$, we have

$$f(x) \geq f(c).$$

Proposition: Let A be an open set and suppose $f : A \to \mathbb{R}$ is differentiable at $c \in A$. If f has a local max or min at c, then $f'(c) = 0$.

Proof: WLOG the c is a local max, and suppose f on $V_\delta(c)$ is at most $f(c)$. Then pick a sequence (ℓ_n) with $c - \delta < \ell_n < c$ that converges to c and a sequence (r_n) with $c < r_n < c + \delta$ that converges to c. Then we have

$$\frac{f(\ell_n) - f(c)}{\ell_n - x} \geq 0 \quad \text{and} \quad \frac{f(r_n) - f(c)}{r_n - c} \leq 0$$

for all n. Since the sequences converge to c, and f is continuous at c (since it's differentiable at c), both quotients converge to $f'(c)$. Note however that the inequalities on the quotients imply that $f'(c) \geq 0$ and $f'(c) \leq 0$. Thus, $f'(c) = 0$. ■

Theorem (Darboux's theorem): Suppose $f : [a, b] \to \mathbb{R}$ is differentiable. If α is between $f'(a)$ and $f'(b)$, then there exists $c \in (a, b)$ where $f'(c) = \alpha$.

Proof: WLOG $f'(b) < \alpha < f'(a)$. Let $g(x) = f(x) - \alpha x$. Then g is differentiable on $[a, b]$ with $g'(x) = f'(x) - \alpha$. Note also that $g'(a) = f'(\alpha) - \alpha > 0$ and $g'(b) = f'(b) - \alpha < 0$. Since $[a, b]$ is compact, by the extreme value theorem, g attains a maximum on $[a, b]$. We need to show that the max does not occur at a or b.

Suppose the max occurred at a. Then $\frac{g(x)-g(a)}{x-a} \leq 0$ for all $x \in (a, b]$. Thus $g'(a) \leq 0$, but this is a contradiction. We can do basically the same thing for b.

Thus g attains its max at $c \in (a, b)$. It's clear the max is also a local max, so by the previous proposition, we have that $g'(c) = 0 \Rightarrow f'(c) = \alpha$. ■

Remark: This is really, really insane. Essentially what this means is that the derivative of a sufficiently pathological differentiable function won't have jump or removable discontinuities or asymptotes, but instead will oscillate infinitely into a point of discontinuity i.e. $x^2 \sin\left(\frac{1}{x}\right)$ at 0, where the function at 0 is defined to be 0. Despite being differentiable at 0 with derivative 0, the derivative is not continous at 0.

5.2. The Mean Value Theorem

Theorem (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 $c \in (a, b)$ where $f'(c) = 0$.

Proof: By the extreme value theorem, f hits a max at $c_1 \in [a, b]$ and a min at $c_2 \in [a, b]$. If either of these are in (a, b), then we're done by the local min/max proposition. Otherwise, c_1 and c_2 are endpoints. WLOG $c_1 = f(a)$ and $c_2 = f(b)$. Then $f(a) \geq f(x) \geq f(b)$ for all $x \in [a, b]$. However, since $f(a) = f(b)$, this implies $f(x)$ is constant, and thus $f'(x) = 0$. ■

Theorem (mean value theorem): Let $f : [a, b] \to \mathbb{R}$ be continuous on $[a, b]$ and differentiable on (a, b). Then there exists some $c \in (a, b)$ where

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

Proof: Let

$$L(x) = \left(\frac{f(b) - f(a)}{b - a}\right)(x - a) + f(a),$$

and define $g(x) = f(x) - L(x)$. Then g is continuous on $[a, b]$ and differentiable on (a, b). Note that $g(a) = g(b)$, so by Rolle's theorem, we have $g'(c) = 0$ for some $c \in (a, b)$. Thus

$$g'(x) = f'(x) - L'(x) = f'(x) - \left(\frac{f(b) - f(a)}{b - a}\right) \Rightarrow 0 = f'(c) - \left(\frac{f(b) - f(a)}{b - a}\right).$$

■

Corollary: Let I be an interval and $f : I \to \mathbb{R}$ be differentiable. If $f'(x) = 0$ for all $x \in I$, then f is constant on I.

Proof: Pick $x, y \in I$ with $x < y$. Since f is differentiable on I, it's also differentiable on $[x, y]$. Thus, by the mean value theorem we have

$$f'(c) = \frac{f(x) - f(y)}{x - y}$$

for some $c \in (x, y)$. By assumption, $f'(c) = 0$, so we have

$$0 = \frac{f(x) - f(y)}{x - y} \Rightarrow f(x) = f(y).$$

■

Corollary: Let I be an interval and $f, g : I \to \mathbb{R}$ be differentiable. If $f'(x) = g'(x)$ for all $x \in I$, then $f(x) = g(x) + C$ for some C.

Proof: Let $h(x) = f(x) - g(x)$. Then $h'(x) = f'(x) - g'(x) = 0$, and so by the previous corollary, we have that h is constant on I, which gives the desired result. ■

Corollary: Let I be an interval and $f : I \to \mathbb{R}$ be differentiable.
a) f is monotone increasing if and only if $f'(x) \geq 0$ for all $x \in I$.
b) f is monotone decreasing if and only if $f'(x) \leq 0$ for all $x \in I$.

Proof: We only prove a), since b) is extremely similar.

First supose f is monotone increasing on I. Then for any $x, c \in I$, we have

$$\frac{f(x) - f(c)}{x - c} \geq 0.$$

Thus the limit of the left side as x approaches c, which is $f'(c)$, which be nonnegative. This holds for all $c \in I$.

Now suppose $f'(x) \geq 0$ for all $x \in I$. Pick $x, y \in I$ with $x < y$. By the mean value theorem, we have

$$\frac{f(y) - f(x)}{y - x} = f'(c) \geq 0$$

for some $c \in (x, y)$. Since the denominator of the quotient is positive, the numerator must be nonnegative, which implies $f(x) \leq f(y)$, as desired. ■

Theorem (Cauchy mean value theorem): If f and g are continuous on $[a, b]$ and differentiable on (a, b), then there exists $c \in (a, b)$ such that

$$[f(b) - f(a)] \cdot g'(c) = [g(b) - g(a)] \cdot f'(c).$$

Proof: If $g(b) = g(a)$, then by Rolle's there's c such that $g'(c) = 0$, so the equation holds.

If $g(b) \neq g(a)$, then define

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

Clearly this is differentiable and continuous. Note that $h(a) = h(b)$, so by Rolle's there is c such that $h'(c) = 0$. Thus

$$h'(x) = f'(x) - \frac{f(b) - f(a)}{g(b) - g(a)} g'(x) \Rightarrow 0 = f'(c) - \frac{f(b) - f(a)}{g(b) - g(a)} g'(c),$$

as desired. ∎

5.3. L'Hôpital's Rule

Theorem (L'Hôpital's rule): Suppose I is an open interval containing a point a, and $f : I \to \mathbb{R}$ and $g : I \to \mathbb{R}$ are differentiable on I, except possibly at a. Suppose also $g'(x) \neq 0$ on I. Then, if

$$\lim_{x \to a} f(x) = 0 \quad \text{and} \quad \lim_{x \to a} g(x) = 0,$$

then

$$\lim_{x \to a} \frac{f(x)}{g(x)} = \lim_{x \to a} \frac{f'(x)}{g'(x)},$$

provided that $\lim_{x \to a} \frac{f'(x)}{g'(x)}$ exists.

Proof:

∎

Theorem (L'Hôpital's rule): Suppose I is an open interval containing a point a, and $f : I \to \mathbb{R}$ and $g : I \to \mathbb{R}$ are differentiable on I, except possibly at a. Suppose also $g'(x) \neq 0$ on I. Then, if

$$\lim_{x\to a^+} f(x) = \infty \quad \text{and} \quad \lim_{x\to a^+} g(x) = \infty,$$

then

$$\lim_{x\to a^+} \frac{f(x)}{g(x)} = \lim_{x\to a^+} \frac{f'(x)}{g'(x)},$$

provided that $\lim_{x\to a^+} \frac{f'(x)}{g'(x)}$ exists. The theorem also works if we approach from the left.

Proof: First a lemma.

Lemma: Suppose $\lim_{x\to c^+} f(x) = \infty$ and $\lim_{x\to c^+} g(x) = \infty$, and suppose $r, s \in \mathbb{R}$. If $\frac{f(x)-r}{g(x)-s}$ is bounded on (c, b) for some b. Then

$$\lim_{x\to c^+} \left[\frac{f(x)-r}{g(x)-s} - \frac{f(x)}{g(x)}\right] = 0.$$

Proof: We can rewrite the inside of the limit as

$$\frac{1}{g(x)}\left(r - s \cdot \frac{f(x)-r}{g(x)-s}\right).$$

Pick $\varepsilon > 0$. Suppose $\frac{f(x)-r}{g(x)-s}$ is bounded on (c, b) by M. Note that $\lim_{x\to c} g(x) = \infty \Rightarrow \lim_{x\to c^+} \frac{1}{g(x)} = 0$. Pick δ such that $c < x < c + \delta$ implies $\left|\frac{1}{g(x)}\right| < \frac{\varepsilon}{|r - s \cdot M|}$ (which we can do because the limit approaches 0). Then we have

$$\left|\frac{1}{g(x)}\left(r - s \cdot \frac{f(x)-r}{g(x)-s}\right)\right| < \left|\frac{\varepsilon}{|r - s \cdot M|}(r - s \cdot M)\right| = \varepsilon.$$

for all $c < x < c + \delta$. This works for any $\varepsilon > 0$, so the limit is indeed 0. ■

We prove the case when the limits approach from the right, since limits approaching from the left is analogous.

Let $\lim_{x\to a^+} \frac{f'(x)}{g'(x)} = L$, and pick $\varepsilon > 0$. By assumption, there exists δ_1 such that

$$a < x < a + \delta_1 \Rightarrow L - \frac{\varepsilon}{2} < \frac{f'(x)}{g'(x)} < L + \frac{\varepsilon}{2}.$$

Now pick $a < x_1 < x_2 < a + \delta_1$. Note that f, g are continous on $[x_1, x_2]$ and differentiable on (x_1, x_2). Thus by Cauchy's mean value theorem, we have

$$\frac{f(x_2) - f(x_1)}{g(x_2) - g(x_1)} = \frac{f'(c)}{g'(c)}$$

for some $c \in (x_1, x_2)$ (note that $g(x_2) - g(x_1)$ can't be 0, since otherwise the regular mean value theorem would imply that $g'(x) = 0$ for some x). Thus, for any $x_1, x_2 \in (a, a + \delta_1)$, we have

$$L - \frac{\varepsilon}{2} < \frac{f(x_2) - f(x_1)}{g(x_2) - g(x_1)} < L + \frac{\varepsilon}{2}.$$

By the lemma, there exists δ_2 such that for all $a < x_2 < a + \delta_2$, we have

$$\frac{f(x_2) - f(x_1)}{g(x_2) - g(x_1)} - \frac{\varepsilon}{2} < \frac{f(x_2)}{g(x_2)} < \frac{f(x_2) - f(x_1)}{g(x_2) - g(x_1)} + \frac{\varepsilon}{2}.$$

Pick $\delta = \min\{\delta_1, \delta_2\}$. Then, for all $a < x < a + \delta$, we have

$$L - \varepsilon < \frac{f(x)}{g(x)} < L + \varepsilon.$$

ε was arbitrary, so we do indeed have $\lim_{x \to a^+} \frac{f(x)}{g(x)} = L$.

5.4. Problems

Problem: Suppose $f : I \to \mathbb{R}$ is differentiable on an interval I. Prove that if f' is bounded, then f is uniformly continuous.

Solution: Consider the difference quotient $\frac{f(x)-f(y)}{x-y}$. Since f is differentiable on I, it's continuous on I, so we can apply the mean value theorem. Thus, for any $x, y \in I$, there exists $c \in I$ such that $\frac{f(x)-f(y)}{x-y} = f'(c)$. Since f' is bounded, the difference quotient must also be bounded, which means

$$\left| \frac{f(x) - f(y)}{x - y} \right| \leq M$$

for some M.

Pick some $\varepsilon > 0$, and let $\delta = \frac{\varepsilon}{M}$. The bound on the difference quotient implies

$$|f(x) - f(y)| \leq M|x - y| < M \cdot \delta = \varepsilon,$$

which implies f is uniformly continuous.

Remark: This solution also shows that f is Lipschitz.

Remark: The converse is not true. Consider $x \sin\left(\frac{1}{x}\right)$ on $[-1, 1]$ with it being defined to be 0 at $x = 0$. The function is continous on $[-1, 1]$ and so is uniformly continous on $[-1, 1]$. However, its derivative is $\sin\left(\frac{1}{x}\right) - \frac{1}{x} \cos\left(\frac{1}{x}\right)$ is unbounded.

Problem: Let I be an interval and $f : I \to \mathbb{R}$ be differentiable. Show that f is Lipschitz on I if and only if f' is bounded on I.

Solution: Suppose f is Lipschitz with constant M. Then

$$\left|\frac{f(x)-f(y)}{x-y}\right| \leq M$$

for all $x, y \in I$. Fix $y = c$ and consider the limit as $x \to c$ of the difference quotient. Clearly it exists since f is differentiable, and since every possible value of the difference quotient is bounded, the derivative at c must be bounded as well. This works for all $c \in I$, so f' is bounded on I.

Now suppose f' is bounded on I. Then, by the mean value theorem, for every $x, y \in I$, we have

$$\left|\frac{f(x)-f(y)}{x-y}\right| = |f'(c)| \leq M,$$

where M is the bound on f'. Thus f is Lipschitz with constant M.

Remark: A nice consequence of this is that if f' is continuous on a closed interval, then by the extreme value theorem it's bounded, so f is also Lipschitz.

Problem: Suppose f and g are differentiable functions with $f(a) = g(a)$ and $f'(x) < g'(x)$ for all $x > a$. Prove that $f(b) < g(b)$ for any $b > a$.

Solution: Consider $h = g - f$. Then $h' = g' - f' > 0$ and $h(a) = 0$. Then by the mean value theorem, for any $b > a$, there exists $c \in (a, b)$ such that

$$\frac{h(b)-h(a)}{b-a} = h'(c) > 0 \Rightarrow h(b) > 0 \Rightarrow g(b) > f(b),$$

as desired.

Problem: Assume that $f(0) = 0$ and $f'(x)$ is increasing. Prove that $g(x) = \frac{f(x)}{x}$ is an increasing function on $(0, \infty)$.

Solution: Note that

$$g'(x) = \frac{f'(x)x - f(x)}{x^2}.$$

Thus we just need to prove $f'(x)x > f(x) \Rightarrow f'(x) > \frac{f(x)}{x}$. Note that the right side is the mean value theorem on $[0, x]$. Thus $\frac{f(x)}{x} = f'(c)$ for some $c < x$, which means $f'(x) > f'(c)$, which is true. Thus g' is greater than 0, which means $\frac{f(x)}{x}$ is increasing, as desired.

6. Integration

6.1. Darboux Integral

Definition (partition): A *partition* of $[a, b]$ is a finite set

$$P = \{x_0, x_1, ..., x_n\}$$

such that $a = x_0$, $b = x_n$, and $x_0 < x_1 < \cdots < x_n$.

We also denote for a subinterval $[x_i, x_{i+1}]$ that

- $m_i = \inf\{f(x) : x \in [x_i, x_{i+1}]\}$
- $M_i = \sup\{f(x) : x \in [x_i, x_{i+1}]\}$

Definition (upper/lower sums): Consider a function $f : [a, b] \to \mathbb{R}$, and consider a partition $P = \{x_0, x_2, ..., x_n\}$ of $[a, b]$. Define the *upper sum* as

$$U(f, P) = \sum_{i=1}^{n} M_i(x_i - x_{i-1})$$

and the *lower sum* as

$$L(f, P) = \sum_{i=1}^{n} m_i(x_i - x_{i+1}).$$

Definition (refinement): Consider a partition of P of $[a, b]$. A partition Q of $[a, b]$ is called a *refinement* of P if $P \subseteq Q$.

Proposition: Consider a function $f : [a, b] \to \mathbb{R}$ and a partition $P = \{x_0, ..., x_n\}$ of $[a, b]$. If Q is a refinment of P, then

$$L(f, P) \leq L(f, Q) \quad \text{and} \quad U(f, P) \geq U(f, Q).$$

Proof: We prove the lower sum case, as the upper sum case is similar. We have that

$$L(f, P) = m_1(x_1 - x_0) + \cdots + m_n(x_n - x_{n-1}).$$

Since Q is a refinment of P, there are $x_{\frac{1}{n'}}, x_{\frac{2}{n'}}, ..., x_{\frac{n'-1}{n'}} \in Q$ such that

$$x_0 < x_{\frac{1}{n'}} < \cdots < x_{\frac{n'-1}{n'}} < x_1.$$

It could happen that there are no elements between x_0 and x_1, but if that's the case, then the contribution of the interval $[x_0, x_1]$ into the lower sum is the same for both P and Q, so it doesn't change the inequality.

Note that every element in $\left[x_{\frac{i}{n'}}, x_{\frac{i+1}{n'}}\right]$ is by definition at least m_1, which implies $m_{\frac{i}{n'}} \geq m_1$. Thus we have

$$\sum_{i=1}^{n'} m_{\frac{i}{n'}}\left(x_{\frac{i}{n'}} - x_{\frac{i-1}{n'}}\right) \geq \sum_{i=1}^{n'} m_1\left(x_{\frac{i}{n'}} - x_{\frac{i-1}{n'}}\right) = m_1(x_1 - x_0).$$

This holds for all of the terms in $L(f, P)$, which implies $L(f, P) \leq L(f, Q)$, as desired. ∎

Proposition: Let $f : [a, b] \to \mathbb{R}$. If P_1 and P_2 are any partitions of $[a, b]$, then

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

Proof: First note that for any partition P, we have

$$L(f, P) = \sum_{i=1}^{n} m_i(x_i - x_{i-1}) \leq \sum_{i=1}^{n} M_i(x_i - x_{i-1}) = U(f, P).$$

Now let $Q = P_1 \cup P_2$, which is clearly a refinement of both of them. Thus, by the previous proposition we have

$$L(f, P_1) \leq L(f, Q) \leq U(f, Q) \leq U(f, P_2),$$

as desired. ∎

6.2. Integrability

Definition (upper/lower integral): Let $f : [a, b] \to \mathbb{R}$ be a bounded function and let $\mathcal{P}$ denote the set of all partitions of $[a, b]$. The *upper integral* of f is defined to be

$$U(f) = \inf\{U(f, P) : P \in \mathcal{P}\},$$

and the *lower integral* of f is defined to be

$$L(f) = \sup\{L(f, P) : P \in \mathcal{P}\}.$$

Lemma (integral bound): Let $f : [a, b] \to \mathbb{R}$ be a bounded function with $m \leq f(x) \leq M$ for all $x \in [a, b]$. Then

$$m(b - a) \leq L(f) \leq U(f) \leq M(b - a).$$

Proof: The middle inequality follows from the last proposition. Let $P_0 = \{a, b\}$ be a partition of $[a, b]$. Then

$$\begin{aligned} L(f) &= \sup\{L(f,P) : P \in \mathcal{P}\} \\ &\geq L(f, P_0) \\ &\geq m(b-a). \end{aligned}$$

Note that we assume m is the infinum of f over $[a,b]$, and if it wasn't, then we just have one more inequality in the chain. The upper inequality holds similarly. ■

Definition (integrable): A bounded function $f : [a,b] \to \mathbb{R}$ is *integrable* if $L(f) = U(f)$, and we write

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

Example: Let $f : [0,1] \to \mathbb{R}$ be the Dirichlet function

$$f(x) = \begin{cases} 1 \text{ if } x \in \mathbb{Q} \\ 0 \text{ if } x \notin \mathbb{Q}. \end{cases}$$

Then f is not integrable. Let P be any partition of $[0,1]$. Note that every subinterval will contain a rational and irrational, since both sets are dense in $\mathbb{R}$. Thus $L(f,P) = 0$ and $U(f,P) = 1$, regardless of what P is. Thus $L(f) \neq U(f)$, and so f is not integrable.

Proposition: Assume that a bounded function $f : [a,b] \to \mathbb{R}$ is integrable and nonnegative on $[a,b]$. Then $\int_a^b f(x)\,dx \geq 0$.

Proof: By the integral bound, we have $0 \cdot (b-a) \leq L(f) = \int_a^b f(x)\,dx$, where the equality comes from f being integrable. ■

Proposition: Let $f : [a,b] \to \mathbb{R}$ be bounded. Then f is integrable if and only if, for all $\varepsilon > 0$ there exists a partition P_ε of $[a,b]$ where

$$U(f, P_\varepsilon) - L(f, P_\varepsilon) < \varepsilon.$$

Remark: This is easily motivated by looking at the definitions of integrability. To be integrable, we require

$$\sup\{L(f,P)\} = \inf\{U(f,P)\},$$

which implies that the elements of each set get arbitrarily close.

Proof: First suppose the condition holds for all $\varepsilon > 0$. We have $L(f, P_\varepsilon) \leq L(f)$ and $U(f) \leq U(f, P_\varepsilon)$. Thus

$$|U(f) - L(f)| \leq U(f) - L(f) \leq U(f, P_\varepsilon) - L(f, P_\varepsilon) < \varepsilon.$$

Since this holds for all $\varepsilon > 0$, we have $U(f) - L(f) = 0$, and so f is integrable.

Now suppose f is integrable, which means $U(f) = L(f) = I$. Let P_1 be a partition such that $I - \frac{\varepsilon}{2} < L(f, P_1)$, which exists since I is a supremum. Similarly, there exists P_2 such that $U(f, P_2) < I + \frac{\varepsilon}{2}$. Let $P_\varepsilon = P_1 \cup P_2$ be a refinement. We have

$$L(f, P_\varepsilon) \geq L(f, P_1) > I - \frac{\varepsilon}{2} \quad \text{and} \quad U(f, P_\varepsilon) \leq U(f, P_2) < I + \frac{\varepsilon}{2}.$$

Subtracting the first inequality from the second yields

$$U(f, P_\varepsilon) - L(f, P_\varepsilon) < \varepsilon,$$

as desired. ■

Corollary: If $f : [a, b] \to \mathbb{R}$ is integrable, then there exists a sequence of partitions (P_n) of $[a, b]$ such that

$$\lim_{n \to \infty} [U(f, P_n) - L(f, P_n)] = 0.$$

Proof: Let P_n be a partition such that $U(f, P_n) - L(f, P_n) < \frac{1}{n}$, which exists by the previous proposition. Since $U(f, P_n) \geq L(f, P_n)$, the sequence is bounded below by 0, and so the squeeze theorem implies the sequence converges to 0. ■

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

Remark: Intuitively we can expect this to hold, since on an arbitrarily small subinterval, we can make $M_i - m_i$ arbitrarily small, and then previous results will give us the desired conclusion.

Proof: Since $[a, b]$ is compact, f is bounded, so we can quote integral results. Compactness also gives uniform continuity.

Pick $\varepsilon > 0$. By unform continuity, there exists δ such that $|x - y| < \delta \Rightarrow |f(x) - f(y)| < \frac{\varepsilon}{b-a}$. Pick n such that $\frac{b-a}{n} < \delta$, and let $x_i = a + i \cdot \frac{b-a}{n}$. Then $P_\varepsilon = \{x_0, ..., x_n\}$ is a partition of $[a, b]$.

Note that on the subinterval $[x_i, x_{i+1}]$, f achieves a min and max by the extreme value theorem, m_i and M_i. Then since $x_{i+1} - x_i = \frac{b-a}{n} < \delta$, the range of f on the subinterval is contained within an interval of length $\frac{\varepsilon}{b-a}$. Thus, $|M_i - m_i| < \frac{\varepsilon}{b-a}$. This holds for any subinterval.

Now we have

$$\begin{aligned} U(f, P_\varepsilon) - L(f, P_\varepsilon) &= \sum_{i=1}^{n} (M_i - m_i)(x_i - x_{i-1}) \\ &< \sum_{i=1}^{n} \frac{\varepsilon}{b - a}(x_i - x_{i-1}) \\ &= \frac{\varepsilon}{b - a} \cdot (b - a) = \varepsilon. \end{aligned}$$

This holds for all $\varepsilon > 0$, so f is indeed integrable. ∎

Lemma: Let $f : [a, b] \to \mathbb{R}$ and $a < c < b$. Then f is integrable on $[a, b]$ if and only if f is integrable on both $[a, c]$ and on $[c, b]$.

Proof: First assume f is integrable on $[a, c]$ and $[c, b]$. Then there exists P_ε^1 and P_ε^2 such that

$$U(f, P_\varepsilon^1) - L(f, P_\varepsilon^1) < \frac{\varepsilon}{2} \text{ and } U(f, P_\varepsilon^2) - L(f, P_\varepsilon^2) < \frac{\varepsilon}{2}.$$

Let $P_\varepsilon = P_\varepsilon^1 \cup P_\varepsilon^2$, and note that it's a partition of $[a, b]$. Note that since the partitions are disjoint except for c, we have $U(f, P_\varepsilon) = U(f, P_\varepsilon^1) + U(f, P_\varepsilon^2)$, and similarly for L. Thus,

$$U(f, P_\varepsilon) - L(f, P_\varepsilon) = U(f, P_\varepsilon^1) - L(f, P_\varepsilon^1) + U(f, P_\varepsilon^2) - L(f, P_\varepsilon^2) < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

Thus f is integrable over $[a, b]$.

Now suppose f is integrable over $[a, b]$. Let P be a partition such that

$$U(f, P) - L(f, P) < \varepsilon.$$

Without loss of generality, $c \in P$, since otherwise we can add it P and both sums get refined, with the difference becoming smaller. Let $P' = P \cap [a, c]$. Then we can write $P = \{x_0, x_1, ..., x_T, x_{T+1}, ..., x_N\}$, where $P' = \{x_0, x_1, ..., x_T\}$. Thus

$$\begin{aligned} U(f, P') - L(f, P') &= \sum_{i=1}^{T} (M_i - m_i)(x_i - x_{i-1}) \\ &\leq \sum_{i=1}^{N} (M_i - m_i)(x_i - x_{i-1}) \\ &= U(f, P) - L(f, P) < \varepsilon. \end{aligned}$$

Thus f is integrable over $[a, c]$, and a similar approach shows that f is integrable over $[c, b]$. ∎

6.3. Lebesgue's Integrability Criterion

This section focuses on the integrability of functions with discontinuities. We first give a few examples of functions with discontinuities that are integrable, and then dive into weeds of Lebesgue's intergrability criterion.

Example *(one discontinuity)*: Let $f : [0, 2] \to \mathbb{R}$, defined by

$$f(x) = \begin{cases} 1 & \text{if } x \neq 1 \\ 0 & \text{if } x = 1. \end{cases}$$

Then f is integrable. Note that the upper sum is always 2, and the lower sum is $2 - \varepsilon$, where ε is the length of the subinterval that contains 1. We can make that subinterval arbitrarily small, so the upper and lower sums get arbitrarily close, meaning the function is integrable.

In general, for any function with one discontinuity, simply make the interval containing the point of discontinuity arbitrarily small.

Example *(finite number of discontinuities)*: We can split the function into separate intervals, each containing a single discontinuity. We know that on these intervals, f is integrable, and by the lemma, f on the overall interval is integrable.

Example *(countable number of discontinuities)*: Since the function's domain is compact, the discontinuities will intuitively cluster around points in the domain. Since partitions must be finite, we can pick a small enough interval around the cluster points that contain countably many of them. Then for the finitely many left discontinuities, we can also pick arbitrarily small intervals.

Example *(discountable number of discontinuities)*: The function that's 1 on the Cantor Set and 0 otherwise is actually integrable. This is especially strange since the Cantor Set is totally disconnected.

Definition (measure zero): A set A has *measure zero* if for all $\varepsilon > 0$ there exists a countable collection $I_1, I_2, I_3, \ldots$ of intervals such that

$$A \subseteq \bigcup_{k=1}^{\infty} I_k \quad \text{and} \quad \sum_{k=1}^{\infty} \mathscr{L}(I_k) < \varepsilon,$$

where $\mathscr{L}(I)$ denotes the length of the interval I.

Remark: Any subset of a measure zero set is measure zero, since the intervals that cover the set will also cover the subset.

Proposition: If a countable collection of sets $S_1, S_2, \ldots$ each has measure 0, then the union of the sets has measure 0.

Proof: For S_i, we can find intervals that cover S_i whose length is less than $\frac{\varepsilon}{2^i}$. Since the union of countably many countable sets is countable, and since that sum of the lenghts of the intervals is $\sum_{i=1}^{\infty} \frac{e}{2^i} = \varepsilon$, the union of the sets does indeed have measure 0. ■

Definition (oscillation on a set): Let f be a function defined on A. The oscillation of f on A is

$$\Omega_f(A) = \sup_{x,y \in A} |f(x) - f(y)|.$$

Definition (oscillation at a point): Let f be a function on A and $c \in A$. Then the oscillation of f at c is

$$\omega_f(c) = \inf_{r>0} \Omega_f(A \cap (c - r, c + r)).$$

Remark: Note that if $B \subseteq A$, then $\Omega_f(B) \leq \Omega_f(A)$. This means for the above definition, we can replace the inf with a limit $r \to 0^+$.

Proposition: Suppose f is defined on A and $c \in A$. Then f is continuous at c if and only if $\omega_f(c) = 0$.

Proof: Suppose f is continuous at c. Then for all $\varepsilon > 0$, there exists $\delta(\varepsilon)$ such that if $x \in A$ and $|x - c| < \delta(\varepsilon)$, then $|f(x) - f(c)| < \frac{\varepsilon}{2}$. Then by the triangle inequality, $|f(x) - f(y)| \leq |f(x) - f(c)| + |f(c) - f(y)| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2}$ for $x, y \in A \cap V_{\delta(\varepsilon)}(c)$. Thus, $0 \leq \Omega_f\left(A \cap V_{\delta(\varepsilon)}(c)\right) \leq \varepsilon$. Then, taking the limit at $\varepsilon \to 0$, by the squeeze theorem we have $\lim_{\varepsilon \to 0} \Omega_f\left(A \cap V_{\delta(\varepsilon)}(c)\right) = 0$, which implies $\omega_f(c) = \inf_{r>0} \Omega_f(A \cap V_r(c)) = 0$.

Now suppose $\omega_f(c) = 0$, and let $\varepsilon > 0$. Then, there exists δ such that $0 \leq \Omega_f(A \cap V_\delta(c)) \leq \frac{\varepsilon}{2}$. Thus $\sup_{x,y \in A \cap V_\delta(c)} |f(x) - f(y)| \leq \frac{\varepsilon}{2} \Rightarrow |f(x) - f(y)| < \varepsilon$ for all $x \in A \cap V_\delta(c)$, which means f is continuous at c. ■

Proposition: Let f be a function with domain $[a, b]$. Then for any $s > 0$, the set

$$A_s = \left\{x \in [a, b] : \omega_f(x) \geq s\right\}$$

is compact.

Proof: Note that clearly A_s is bounded, so we just need to show that it's closed. We do this by showing A_s^c is open relative to $[a, b]$.

Let $x_0 \in A_s^c$. Then $\omega_f(x_0) = t < s$. This means

$$t = \lim_{r \to 0^+} \sup_{x,y \in V_r(x_0) \cap [a,b]} |f(x) - f(y)|.$$

Then, letting $\varepsilon = \frac{s-t}{2}$, there exists δ such that $0 < r \leq \delta$ implies

$$\left| \sup_{x,y \in V_r(x_0) \cap [a,b]} |f(x) - f(y)| - t \right| < \varepsilon \Rightarrow \sup_{x,y \in V_r(x_0) \cap [a,b]} |f(x) - f(y)| < t + \varepsilon = \frac{t+s}{2}.$$

Pick $y_0 \in V_{\frac{\delta}{2}}(x_0)$. Then $V_{\frac{\delta}{2}}(y_0) \subset V_\delta(x_0)$, and so for all $0 < r' < \frac{\delta}{2}$, we have

$$\begin{aligned} \sup_{x,y \in V_{r'}(y_0) \cap [a,b]} |f(x) - f(y)| &\leq \sup_{x,y \in V_{\frac{\delta}{2}}(y_0) \cap [a,b]} |f(x) - f(y)| \\ &\leq \sup_{x,y \in V_\delta(x_0) \cap [a,b]} |f(x) - f(y)| < \frac{t+s}{2}. \end{aligned}$$

Thus,

$$\lim_{r \to 0^+} \sup_{x,y \in V_r(y_0) \cap [a,b]} |f(x) - f(y)| \leq \frac{t+s}{2} < s.$$

This means $V_{\frac{\delta}{2}}(x) \subset A_s^c$, and so A_s^c is open. ■

Theorem (Lebesgue's integrability criterion): A bounded function f on $[a, b]$ is integrable if and only if the set of discontinuities D has measure zero.

Proof: Suppose f is integrable. Let D_k be the set of points such that $\omega_f(x) \geq \frac{1}{2^k}$. Let P_k be a partition $\{x_0, ..., x_n\}$ such that

$$U(f, P_k) - L(f, P_k) < \frac{\varepsilon}{4^k}.$$

Suppose $x \in D_k \cap (x_{j-1}, x_j)$. Then there exists δ such that $V_\delta(x) \subseteq (x_{j-1}, x_j)$. Then we have

$$\frac{1}{2^k} \leq \omega_f(x) \leq \Omega_f(V_\delta(x)) \leq M_k - m_k,$$

where these all follow by definition. Let T be the set of j such that $D_k \cap (x_{j-1}, x_j) \neq \emptyset$. Then we have

$$\frac{1}{2^k} \sum_{j \in T} (x_j - x_{j-1}) \leq \sum_{j=1}^{n} (M_j - m_j)(x_j - x_{j-1}) = U(f, P_k) - L(f, P_k) \leq \frac{\varepsilon}{4^k}.$$

Note that $D_k \subseteq \bigcup_{j \in T} (x_{j-1}, x_j) \cup \bigcup_{j=0}^{n} \{x_j\}$. Note that the length of those intervals totaled is $\sum_{j \in T} (x_j - x_{j-1}) \leq \frac{\varepsilon}{2^k}$. Thus D_k is contained in a union of intervals that can get arbitrarily small, which implies D_k has measure 0. Then the collection of D_k is countable and each one has measure zero, then the union of all of them, which is equal to D, has measure zero.

Now suppose the set of discontinuities D has measure 0. Let $M = \sup_{x \in [a,b]} f(x)$ and $m = \inf_{x \in [a,b]} f(x)$. Note that if $M = m$, then f is constant, and so clearly integrable. Thus we can assume $M > m$. Define $A_s = \{x \in [a, b] \mid \omega_f(x) \geq s\}$ with $s > 0$. Then $A_s \subseteq A$ and so $m(A_s) = 0$, where $m(S)$ denotes the measure of S.

Let $\varepsilon > 0$. Since $A_{\frac{\varepsilon}{2(b-a)}}$ has measure zero, there exist open intervals $I_1, I_2, ...$ such that

$$A_{\frac{\varepsilon}{2(b-a)}} \subset \bigcup_{k=1}^{\infty} I_k \quad \text{and} \quad \sum_{k=1}^{\infty} m(I_k) < \frac{\varepsilon}{2(M - m)}.$$

Since $A_{\frac{\varepsilon}{2(b-a)}}$ is compact and the I_k's cover it, then there's a finite subcover $I_1, I_2, ..., I_N$.

If $x \in [a, b] \setminus \left(\bigcup_{k=1}^{N} I_k\right) \subset [a, b] \setminus A_{\frac{\varepsilon}{2(b-a)}}$, then $\omega_f(x) < \frac{\varepsilon}{2(b-a)}$. Thus, for each x there exists δ_x such that $y, z \in V_{\delta_x}(x) \Rightarrow |f(y) - f(z)| < \frac{\varepsilon}{2(b-a)}$. Since $[a, b] \setminus \left(\bigcup_{k=1}^{N} I_k\right)$ is compact (the union of open intervals is open), and since the $V_{\delta_x}(x)$'s cover this set, there exists a finite subcover $\{(x'_1 - \delta_1, x'_1 + \delta_1), ..., (x'_k - \delta_k, x'_k + \delta_k)\}$.

Now we construct a partition P such that $U(f, P) - L(f, P) < \varepsilon$. Note that

$$(x'_1 - \delta_1, x'_1 + \delta_1), ..., (x'_k - \delta_k, x'_k + \delta_k), I_1, ..., I_N$$

is a finite cover of $[a, b]$. Let $P = \{x_0, x_1, ..., x_n\}$ be a partition such that each $[x_{i-1}, x_i]$ is entirely contained in one of those intervals.

Let C_1 be the set of subintervals formed by P that are contained in some I_k, and let C_2 be the set of subintervals in the other open intervals. We have

$$\begin{aligned}\sum_{C_1}(M_i - m_i)(x_i - x_{i-1}) &< (M-m)\sum_{C_1}(x_i - x_{i-1}) \\ &< (M-m)\sum_{C_1} m(I_k) \\ &< (M-m)\cdot\frac{\varepsilon}{2(b-a)} \\ &= \frac{\varepsilon}{2}.\end{aligned}$$

Since $[x_{i-1}, x_i] \subset \left(x'_j - \delta_j, x'_j + \delta_j\right)$, we have $y, z \in [x_{i-1}, x_i] \Rightarrow |f(y) - f(z)| < \frac{\varepsilon}{2(b-a)}$. Then

$$\sum_{C_2}(M_i - m_i)(x_i - x_{i-1}) \leq \frac{\varepsilon}{2(b-a)}\sum_{C_2}(x_i - x_{i-1}) < \frac{\varepsilon}{2(b-a)}\cdot(b-a) = \varepsilon.$$

Adding the two sums yields $U(f,P) - L(f,P) < \varepsilon$. Thus, f is integrable on $[a,b]$. ■

6.4. Integral Properties

Definition:

$$\int_a^a f(x)\,dx = 0$$

$$\int_a^b f(x)\,dx = -\int_b^a f(x)\,dx.$$

Remark: These are definitions since we only defined integrals for $a < b$.

Proposition (additivity of bounds): Assume that $f : [a,b] \to \mathbb{R}$ is integrable. If $a < c < b$, then

$$\int_a^b f(x)\,dx = \int_a^c f(x)\,dx + \int_c^b f(x)\,dx.$$

Proof: Since f is integrable over $[a,b]$, then it's Let P_1 and P_2 be partitions such that $U(f,P_1) - L(f,P_1) < \frac{\varepsilon}{2}$ and $U(f,P_2) - L(f,P_2) < \frac{\varepsilon}{2}$. Let $P = P_1 \cup P_2$. Then

$$U(f,P) - L(f,P) = [U(f,P_1) + U(f,P_2)] - [L(f,P_1) + L(f,P_2)] < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

We can rewrite the inequality as $U(f,P) - \varepsilon < L(f,P) < \int_a^b f(x)\,dx = I < U(f,P) < L(f,P) + \varepsilon$, which implies

$$[U(f,P_1) + U(f,P_2)] - \varepsilon < I < [L(f,P_1) + L(f,P_2)] + \varepsilon.$$

Since integrals are less than upper sums and greater than lower sums, we have

$$\left[\int_a^c f(x)\,dx + \int_c^b f(x)\,dx\right] - \varepsilon < I < \left[\int_a^c f(x)\,dx + \int_c^b f(x)\,dx\right] + \varepsilon.$$

Since ε was arbitrary, this implies the desired equality. ■

Proposition (linearity of integral operator): Let $f, g : [a, b] \to \mathbb{R}$ be integrable. Then kf is integrable and

$$\int_a^b k \cdot f(x)\,dx = k \int_a^b f(x)\,dx$$

for all $k \in \mathbb{R}$, and $f + g$ is integrable with

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

Proof: If $k > 0$, then $U(kf, P) - L(kf, P) = k(U(f, P) - L(f, P))$, and we can make the second term as small as we want, so kf is integrable. Then we have

$$\begin{aligned}\int_a^b k \cdot f(x)\,dx &= \inf\{U(kf, P) : P \in \mathcal{P}\} \\ &= \inf\{k \cdot U(f, P) : P \in \mathcal{P}\} \\ &= k \cdot \inf\{U(f, P) : P \in \mathcal{P}\} \\ &= k \cdot U(f) = k \int_a^b f(x)\,dx.\end{aligned}$$

If $k = 0$, then the result is obvious. If $k < 0$, then in a subinterval of a partition, M_i and m_i switch, so $U(kf, P) = k \cdot L(f, P)$ and vice versa. Then the rest follows similarly.

Note that $U(f + g, P) \le U(f, P) + U(g, P)$, since

$$\sup\{f(x) + g(x) : x \in [x_{i-1}, x_i]\} \le \sup\{f(x) : x \in [x_{i-1}, x_i]\} + \sup\{g(x) : x \in [x_{i-1}, x_i]\},$$

and similarly $L(f + g, P) \ge L(f, P) + L(g, P)$. Then we have $U(f + g, P) - L(f + g, P) \le [U(f, P) - L(f, P)] + [U(g, P) - L(f, P)]$, and we can make the right side arbitrarily small, so $f + g$ is integrable.

Note that

$$L(f, P) + L(g, P) \le L(f + g, P) \le L(f + g) = U(f + g) \le U(f + g, P) \le U(f, P) + U(g, P).$$

There exists a sequence of partitions P_n^1 such that $U(f, P_n^1), L(f, P_n^1) \to I_f$, and a sequence of partitions P_n^2 such that $U(g, P_n^2), L(g, P_n^2) \to I_g$. Taking $P_n = P_n^1 \cup P_n^2$ allows both to happen simultaneously. Then taking the limit of inequality using this partition yields

$$\int_a^b f(x)\,dx + \int_a^b g(x)\,dx \le I_{f+g} \le \int_a^b f(x)\,dx + \int_a^b g(x)\,dx,$$

which implies the desired result. ■

Corollary: If $f, g : [a, b] \to \mathbb{R}$ are integrable and $f(x) \leq g(x)$ for all $x \in [a, b]$, then

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

Proof: We have

$$\int_a^b (g(x) - f(x)) \geq 0 \Rightarrow \int_a^b g(x)\,dx \geq \int_a^b f(x)\,dx.$$

■

Corollary: If $f : [a, b] \to \mathbb{R}$ is integrable, then $|f|$ is also integrable, and

$$\left| \int_a^b f(x)\,dx \right| \leq \int_a^b |f(x)|\,dx.$$

Proof: Let P be a partition such that $U(f, P) - L(f, P) < \varepsilon$. Let M_i' be the supremum of $|f(x)|$ on a subinterval, and m_i' be the infinum of $|f(x)|$ on a subinterval. Then $M_i - m_i \geq M_i' - m_i'$, since if M and m have the same sign, the two sides are equal, and if they have differents signs, the right side is smaller. Thus $U(|f|, P) - L(|f|, P) \leq U(f, P) - L(f, P) < \varepsilon$, and so $|f|$ is integrable.

Note that $-|f(x)| \leq f(x) \leq |f(x)|$, and so from the previous corollary we obtain

$$-\int_a^b |f(x)|\,dx \leq \int_a^b f(x)\,dx \leq \int_a^b |f(x)|\,dx,$$

which gives us the desired result. ■

Theorem (integral mean value theorem): If f is continuous on $[a, b]$, then there exists $c \in [a, b]$ such that

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

Proof: Let M, m denoted the max, min of f on $[a, b]$ respectively. Letting the integral equal I, we have $m(b - a) \leq I \leq M(b - a)$. Dividing by $(b - a)$ yields $m \leq \frac{1}{b-a} \cdot I \leq M$, and then we're done by the intermediate value theorem. ■

6.5. Fundamental Theorem of Calculus

Theorem (ftc part 1): If $f : [a, b] \to \mathbb{R}$ is integrable, and $F : [a, b] \to \mathbb{R}$ satisfies $F'(x) = f(x)$ for all $x \in [a, b]$, then

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

Proof: Let P be a partition. Note that by the derivative mean value theorem, we have

$$F'(c_i) = \frac{F(x_i) - F(x_{i-1})}{x_i - x_{i-1}} \Rightarrow F(x_i) - F(x_{i-1}) = F'(c_i)(x_i - x_{i-1}) = f(c_i)(x_i - x_{i-1}).$$

for some $c_i \in (x_{i-1}, x_i)$. Then we have

$$L(f, P) \leq \sum_{i=1}^n m_i(x_i - x_{i-1}) \leq \sum_{i=1}^n f(c_i)(x_i - x_{i-1}) \leq \sum_{i=1}^n M_i(x_i - x_{i-1}) \leq U(f, P).$$

Note that the middle sum is equal to $\sum_{i=1}^n F(x_i) - F(x_{i-1})$, which telescopes to $F(b) - F(a)$. Since F is integrable, there exists a sequence of partitions for which both the upper and lower sums approach $\int_a^b f(a)\,dx$. Thus by the squeeze theorem we have

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

■

Theorem (ftc part 2): Let $g : [a, b] \to \mathbb{R}$ be integrable and define $G : [a, b] \to \mathbb{R}$ by

$$G(x) = \int_a^x g(t)\,dt.$$

Then G is continuous. Moreover, if g is continuous, then G is differentiable and $G'(x) = g(x)$.

Proof: First assume g is integrable. If $g = 0$, then G is also 0, which is continuous. Thus we can assume $M = \sup\{|g(x)| : x \in [a, b]\}$ is greater than 0.

Pick $x_0 \in [a, b]$. We show G is continuous at x_0. Suppose $\varepsilon > 0$ and let $\delta = \frac{\varepsilon}{M}$. If $|x - x_0| < \delta$, then

$$\begin{aligned} |G(x) - G(x_0)| &= \left| \int_a^x g(t)\,dt - \int_a^{x_0} g(t)\,dt \right| \\ &= \left| \int_{x_0}^x g(t)\,dt \right| \\ &\leq |M(x - x_0)| \\ &< |M \cdot \delta| = \varepsilon. \end{aligned}$$

Thus G is continuous at x_0.

Now suppose g is also continuous. Pick $c \in [a, b]$. We need to show

$$\lim_{x \to c} \frac{G(x) - G(c)}{x - c} = g(c).$$

Pick some sequence $x_n \to c$. Note that

$$G(x_n) - G(c) = \int_a^{x_n} g(t)\, dt - \int_a^c g(t)\, dt = \int_c^{x_n} g(t)\, dt = g(c_n)(x_n - c),$$

where the last equality comes from the integral mean value theorem (since g is continuous) and c_n is in between c and x_n. Thus we have

$$\frac{G(x_n) - G(c)}{x_n - c} = g(c_n).$$

Note that by the squeeze theorem, $c_n \to c$. Thus taking the limit of both sides as $n \to \infty$ yields $G'(c) = g(c)$, as desired. ■

6.6. Integration Rules

Proposition (integration by parts): If f and g are differentiable with continuous derivatives on $[a, b]$. Then $f'g$ and fg' are integrable and

$$\int_a^b f(x)g'(x)\, dx = f(b)g(b) - f(a)g(a) - \int_a^b f'(x)g(x)\, dx.$$

Proof: Note that since the dertivatives are continuous, $f'g$ and fg' are both continuous, and thus integrable. By the product rule, we have $(f(x)g(x))' = f'(x)g(x) + f(x)g'(x)$. Integrating, we get

$$\int_a^b (f(x)g(x))'\, dx = f(a)g(a) - f(b)g(b) = \int_a^b f'(x)g(x)\, dx + \int_a^b f(x)g'(x)\, dx,$$

as desired. ■

Proposition (u-sub): Suppose g is a function whose derivative g' is continuous on $[a, b]$, and suppose that f is a function that is continuous on $g([a, b])$. Then

$$\int_a^b f(g(x))g'(x)\, dx = \int_{g(a)}^{g(b)} f(u)\, du.$$

Proof: Note that by IVT, $g([a, b])$ is either in interval or a single point. In the case of a single point, g is constant, and so $g' = 0$, and both integrals are equal to 0, so we can assume $g([a, b]) = [c, d]$ for $c \neq d$.

Define

$$F(x) = \int_{g(a)}^{x} f(t)\, dt,$$

and note that $F(g(a)) = 0$. By the chain rule and FTC, we have $\frac{d}{dx} F(g(x)) = F'(g(x))g'(x) = f(g(x))g'(x)$. Since g' is continuous, the right hand side is continuous and thus integrable. Integrating both sides and using FTC yields

$$\int_a^b f(g(x))g'(x) dx = \int_a^b \frac{d}{dx} F(g(x))\, dx = F(g(b)) - F(g(a)) = \int_{g(a)}^{g(b)} f(u)\, du,$$

as desired. ■

6.7. Problems

Problem: Show that $\int_0^\pi \frac{\sin(xt)}{t}\, dt$ varies continuously in x.

Solution: Let $I(x)$ denote the integral above, and fix $c \in \mathbb{R}$. We have

$$\begin{aligned} \left| \frac{\sin(xt)}{t} - \frac{\sin(ct)}{t} \right| &= \frac{2}{t} \left| \sin\left(\frac{t(x-c)}{2}\right) \cos\left(\frac{t(x+c)}{2}\right) \right| \\ &\leq \frac{2}{t} \left| \sin\left(\frac{t(x-c)}{2}\right) \right| \\ &\leq \frac{2}{t} \left| \frac{t(x-c)}{2} \right| = |x-c|. \end{aligned}$$

Thus

$$\begin{aligned} |I(x) - I(c)| &= \left| \int_0^\pi \frac{\sin(xt)}{t} - \frac{\sin(ct)}{t}\, dt \right| \\ &\leq \int_0^\pi \left| \frac{\sin(xt)}{t} - \frac{\sin(ct)}{t} \right| dt \\ &\leq \int_0^\pi |x-c|\, dt = \pi |x-c|. \end{aligned}$$

Thus I is Lipschitz, and so continuous.

7. Sequences and Series of Functions

7.1. Functional Convergence and Properties

A lot of things port over from regular sequences, but we need an extra notion of functions getting arbitrarily across the domain in order to get properties we like with convergence.

Definition (pointwise convergence): Suppose (f_k) is a sequence of functions defined on $A \subseteq \mathbb{R}$. The sequence *converges pointwise* to a function $f : A \to \mathbb{R}$ if, for each $x_0 \in A$,

$$\lim_{k\to\infty} f_k(x_0) = f(x_0).$$

Definition (uniform convergence): Let (f_k) be a sequence of functions defined on $A \subseteq \mathbb{R}$. Then (f_k) *converges uniformly* on A to a function f if, for every $\varepsilon > 0$, there exists an $N \in \mathbb{N}$ such that $|f_k(x) - f(x)| < \varepsilon$ for all $k \geq N$ and for all $x \in A$.

Example: Consider $f_k(x) = \frac{x^2+kx}{k}$. As k increases, f_k stay a parabola. However, we have

$$\lim_{k\to\infty} \frac{x_0^2 + kx_0}{k} = \lim_{k\to\infty} \frac{x_0^2}{k} + x_0 = x_0,$$

and so f_k converges pointwise to $f(x) = x$. However, it doesn't converge uniformly to x, x and a parabola can differ an arbitrarily large amount.

Proposition: Suppose $f_k : A \to \mathbb{R}$ converges uniformly to f. Then (f_k) converges to f pointwise.

Proof: Pick $x_0 \in A$. We want to show $\lim_{n\to\infty} f_n(x_0) = f(0)$. By uniform convergence, there exists N such that $k \geq N$ implies $|f_k(x) - f(x)| < \varepsilon$ for all $x \in A$. In particular, this holds for x_0, and since ε was arbitrary, the limit does indeed hold. ■

Definition (Cauchy sequence): Let $f_k : A \to \mathbb{R}$. Then (f_k) is Cauchy if for all ε, there exists N such that for all $m, n \geq N$,

$$|f_m(x) - f_n(x)| < \varepsilon$$

for all $x \in A$.

Proposition: Let $f_k : A \to \mathbb{R}$. Then the sequence (f_k) converges uniformly if and only if (f_k) is Cauchy.

Proof: Suppose (f_k) is Cauchy. Then there exists N such that $m, n \geq N \Rightarrow |f_n(x) - f_m(x)| < \frac{\varepsilon}{2}$ for all $x \in A$. Note for a fixed x_0, we get a convergent value $f(x_0)$ since the regular sequence is Cauchy and so converges. We claim this f is what the sequence converges to. We already showed that the sequence converges pointwise to f. Thus, we can fix x_0 and the limit as m approaches infinity to get

$$\lim_{m \to \infty} |f_n(x_0) - f_m(x_0)| = |f_n(x_0) - f(x_0)| \leq \frac{\varepsilon}{2} < \varepsilon.$$

Since this holds for any $x_0 \in A$, and for all $n \geq N$, we see that (f_k) converges uniformly to f, as desired.

Now suppose (f_k) converges uniformly to f. There exists N such that $i, j \geq N$ implies $|f_i(x) - f(x)| < \frac{\varepsilon}{2}$ and $\left|f(x) - f_j(x)\right| < \frac{\varepsilon}{2}$ for all $x \in A$. Adding them together and using the triangle inequality yields $\left|f_i(x) - f_j(x)\right| < \varepsilon$ for all $x \in A$. Thus (f_k) is Cauchy. ■

Continuity

Example: Consider $f_k(x) = x^k$ on $[0, 1]$. Note that f_k converges pointwise to a function that's 0 everywhere except 1. This shows that a sequence of continuous functions can converge pointwise to a noncontinuous function.

Proposition: Assume each $f_k : A \to \mathbb{R}$ is continuous at some $c \in A$. If (f_k) converges uniformly to f, then f is continuous at c.

Proof: From uniform convergence, we know there exists some N such that

$$|f_k(x) - f(x)| < \frac{\varepsilon}{3}$$

for all $k \geq N$ and for all $x \in A$, and so holds for N. In particular, $|f_N(c) - f(c)| < \frac{\varepsilon}{3}$. Since f_N is continuous at c, there exists δ such that $|x - c| < \delta \Rightarrow |f_N(x) - f_N(c)| < \frac{\varepsilon}{3}$. Then we have

$$|f(x) - f(c)| \leq |f(x) - f_N(x)| + |f_N(x) - f_N(c)| + |f_N(c) - f(c)| < \frac{\varepsilon}{3} + \frac{\varepsilon}{3} + \frac{\varepsilon}{3} = \varepsilon$$

for all $|x - c| < \delta$ and $x \in A$. Thus f is continuous at c. ■

Boundedness

Example: Consider

$$f_k(x) = \begin{cases} \frac{1}{x} & \text{if } \ x \in \left[\frac{1}{k}, 1\right] \\ 0 & \text{if } \ x \in \left(0, \frac{1}{k}\right). \end{cases}$$

Clearly each f_k is bounded, but they converge pointwise to $\frac{1}{x}$, which is unbounded on $(0, 1]$, so pointwise convergence does not necessarily maintian boundedness.

Proposition: Assume that each $f_k : A \to \mathbb{R}$ is bounded and $f_k \to f$ uniformly. Then f is also bounded.

Proof: By uniform convergence, there exists N such that $k \geq N$ implies $|f_k(x) - f(x)| < 1$ for all $x \in A$. In particular, this implies

$$|f_N(x) - f(x)| < 1 \Rightarrow f_N(x) - 1 < f(x) < f_N(x) + 1.$$

Since f_N is bounded, the left and right sides are bounded, and so f is bounded as well. ∎

Unboundedness

Example: Consider $f_k : (0, 1] \to \mathbb{R}$ defined by

$$f_k(x) = \begin{cases} \frac{1}{x} & \text{if } \ x \in \left(0, \frac{1}{k}\right] \\ 0 & \text{if } \ x \in \left(\frac{1}{k}, 1\right]. \end{cases}$$

Note each function is unbounded on $\left(0, \frac{1}{k}\right]$, but it converges pointwise to 0. Thus pointwise convergence does not necessarily preserve unboundedness.

Proposition: Suppose $f_k : A \to \mathbb{R}$ is unbounded and $f_k \to f$ uniformly. Then f is unbounded.

Proof: By uniform convergence, there exists N such that $k \geq N$ implies $|f_k(x) - f(x)| < 1$ for all $x \in A$. In particular, this implies

$$|f_N(x) - f(x)| < 1 \Rightarrow f_N(x) - 1 < f(x) < f_N(x) + 1.$$

If f_N is unbounded above, then the left side of the inequality is unbounded, and so f is unbounded above. Similarly, if f_N is unbounded below, the right side is unbounded, and so f is unbounded below. ∎

Uniform Continuity

Obviously pointwise converging functions won't necessarily converge to a uniformly continuous function since they sometimes don't even converge to a continuous function.

Proposition: Suppose each $f_k : A \to \mathbb{R}$ is uniformly continuous and uniformly converges to f. Then f is uniformly continuous.

Proof: By uniform convergence, there exists N such that $k \geq N \Rightarrow |f_k(x) - f(x)| < \frac{\varepsilon}{3}$ for all $x \in A$. In particular $|f_N(x) - f(x)| < \frac{\varepsilon}{3}$. By uniform continuity, there also exists δ such that $|x - y| < \delta \Rightarrow |f_N(x) - f_N(y)| < \frac{\varepsilon}{3}$. Then, for $|x - y| < \delta$, we have

$$\begin{aligned} |f(x) - f(y)| &= |f(x) - f_N(x) + f_N(x) - f_N(y) + f_N(y) - f(y)| \\ &< |f_N(x) - f(x)| + |f_N(x) - f_N(y)| + |f_N(y) - f(y)| \\ &< \frac{\varepsilon}{3} + \frac{\varepsilon}{3} + \frac{\varepsilon}{3} = \varepsilon, \end{aligned}$$

which implies f is uniformly continuous. ∎

Differentiability

Unfortunately, uniform convergence and differentiabilty don't play as nicely as we'd like, at least when we only assume the functions are differentiable and nothing more.

Example: Consider $f_k : [-1, 1] \to \mathbb{R}$ defined by $f_k(x) = x^{1+\frac{1}{2k-1}}$. We can show that $f_k \to |x|$ uniformly, which is not differentiable, even though each f_k is differentiable at 0.

Example: Consider $f_k = \frac{x}{1+kx^2}$.

Proposition: Suppose $f_k : [a, b] \to \mathbb{R}$ and assume each f_k is differentiable. If (f_n') converges uniformly to g, and there exists some $x_0 \in [a, b]$ such that $(f_k(x_0))$ converges, then (f_k) converges uniformly to some f with $f' = g$.

Remark: The condition on (f_n) converging at some point is needed so that the sequence of functions doesn't blow up to infinity because of some increasing constant that disappears under differentiation.

Proof: First we show that f uniformly converges. We have

$$\begin{aligned} |f_n(x) - f_m(x)| &\leq |(f_n(x) - f_m(x)) - (f_n(x_0) - f_m(x_0))| + |f_n(x_0) - f_m(x_0)| \\ &= |(x - x_0)(f_n'(c) - f_m'(c))| + |f_n(x_0) - f_m(x_0)| \\ &\leq |a - b||f_n'(c) - f_m'(c)| + |f_n(x_0) - f_m(x_0)|, \end{aligned}$$

where the equality came from using the mean value theorem on $f_n - f_m$ with c between x and x_0. Since (f_n') converges uniformly, the sequence is uniformly Cauchy, so there exists N_1 such that $n, m \geq N_1 \Rightarrow |f_n'(c) - f_m'(c)| < \frac{\varepsilon}{2|a-b|}$. Since $(f_k(x_0))$ converges, it's also Cauchy, so there exists N_2 such that $n, m \geq N_2 \Rightarrow |f_n(x_0) - f_m(x_0)| < \frac{\varepsilon}{2}$. Letting $N = \max\{N_1, N_2\}$, we have for any $n, m \geq N$ that

$$|a - b||f_n'(c) - f_m'(c)| + |f_n(x_0) - f_m(x_0)| < |a - b| \cdot \frac{\varepsilon}{2|a - b|} + \frac{\varepsilon}{2} = \varepsilon.$$

Thus (f_k) is uniformly Cauchy, and so uniformly converges to some function f.

Next we show that $f' = g$. We have

$$\left|\frac{f(x) - f(c)}{x - c} - g(c)\right| \leq \left|\frac{f(x) - f(c)}{x - c} - \frac{f_n(x) - f_n(c)}{x - c}\right| + \left|\frac{f_n(x) - f_n(c)}{x - c} - f_n'(c)\right| + |f_n'(c) - g(c)|.$$

Consider

$$\frac{f_m(x) - f_m(c)}{x - c} - \frac{f_n(x) - f_n(c)}{x - c}.$$

As we did in the first part, we can use the mean value theorem on $f_m(x) - f_n(x)$ to obtain

$$\frac{f_m(x) - f_m(c)}{x - c} - \frac{f_n(x) - f_n(c)}{x - c} = f_m'(y) - f_n'(y)$$

for some y in between x and c. Since (f_n') converges uniformly, there exists N_1 such that $n, m \geq N_1$ implies

$$\left|\frac{f_m(x) - f_m(c)}{x - c} - \frac{f_n(x) - f_n(c)}{x - c}\right| = |f_m'(y) - f_n'(y)| < \frac{\varepsilon}{3}.$$

Letting $m \to \infty$ yields that for any $n \geq N_1$, we have

$$\left|\frac{f(x) - f(c)}{x - c} - \frac{f_n(x) - f_n(c)}{x - c}\right| < \frac{\varepsilon}{3}.$$

Since (f_n') converges uniformly to g, there exists N_2 such that $n \geq N_2 \Rightarrow |f'(c) - g(c)| < \frac{\varepsilon}{3}$. Let $N = \max\{N_1, N_2\}$. Since f_N is differentiable, there exists δ such that $0 < |x - c| < \delta$ implies

$$\left|\frac{f_N(x) - f_N(c)}{x - c} - f_N'(c)\right| < \frac{\varepsilon}{3}.$$

Combining everything with the initial inequality yields

$$\begin{aligned}\left|\frac{f(x) - f(c)}{x - c} - g(c)\right| &\leq \left|\frac{f(x) - f(c)}{x - c} - \frac{f_N(x) - f_N(c)}{x - c}\right| + \left|\frac{f_N(x) - f_N(c)}{x - c} - f_N'(c)\right| + |f_N'(c) - g(c)| \\ &< \frac{\varepsilon}{3} + \frac{\varepsilon}{3} + \frac{\varepsilon}{3} = \varepsilon.\end{aligned}$$

Thus f is differentiable at c with derivative $g(c)$, as desired. ■

Remark: What this result is essentially saying is that

$$\lim_{n \to \infty} \lim_{x \to c} \frac{f_n(x) - f_n(c)}{x - c} = \lim_{x \to c} \lim_{n \to \infty} \frac{f_n(x) - f_n(c)}{x - c}.$$

Integrability

First we prove a proposition that we use in the next result.

Proposition: For bounded f and g on $[a, b]$, we have

$$U(f + g) \leq U(f) + U(g)$$

and

$$L(f + g) \geq L(f) + L(g).$$

Proof: We prove the upper sum case, as the lower sum case follows similarly. For any partition P, we have

$$U(f + g) \leq U(f + g, P) \leq U(f, P) + U(g, P).$$

Since $U(f)$ is an infinum, there exists a sequence of partitions such that $U(f, P_n)$ approaches $U(f)$. Similarly, there exists such a sequence of partitions for $U(g)$. Taking the union of each term in the sequence of partitions gives a sequence for which both terms converge to their upper sums. Since the inequality above holds for all partitions, we obtain $U(f+g) \leq U(f) + U(g)$, as desired. ■

Proposition: Suppose each $f_k : [a,b] \to \mathbb{R}$ is integrable. If (f_k) converges uniformly to f, then f is integrable, and

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

Proof: Since each f_k is integrable, each is bounded, which implies f is bounded by our boundedness results.

Now we can prove that $L(f) = U(f)$. By uniform convergence, there exists N such that $k \geq N$ implies

$$|f_k(x) - f(x)| < \frac{\varepsilon}{2(b-a)}$$

for all $x \in [a,b]$.

Then we have

$$\begin{aligned} U(f) - L(f) &= U(f - f_N + f_N) - L(f - f_N + f_N) \\ &\leq U(f - f_N) + U(f_N) - L(f - f_N) - L(f_N), \end{aligned}$$

where the inequality comes from the previous proposition. Since f_N is integrable, $U(f_N) = L(f_N)$, we get $U(f) - L(f) \leq U(f - f_N) - L(f - f_N)$. From uniform convergence, we have $-\frac{\varepsilon}{2(b-a)} < f_N(x) - f(x) < \frac{\varepsilon}{2(b-a)}$. Then we get

$$U(f) - L(f) \leq U(f - f_N) - L(f - f_N) < U\left(\frac{\varepsilon}{2(b-a)}\right) - L\left(-\frac{\varepsilon}{2(b-a)}\right) = \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

Thus $0 \leq U(f) - L(f) < \varepsilon$, and so $U(f) - L(f) = 0$. Thus f is integrable.

Now we prove the integral converges to the integral of the convergent function. By uniform convergence, there exists N such that $k \geq N \Rightarrow |f_k(x) - f(x)| < \frac{\varepsilon}{b-a}$ for all $x \in [a,b]$. Thus

$$f_k(x) - \frac{\varepsilon}{b-a} < f(x) < f_k(x) + \frac{\varepsilon}{b-a}$$

for all $k \geq N$ and $x \in [a,b]$. Integrating both sides yields

$$\int_a^b f_k(x) - \frac{\varepsilon}{b-a}\,dx = \int_a^b f_k(x)\,dx - \varepsilon < \int_a^b f(x)\,dx < \int_a^b f_k(x) + \frac{\varepsilon}{b-a}\,dx = \int_a^b f_k(x)\,dx + \varepsilon$$

for all $k \geq N$, which implies

$$\left|\int_a^b f_k(x)\,dx - \int_a^b f(x)\,dx\right| < \varepsilon,$$

and so the sequence does converge to $\int_a^b f(x)\,dx$. ■

Remark: What this result is saying is that

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

Arzela-Ascoli Theorem

Definition (uniformly bounded): Let $\mathcal{F}$ be a family of functions with each $f : U \to \mathbb{R}$. Then $\mathcal{F}$ is uniformly bounded if there exists some M such that for all $f \in \mathcal{F}$ and for all $x \in U$, we have $|f(x)| \leq M$.

Definition (equicontinuity): Let $\mathcal{F}$ be a familiy of functions with each $f : U \to \mathbb{R}$.

- $\mathcal{F}$ is *equicontinuous* at $x_0 \in U$ if for every $\varepsilon > 0$, there exists $\delta > 0$ such that for all $f \in \mathcal{F}$ and all z with $|x - x_0| < \delta$, then $|f(x) - f(x_0)| < \varepsilon$ (also called pointwise equicontinuity).
- $\mathcal{F}$ is *uniformly equicontinuous* if, for every $\varepsilon > 0$, there exists $\delta > 0$ such that for all $f \in \mathcal{F}$ and all x, y with $|x - y| < \delta$, then $|f(x) - f(y)| < \varepsilon$.

Theorem (Arzela-Ascoli theorem): Let $X \subseteq \mathbb{R}$ be a bounded set, and let $\mathcal{F}$ be an infinite family of uniformly bounded and uniformly equicontinuous functions $f : X \to \mathbb{R}$. Then there exists a uniformly convergent subsequence $f_1, f_2, \ldots \in \mathcal{F}$.

Proof: We first find a countable and dense subset $z_1, z_2, \ldots$ of X. We can extract countably many of the functions in $\mathcal{F}$ and make a sequence out of them, say $g_1, g_2, \ldots$. Because $\mathcal{F}$ is uniformly bounded, say by M, some subsequence $g_{1,1}, g_{1,2}\ldots$ exists such that $g_{1,1}(z_1), g_{1,2}(z_1), \ldots$ forms a convergent subsequence by Bolzano-Weierstrass.

Now, out of $g_{1,1}, g_{1,2}\ldots$, find a subsequence $g_{2,1}, g_{2,2}, \ldots$ such that $g_{2,1}(z_2), g_{2,2}(z_2)$ for a convergent sequence. Note also that $g_{2,1}(z_1), g_{2,2}(z_1)$ also converges, since its a subsequence of a convergent sequence. We can keep doing this, and so we have that $g_{k,1}(z_k), g_{k,2}(z_k), \ldots$ converges for all k.

Let $f_n = g_{n,n}$. Note that for each k, the sequence $f_1(z_k), f_2(z_k)\ldots$ is a convergent sequence, since $f_k(z_k), f_{k+1}(z_k)\ldots$ is a subsequence of the convergent sequence $g_{k,1}(z_k), g_{k,2}(z_k)$ (remember that each $(g_{k,i})$ is a subsequence of $(g_{k-1,i})$), and the first finitely many terms don't matter. Thus $f_1(z_k), f_2(z_k), \ldots$ converges.

Now we show that $f_1, f_2, \ldots$ is uniformly convergent. Let $\varepsilon > 0$, and choose $\delta > 0$ such that $z, w \in X$ are such that $|z - w| < \delta \Rightarrow |f(z) - f(w)| < \frac{\varepsilon}{3}$ for all $f \in \mathcal{F}$, in particular $|f_n(z) - f_n(w)| < \frac{\varepsilon}{3}$ (which exists by uniform equicontinuity). Find some N such that for every $z \in X$, there exists some n with $1 \leq n \leq N$ such that $|z - z_n| < \delta$ (we can show N must be finite by first taking the closure of X, which makes it compact since X is bounded, and then taking the union of all $B_\delta(z_n)$. Since $z_1, z_2, \ldots$ is dense, this union must cover X. Thus there exists a finite subcover of the closure. Since they cover $\overline{X}$ and $X \subseteq \overline{X}$, they also cover X. Thus there is a maximal n among the cover, and we can set N to be that maximum). Because $f_1, f_2, \ldots$ converges at each z_n, there is some K such that if $\ell, k \geq K$, then

$$|f_\ell(z_n) - f_k(z_n)| < \frac{\varepsilon}{3} \quad \text{for all } n \text{ with } 1 \leq n \leq N.$$

Now pick $z \in X$. There exists some $n \leq N$ such that $|z - z_n| < \delta$, so if $\ell, m > K$, then

$$|f_k(z) - f_\ell(z)| \leq |f_k(z) - f_k(z_n)| + |f_k(z_n) - f_\ell(z_n)| + |f_\ell(z_n) - f_\ell(z)|.$$

The first and third terms are $< \frac{\varepsilon}{3}$ because of the choice of δ, and the second term is $< \frac{\varepsilon}{3}$ by the above. Thus we have $|f_k(z) - f_\ell(z)| < \varepsilon$ for all $z \in X$ and all $k, \ell \geq K$. Thus $f_1, f_2, \ldots$ is uniformly convergent on X.

■

7.2. Series of Functions

Definition (series of functions): Let (f_k) be a sequence of functions defined on a set A and let $s_n = \sum_{i=1}^n f_i$. The series $\sum_{i=1}^n f_i$ converges pointwise to $f : A \to \mathbb{R}$ is (s_n) converges pointwise to f, and it converges uniformly to f is (s_n) converges uniformly to f.

Proposition: Let $f_k : A \to \mathbb{R}$. Then $\sum_{n=1}^\infty f_n$ converges uniformly on A if and only if for every $\varepsilon > 0$ there exists N such that

$$\left| \sum_{k=m}^n f_k(x) \right| < \varepsilon$$

for all $n \geq m \geq N$ and for all $x \in A$.

Proof: Just apply Cauchy convergence on the partial sums. ■

This method for determining convergence sucks, since the functions can be crazy, so its easier to use the following slightly weaker result.

Proposition (Weierstrass M-test): Let $f_k : A \to \mathbb{R}$ and suppose for each k there exists M_k such that $|f_k(x)| \leq M_k$ for all $x \in A$. If $\sum_{k=1}^\infty M_k$ converges, the $\sum_{k=1}^\infty f_k(x)$ converges uniformly on A.

Proof: Since $\sum_{k=1}^\infty M_k$ converges, the partial sums are Cauchy, so there exists N such that for all $n \geq m \geq N$ we have

$$\sum_{k=m}^n M_k < \varepsilon.$$

Then we have

$$\left|\sum_{k=m}^{n} f_k(x)\right| \le \sum_{k=m}^{n} |f_k(x)| \le \sum_{k=m}^{n} M_k < \varepsilon.$$

Thus the partial sums of $\sum_{k=1}^{\infty} f_k(x)$ are Cauchy, so the sum converges uniformly. ■

7.3. Power Series

Definition (formal power series): A *formal power series centered at* c is a series of the form

$$\sum_{n=0}^{\infty} a_n (x-c)^n$$

with each $a_n \in \mathbb{R}$.

Proposition: Let

$$R = \frac{1}{\limsup_{n\to\infty} |a_n|^{\frac{1}{n}}}$$

where $R = \infty$ is the denominator is 0 and $R = 0$ is the denominator is infinity. Then the power series with coefficients a_n centerd at c has a radius of convergence R, and interval of convergence $(c-R, c+R)$ (the endpoints could possibly be included in the interval convergence, depending on the coefficients).

Proof: Follows from using the root test on the power series. ■

Proposition: For any $0 < r < R$, the series $\sum_{n=0}^{\infty} a_n (x-c)^n$ converges uniformly on the compact interval $[c-r, c+r]$.

Proof: For any $c + \ell$ in the interval, we have

$$|a_n (c+\ell-c)^n| < \left|a_n r^n \cdot \frac{\ell^n}{r^n}\right| \le |a_n r^n| = M_n.$$

Since $c + r$ is within the interval of convergence, the power series absolutely converges at $x = c + r$, and the terms M_n are the terms of the power series at $x = c + r$. Thus by the Weierstrass M-test, the power series converges uniformly on $[c-r, c+r]$. ■

Remark: Since each term of the power series is continuous, this implies that a power series is continuous on its interval of convergence.

Remark: Note that uniform convergence does not necessarily extend to the whole interval of convergence, since at the endpoints the series can diverge, for example $1 + x + x^2 + \cdots$.

Theorem (Abel's theorem): Suppose $\sum_{n=0}^{\infty} a_n (x-c)^n$ be a power series that converges at $c + R$ with $R > 0$. Then the series converges uniformly on $[c, c+R]$. We have a similar result for $c - R$.

Proof: Without loss of generality, suppose $c = 0$ and $R = 1$, and pick $\varepsilon > 0$. We need to find N such that $n > m \geq N \Rightarrow |a_m x^m + \cdots a_n x^n| < \varepsilon$ for all $x \in [0, 1]$ (this follows from the equivalence of uniform convergence and a sequence of functions be Cauchy, in this case the partial sums). From the convergence at 1, we know that $|a_m + \cdots + a_n|$ gets arbitrarily small, say less than $\frac{\varepsilon}{2}$. Note also that (x^n) is monotone decreasing for $x \in [0, 1]$. Then by Abel's lemma, we have

$$\left| \sum_{k=m}^{n} a_k x^k \right| \leq \frac{\varepsilon}{2} \cdot x^m < \varepsilon.$$

■

Remark: This is a significant strengthening of the previous proposition, since now the boundary points can possibly be included as well, depending on whether they converge. This theorem also implies that if a series converges at an endpoint, then it's continuous there.

Proposition: Let $\sum_{n=0}^{\infty} a_n (x-c)^n$ be a power series with $R > 0$. Then the power series is differentiable on $(c - R, c + R)$ with derivative $\sum_{n=0}^{\infty} n a_n (x-c)^{n-1}$.

Proof: We invoke the result about uniformly converging derivatives. Let s_n be the partial sums of the first power series, and t_n be the partial sums of the second. Clearly $s_n' = t_n$, and s_n converges at some point since $R > 0$. We just need to show that t_n converges uniformly and we're done.

We show that $\limsup_{n\to\infty} |n a_n|^{\frac{1}{n}} = R$, which then implies that s_n converges uniformly on $[c - r, c + r]$ for all $0 < r < R$, which is what we need. Note that $\limsup_{n\to\infty} |n a_n|^{\frac{1}{n}} = \limsup_{n\to\infty} n^{\frac{1}{n}} |a_n|^{\frac{1}{n}}$. The first term in the product gets arbitrarily close to 1, and the second term has limsup $\frac{1}{R}$, so indeed the limsup of $|n a_n|^{\frac{1}{n}}$ is $\frac{1}{R}$. ■

Proposition: Let $\sum_{n=0}^{\infty} a_n(x-c)^n$ be a power series with $R > 0$. If $[a, b] \subseteq (c - R, c + R)$, then

$$\int_a^b \left(\sum_{n=0}^{\infty} a_n (x-c)^n \right) dx = \sum_{n=0}^{\infty} a_n \frac{(b-c)^{n+1} - (a-c)^{n+1}}{n+1}.$$

Proof: We have

$$\sum_{k=0}^{\infty} \int_a^b a_k (x-c)^k \, dx = \lim_{n\to\infty} \sum_{k=0}^{n} \int_a^b a_k (x-c)^k \, dx = \lim_{n\to\infty} \int_a^b \sum_{k=0}^{n} a_k (x-c)^k \, dx = \lim_{n\to\infty} \int_a^b s_n \, dx$$
$$= \int_a^b \lim_{n\to\infty} s_n \, dx = \int_a^b \sum_{k=0}^{\infty} a_k (x-c)^k \, dx.$$

We're allowed to bring the limit inside the integral in the second line since the sequence of partial sums converges uniformly on $[a, b]$. Then the first term in the string of equalities is equal to the desired sum by just integrating. ■

Remark: What both of these results say is that we can differetiate/integrate power series term by term, which is really useful. On top of that, they keep the same interval of convergence modulo endpoints.

7.4. Taylor and Maclaurin Series

Definition (Taylor/Maclaurin series): Suppose $f^{(k)}(c)$ exists for all $k \in \mathbb{N}$. The *Taylor series* of f about c is

$$\sum_{k=0}^{\infty} \frac{f^{(k)}(c)}{k!} (x-c)^k.$$

If $c = 0$, then the series is called a *Maclaurin series.*

For a Taylor series, we define the Taylor polynomial of degree n at c to be

$$T_{x=c}^{n}(f) = \sum_{k=0}^{\infty} \frac{f^{(k)}(c)}{k!} (x-c)^k.$$

Definition (error function): The Error function $E_n(x)$ for a function f is defined by

$$E_n(x) = f(x) - T_{x=c}^{n}(f).$$

Lemma: Suppose f is infinitely differentiable in an interval I and $c \in I$. Then for $x \in I$,

$$f(x) = \sum_{k=0}^{\infty} \frac{f^{(k)}(c)}{k!}(x-c)^k \quad \text{if and only if} \quad E_n(x) \to 0 \text{ pointwise.}$$

Proof: For a fixed $x \in I$, we have

$$\begin{aligned} f(x) = \sum_{k=0}^{\infty} \frac{f^{(k)}(c)}{k!}(x-c)^k &\iff f(x) = \lim_{n\to\infty} \sum_{k=0}^{n} \frac{f^{(k)}(c)}{k!}(x-c)^k \\ &\iff f(x) = \lim_{n\to\infty} T^n_{x=c}(f) \\ &\iff \lim_{n\to\infty} [f(x) - T^n_{x=c}(f)] = 0 \\ &\iff \lim E_n(x) = 0. \end{aligned}$$

■

Theorem (integral error function): Suppose f is infinitely differentiable in an interval I and $c \in T$. Then for $x \in I$ we have

$$E_n(x) = \frac{1}{n!} \int_c^x (x-t)^n f^{(n+1)}(t)\, dx.$$

Proof: We proceed by induction. For the base case, we need to show that $E_1(x)$ equals $\int_c^x (x-t) f''(t)$. To do this, we note that

$$E_1(x) = f(x) - T^1_{x=c}(x) = f(x) - f(c) - f'(c)(x-c).$$

Rewriting the right side yields

$$\int_c^x (f'(t) - f'(c))\, dt.$$

Integrating by parts using $u = f'(t) - f'(c)$ and $v = t - x$ yields

$$(f'(t) - f'(c))(t-x)\Big|_c^x + \int_c^x (x-t) f''(t)\, dt = \int_c^x (x-t) f''(t)\, dt,$$

as desired.

Now suppose the result holds for k. We have

We have

$$E_{k+1}(x) = E_k(x) - \frac{f^{(k+1)}(c)}{(k+1)!}(x-c)^{k+1} = \frac{1}{k!} \int_c^x (x-t)^k f^{(k+1)}(t)\, dt - \frac{f^{(k+1)}(c)}{(k+1)!}(x-c)^{k+1}.$$

Integrating using integration by parts with $u = f^{(k+1)}(t)$ and $v = -\frac{(x-t)^{k+1}}{k+1}$ yields

$$\frac{1}{k!}\left(-\frac{f^{(k+1)}(t)}{k+1}(x-t)^{k+1}\Big|_c^x + \int_c^x \frac{f^{(k+2)}(t)(x-t)^{k+1}}{k+1}\,dt\right)$$
$$= \frac{f^{(k+1)}(c)}{(k+1)!}(x-c)^{k+1} + \frac{1}{(k+1)!}\int_c^x f^{(k+2)}(t)(x-t)^{k+1}.$$

Thus we have

$$E_{k+1} = \frac{1}{(k+1)!}\int_c^x f^{(k+2)}(t)(x-t)^{k+1}\,dt,$$

as desired. ∎

Theorem (Lagrange error function): Suppose f is infinitely differentiable on I and $c \in I$. Then for any other $x_0 \in I$ there exists a_n between x_0 and c such that

$$f(x_0) = \sum_{k=0}^{n-1} \frac{f^{(k)}(c)}{k!}(x_0-c)^k + \frac{f^{(n)}(\alpha_n)}{n!}(x_0-c)^n.$$

That is,

$$E_{n-1}(x_0) = \frac{f^{(n)}(\alpha_n)}{n!}(x_0-c)^n.$$

Proof: Without loss of generality that $c < x_0$. By the extreme value theorem, $f^{(n)}$ has a min and max on $[c, x_0]$, m and M respectively. Then using the integral error function, we have

$$\frac{m}{n!}(x_0-c)^n = \frac{m}{(n-1)!}\int_c^{x_0}(x_0-t)^{n-1}\,dt \le E_{n-1}(x_0) \le \frac{M}{(n-1)!}\int_c^{x_0}(x_0-t)^{n-1}\,dt = \frac{M}{n!}(x_0-c)^n.$$

Note that $\frac{f^{(n)}(t)(x_0-c)^n}{n!}$ is bounded between $\frac{m}{n!}(x_0-c)^n$ and $\frac{M}{n!}(x_0-c)^n$. Thus by the intermediate value theorem, there exists $\alpha_n \in [c, x_0]$ such that

$$E_{n-1}(x_0) = \frac{f^{(n)}(\alpha_n)}{n!}(x_0-c)^n,$$

as desired. ∎

Proposition (Cauchy error form): Suppose f is $N+1$ times differentiable on $(-R, R)$. Then, for $x \in (-R, R)$, there exists c between 0 and x such that

$$E_N(x) = \frac{f^{N+1}(c)}{N!}(x-c)^N x.$$

Proof: Let

$$S_N(x, a) = \sum_{n=0}^{N} \frac{f^n(a)}{n!}(x-a)^n,$$

and let $E_N(x, a) = f(x) - S_N(x, a)$. Note that $E_N(x, a)$ is differentiable with respect to a, from which we get

$$\frac{d}{da} E_N(x, a) = -f'(a) + \sum_{n=1}^{N} \left(\frac{f^n(a)}{(n-1)!}(x-a)^{n-1} - \frac{f^{n+1}(a)}{n!}(x-a)^n \right),$$

which telescopes to $-\frac{f^{N+1}(a)}{N!}(x-a)^N$. Then from the mean value theorem, we have

$$\frac{E_N(x, x) - E_N(x, 0)}{x} = E'(x, c)$$

for some c between 0 and x. Note that $S_N(x, x) = f(x)$, so $E_N(x, x) = 0$. Thus writing in terms of $E_N(x, 0) = E_N(x)$, we get the desired conclusion ■

7.5. Weierstrass Approximation Theorem

Lemma: The function $f(x) = \sqrt{1-x}$ has a power series representation that converges uniformly to it on the interval $[-1, 1]$.

Proof:

■

Lemma: For $\varepsilon > 0$, there exists a polynomial $p(x)$ for which

$$||x| - p(x)| < \varepsilon$$

for all $x \in [a, b]$.

Proof: We can assume without loss of generality that we're working on the interval $[-1, 1]$, since for any other interval, we can simply scale sufficiently to arrive at a subset of $[-1, 1]$.

Note that $|x| = \sqrt{1-(1-x^2)}$, and since $1 - x^2 \in [-1, 1]$ for $x \in [-1, 1]$, can expand this using the Taylor series for $\sqrt{1-x}$ about 0, and plugging in $1 - x^2$. Since the series converges uniformly on $[-1, 1]$, we can cutoff the series at some point and obtain a polynomial that approximates with error less than ε, as desired. ■

Definition (polygonal function): A continuous function $\varphi : [a, b] \to \mathbb{R}$ is *polygonal* if there exists a partition of $[a, b]$ such that φ is linear on each subinterval of the partition.

Lemma: Let $f : [a, b] \to \mathbb{R}$ be continuous. Given $\varepsilon > 0$, there exists a polygonal function φ satisfying

$$|f(x) - \varphi(x)| < \varepsilon$$

for all $x \in [a, b]$.

Proof: Since $[a, b]$ is compact, f is uniformly continuous on its domain. ■

Theorem (Weierstrass approximation theorem): Let $f : [a, b] \to \mathbb{R}$ be continuous. Given $\varepsilon > 0$, there exists a polynomial $p(x)$ satisfying

$$|f(x) - p(x)| < \varepsilon$$

for all $x \in [a, b]$.

Proof: As before, without loss of generality we show the result for $[-1, 1]$. ■

Corollary: Given a continous function $f : [a, b] \to \mathbb{R}$, there exists a sequence of polynomials (p_n) such that $p_n \to f$ uniformly.

Proof: Obvious. ■

Corollary: Suppose $f : [a, b] \to \mathbb{R}$ is continuously differentiable. Then there exists a polynomial $p(x)$ such that

$$|f(x) - p(x)| < \varepsilon \quad \text{and} \quad |f'(x) - p'(x)| < \varepsilon$$

for all $x \in [a, b]$.

Proof: Since f' is continuous, there exists p such that

$$|f'(x) - p(x)| < \frac{\varepsilon}{b - a}.$$

Then for $x \in [a, b]$, we have

$$\begin{aligned} \varepsilon > \varepsilon \cdot \frac{x - a}{b - a} &> \int_a^x \frac{\varepsilon}{b - a}\, dt \\ &> \int_a^x |f'(t) - p(t)|\, dt \geq \left| \int_a^x f'(t) - p(t)\, dt \right| = |f(x) - P(x) - f(a) + P(a)|, \end{aligned}$$

where $P'(x) = p(x)$, and where we used the fundamental theorem of calculus in the last equality. Note that we can force $P(a) = f(a)$, since that won't change $p(x)$. Thus, $P(x)$ is our desired polynomial. ■

7.6. Problems

8. Metric Spaces

8.1. Basic Notions

Basic analysis but $|\cdot|$ is replaced with d.

Definition (metric space): A *metric space* (M, d) is a space M of objects, together with a *distance function* or *metric* $d : M \times M \to [0, \infty)$ which satisifes the following three conditions:

a) For any $x, y \in M$, we have $d(x, y) = 0$ if and only if $x = y$.
b) For any $x, i \in M$, we have $d(x, y) = d(y, x)$.
c) For any $x, y, z \in M$, we have $d(x, z) \leq d(x, y) + d(y, z)$.

Remark: We often want to consider a subset of M with the metric d, in which case we say that the subset E *inherits* the metric d from M, writing $d|_{E \times E}$ or d_E.

Example: The standard metric used on the reals is the absolute value metric, namely $d(x, y) = |x - y|$.

Example *(sup norm)*: For $x, y \in \mathbb{R}^n$, define $d_{l^\infty} : \mathbb{R}^n \times \mathbb{R}^n \to [0, \infty)$ as

$$d_{l^\infty}(x, y) = \sup\{|x_i - y_i| : 1 \leq i \leq n\}.$$

Example *(discrete metric)*: For an arbitrary set M, let $d_{\text{disc}} : M \times M \to [0, \infty)$ be defined as $d_{\text{disc}}(x, y) = 0$ if $x = y$, and $d_{\text{disc}}(x, y) = 1$ otherwise.

From here, we can basically redo everything till continuity with respect to an arbitrary metric (in fact, in certain metrics we can go beyond).

Definition (convergence in metric spaces): Suppose (x_n) is a sequence in the metric space (M, d). Then $(x_n) \to x$ if, for all $\varepsilon > 0$, there exists N such that

$$n \geq N \Rightarrow d(x_n, x) < \varepsilon.$$

Proposition: If a sequence in a metric space converges to two different limits, then the limits are the same.

Proof: Suppose that in (M, d), the sequence (x_n) converges to L_1 and L_2. We have

$$d(L_1, L_2) \leq d(L_1, x_n) + d(x_n, L_2).$$

By definition of convergence, the right side gets arbitrarily close to 0 as $n \to \infty$. Thus $d(L_1, L_2) = 0 \Rightarrow L_1 = L_2$. ■

Definition (open ball): Let (M, d) be a metric space, let x_0 be a point in M, and let $r > 0$. The *open ball* $B_{(M,d)}(x_0, r)$ in M, centered at x_0 with radius r with respect to d is the set

$$B_{(M,d)}(x_0, r) = \{x \in M : d(x, x_0) < r\}.$$

When the space and metric function are clear, we abbreviate it as $B_r(x_0)$.

Definition (interior, exterior, boundary): Let (M, d) be a metric space and let E be a subset of X. We say a point $x_0 \in X$ is an *interior point* of E is there exists $r > 0$ such that $B_r(x_0) \subseteq E$. We say that $x_0 \in X$ is an *exterior point* if there exists $r > 0$ such that $B_r(x_0) \cap E = \emptyset$. We say that $x_0 \in X$ is a *boundary point* if it's neither an interior or exterior point.

The set of all interior points of E is denoted $\operatorname{int}(E)$, the set of all exterior points of E is denoted $\operatorname{ext}(E)$, and the set of boundary points of E is denoted ∂E.

Definition (adherent point): Let (M, d) be a metric space, let E be a subset of M, and let x_0 be a point in M. We say x_0 is an *adherent point* of E is for every radius $r > 0$, the ball $B_r(x_0)$ has nonempty intersection with E.

Definition (limit point of a set): Let (M, d) be a metric space, let E be a subset of M, and let x_0 be a point in M. We say x_0 is a *limit point* of E if there exists a sequence (a_n) in $E \setminus \{x_0\}$ such that $a_n \to x_0$.

Definition (closure): Let (M, d) be a metric space and let E be a subspace of M. The *closure* of E, denoted as $\overline{E}$, is the set of all adherent points of R.

Proposition: Let (M, d) be a metric space and let E be a subspace of M. Let E' be the set of all limit points of E. Then $\overline{E} = E'$.

Proof: Suppose x_0 is a limit point of E. Thus there's a sequence $(a_n) \in E \setminus \{x_0\}$ such that $a_n \to x_0$. Pick $\varepsilon > 0$. Then from the definition of convergence, there exists N such that $n \geq N \Rightarrow d(a_n, x_0) < \varepsilon$. Taking $n = N$, we can clearly see that $a_n \in B_\varepsilon(x_0) \cap E$, and this holds for any ε, so clearly x_0 is an adherent point of E. Thus $E' \subseteq \overline{E}$.

Now suppose x_0 is an adherent point of E. Suppose there exists some $\varepsilon > 0$ such that $B_\varepsilon(x_0) \cap E = \{x_0\}$. Then clearly x_0 is not a limit point. However, from the intersection we see that $x_0 \in E$, so x_0 is in both $\overline{E}$ and E'.

In the other case, for all $\varepsilon > 0$, the intersection $B_\varepsilon(x_0) \cap E$ has a point that isn't x_0. Pick $\varepsilon = \frac{1}{n}$, and choose the point in the intersection that isn't x_0. Then we have a sequence that converges to x_0, and so x_0 is a limit point. Thus $\overline{E} \subseteq E'$. ■

Remark: While in regular $\mathbb{R}$, the limit point definition of closure is easier to use, in arbitrary metric spaces, it's easier to use the adherent definition, since to be adherent you need to be in the space, and so for arbitrary spaces you only need to focus points within the space.

Proposition: Let (M, d) be a metric space, and let E be a subset of M. Then every adherent point of E is either an interior point or a boundary point.

Proof: Follows from definitions ■

Definition (open and closed sets): Let (M, d) be a metric space, and let E be a subset of X. We say E is *closed* if it contains all of its boundary points. We say that E is *open* if it contains none of its boundary points.

Corollary: E is closed if and only if $E = \overline{E}$.

Proof: Obvious. ■

Remark: The notion of open sets here is equivalent to every point having a neighborhood within the set. Similarly, the notion of closed sets here is equivalent to the complement being open.

8.2. Cauchy Sequences and Complete Metric Spaces

Definition (subsequence): Suppose (x_n) is a sequence in a metric space (M, d). Suppose (n_i) is a strictly increasing sequence of integers. Then $\left(x_{n_i}\right)$ is a *subsequence* of (x_n).

Proposition: Suppose $x_n \to x$ in a metric space (M, d). Then every subsequence converges to x.

Proof: Suppose $\left(x_{n_i}\right)$ is a subsequence, and pick $\varepsilon > 0$. There exists N such that $n \geq N \Rightarrow d(x_n, x) < \varepsilon$. Clearly there exists I such that $i \geq I \Rightarrow n_i \geq N$, and thus $d\left(x_{n_i}, x\right) < \varepsilon$. Thus $\left(x_{n_i}\right) \to x$. ■

Definition (limit point of a sequence): Suppose (x_n) is a sequence in (M, d), and let $L \in M$. We say L is a *limit point* of (x_n) if for every $N > 0$ and $\varepsilon > 0$, there exists $n \geq N$ such that $d(x_n, L) < \varepsilon$.

Proposition: Suppose (x_n) is a sequence in (M, d), and let $L \in M$. Then L is a limit point of (x_n) if and only if there exists a subsequence converging to L.

Proof: Follows easily from definitions. ■

Definition (Cauchy sequence): Let (x_n) be a sequence in (M, d). We say this sequence is a *Cauchy sequence* if for every $\varepsilon > 0$, there exists N such that $i, j \geq M \Rightarrow d\left(x_i, x_j\right) < \varepsilon$.

Proposition: Suppose the sequence (x_n) in (M, d) converges to x. Then the sequence is Cauchy.

Proof: Pick $\varepsilon > 0$. From convergence, there exists N such that $n \geq N \Rightarrow d(x_n, x) < \frac{\varepsilon}{2}$. Pick $i, j \geq N$. Then we have

$$d\left(x_i, x_j\right) \leq d(x_i, x) + d\left(x, x_j\right) < \varepsilon,$$

so the sequence is Cauchy. ■

Unlike in $\mathbb{R}$, being Cauchy doesn't imply convergence, because a metric space doesn't necessarily need to be complete.

Example: Consider $(\mathbb{Q}, |\cdot|)$. Then the sequence

$$3, 3.1, 3.14, 3.141, 3.14159, \ldots$$

Clearly this sequence is Cauchy, but it converges to $\pi \notin \mathbb{Q}$.

Proposition: Suppose (x_n) is Cauchy in (M, d), and some subsequence $\left(x_{n_i}\right)$ converges to x. Then $x_n \to x$.

Proof: Pick $\frac{\varepsilon}{2}$. From convergence, there exists I such that $i \geq I \Rightarrow d\left(x_{n_i}, x\right) < \frac{\varepsilon}{2}$. From Cauchy, there exists N such that $n, m \geq N \Rightarrow d(x_n, x_m) < \frac{\varepsilon}{2}$. Let $N' = \max\{N, n_I\}$. Then for $n, m \geq N'$ we have

$$d(x_n, x) \leq d(x_n, x_m) + d(x_m, x).$$

Letting $m = n_j$, where $n_j \geq N'$, we obtain

$$d\left(x_n, x_{n_j}\right) + d\left(x_{n_j}, x\right) \leq \varepsilon.$$

Thus x_n converges to x. ∎

Proposition: Every Cauchy sequence has at most one limit point.

Proof: Suppose the sequence $(a_n) \in (M, d)$ has two limit points x, y. Then there exist two subsequence of (a_n), say (x_n) and (y_n), that converge to x and y respectively. From the previous proposition, this implies that $a_n \to x$ and $a_n \to y$. But since limits are unique, this implies that $x = y$. ∎

Definition (complete metric space): A metric space (M, d) is *complete* if every Cauchy sequence in (M, d) converges to a point in M.

Proposition:
a) Let (X, d) be a metric space, and let $(Y, d|_{Y \times Y})$ be a subspace of (X, d). If $(Y, d|_{Y \times Y})$ is complete, then Y must be closed in X.
b) Let (X, d) be a complete metric space, and suppose Y is a closed subset of X. Then the subspace $(Y, d|_{Y \times Y})$ is complete.

Proof:

a) Let $d_Y = d|_{Y \times Y}$. Let $y_0 \in X$ be an adherent point of Y. If there exists some r such that $B_r(y_0) \cap Y = \{y_0\}$, then we must necessarily have that $y_0 \in Y$. Otherwise, for each integer n, there exists some point not equal to y_0 in the intersection of $B_{\frac{1}{n}}(y_0) \cap Y$. Let this point be y_n. Then, with respect to (X, d), we have $\lim_{n \to \infty} y_n = y_0$.

   Pick $\varepsilon > 0$, and let $N > \frac{2}{\varepsilon}$. Then $\forall i, j \geq N$, we have

   $$d\left(y_i, y_j\right) \leq d(y_i, y_0) + d\left(y_j, y_0\right) < \frac{2}{N} < \varepsilon.$$

   Thus, with respect to (X, d), the sequence (y_n) is Cauchy. However, since $y_i, y_j \in Y$, we also have $d\left(y_i, y_j\right) = d_Y\left(y_i, y_j\right)$, and thus is also Cauchy with respect to (Y, d_Y). Then from completeness, the sequence must converge to $y' \in Y$. However, this implies that $y_n \to y$ with respect to (X, d) as well, and since limits are unique, we must have $y_0 = y' \in Y$. Thus, Y contains all of its adherent points, and therefore is closed in X.

b) Let $d_Y = d|_{Y\times Y}$. Suppose $(y_n) \in Y$ is a Cauchy sequence. Then from completeness, it converges to $x \in X$ with respect to (X, d). Thus, for any $\varepsilon > 0$, there exists N such that $n \geq N \Rightarrow d(y_n, x) < \varepsilon$. Thus $B_\varepsilon(x) \cap Y \neq \emptyset$ for all $\varepsilon > 0$, which implies that x is an adherent point of Y. Since Y is closed, it must contain x. But then $d_Y(y_n, x) < \varepsilon$ for all $n \geq N$, and so $\lim_{N\to\infty} y_n \to x$ with respect to (Y, d_Y). Thus the Cauchy sequence (y_n) converges to a point in Y, and this Y is complete.

■

8.3. Compact Metric Spaces

Definition (compact): A metric space (M, d) is said to be *compact* if every sequence in (M, d) has a convergent subsequence. A subset Y of M is said to be compact if $(Y, d|_{Y\times Y})$ is compact.

Remark: This is one of the equivalent definitions of compactness for $\mathbb{R}$.

Remark: From this definition it easily follows that a metric space is complete if and only if every sequence has a limit point.

Definition (bounded): Let (M, d) be a metric space, and let Y be a subset of M. We say that Y is *bounded* if for every $x \in M$, there exists some finite r such that $Y \subseteq B_r(x)$. We call the metric space (M, d) bounded if M is bounded.

Example: Consider $\mathbb{R}$ with the following metric:

$$d(x, y) = \frac{|x - y|}{1 + |x - y|}.$$

Then $0 \leq d(x, y) < 1$ for all $x, y \in \mathbb{R}$. Thus, given any point $x \in \mathbb{R}$, we have $\mathbb{R} \subseteq B_2(x)$, so $(\mathbb{R}, d)$ is bounded.

Theorem (one direction of Heine-Borel): Let (M, d) be a compact metric space. Then (M, d) is complete and bounded.

Remark: This is equivalent to one half of Heine-Borel on the reals, except closed is replaced with complete, since on $\mathbb{R}$, being closed and complete are equivalent.

Proof: First suppose M is not complete. Then there exists some Cauchy sequence $(a_n) \in M$ that doesn't converge. We know that (a_n) has at most one limit point, but since it doesn't converge, it can't have any (otherwise some subsequence would converge to the limit point, which would imply the whole sequence converges). Now suppose some subsequence of (a_n) converged to some point L. Then L would be a limit point, contradiction. Then the sequence (a_n) has no convergent subsequences, and thus M is not complete.

Now suppose M is not bounded. Thus there exists some $x \in M$ such that for all r, M is not contained in $B_r(x)$. Let a_n denote an element in M but not in $B_n(x)$. Then we have $d(a_n, x) \geq n$ for all $n \in \mathbb{N}$. Consider some subsequence $\left(a_{n_i}\right)$. For any $L \in M$, we have $d\left(a_{n_i}, x\right) \leq d\left(a_{n_i}, L\right) + d(L, x) \Rightarrow d\left(a_{n_i}, x\right) - d(L, x) \leq d\left(a_{n_i}, L\right)$. The second term in the left is constant, and the first term is unbounded. Thus the left is unbounded, which means $\left(a_{n_i}\right)$ cannot converge to L. This holds for any subsequence and any $L \in M$. Thus (a_n) has no convergent subsequence, so M is not compact. ■

Unfortunately, the other direction of Heine-Borel doesn't hold on general metric spaces.

Example: Consider $\mathbb{Z}$ with the discrete metric. Then it's both complete and bounded, but the sequence $1, 2, 3, \ldots$ has no convergent subsequence.

Thankfully, we have the following:

Theorem (Heine-Borel in Euclidean spaces): Let $(\mathbb{R}^n, d)$ be a Euclidean space with either the Euclidean metric, taxicab metric, or supnorm metric. Let E be a subset of $\mathbb{R}^n$. Then E is compact if and only if it's closed and bounded.

Proof: We already showed one direction in general, so now assume E is closed and bounded. Consider some sequence $(a_n) \in E$. Look at the sequence $\left(a_{n,1}\right) \in \mathbb{R}$ formed by the first coordinates of this sequence. Since E is bounded, this sequence of reals is bounded, and thus by Bolzano-Weierstrass, some subsequence converges to a real number. Now throw out every element in (a_n) whose first coordinate isn't part of this subsequence. Thus in the new sequence (a'_n), the first coordinate converges. Repeat this procedure for every other coordinate, and we obtain a subsequence of $(a_n) \in E$ that converges to some point in $\mathbb{R}^n$ (since everything we were doing was respect to $(\mathbb{R}^n, d)$). Thus the subsequence is Cauchy with respect to $(\mathbb{R}^n, d)$, and since all elements come from E, is Cauchy with respect to (E, d_E). Since $\mathbb{R}^n$ is complete and E is closed, E is complete as well. Since the subsequence is Cauchy in (E, d_E), it therefore must converge in (E, d_E). Thus E is compact. ■

We can get a stronger version of Heine-Borel by replacing bounded with totally bounded.

Definition (totally bounded): A metric space (X, d) is *totally bounded* if for every $\varepsilon > 0$, there exists a finite number of balls $B_\varepsilon(x_1), \ldots, B_\varepsilon(x_n)$ that cover X.

Example: The set $\{1, 2, \ldots\}$ with the discrete metric is not totally bounded, since for $\varepsilon = \frac{1}{2}$, a ball centered at a point in the set only contains that point, so we can't cover the set with finitely many balls.

Theorem: A metric space (M, d) is compact if and only if it's complete and totally bounded.

Proof: We previously showed that compact must implies complete, so suppose M is not totally bounded. Then there exists ε such that no finite set of ε balls can cover M. We now construct a sequence with no Cauchy subsequence, which contradicts compactness. Pick a point x_1 in M, and construct an ε ball around it. By the lack of total boundedness, there exists a point in M not covered by the ball. Let this point be x_2, and contruct another ε ball around it. Again by the lack of total boundedness, we can pick x_3 in M not covered by the balls. We can keep doing this and get a sequence (x_n), where between any two points, we have $d(x_i, x_j) \geq \varepsilon$, so clearly no subsequence is Cauchy.

Now suppose M is complete and totally bounded, and pick a sequence $(x_n) \in M$. From total boundedness, there are finitely many balls of radius 1 needed to cover M, so there must be a ball that contains infintely many terms of (x_n). Label this subsequence $(x_{n,1})$. Again by total boundedness, there exists finitely many balls of size $\frac{1}{2}$ that cover M, so there exists a subsequence $(x_{n,2})$ of $(x_{n,1})$ such that all the terms are contained in a single ball of size $\frac{1}{2}$. We keep doing this, and consider the sequence $(x_{n,n})$. Since $x_{j,j}$ comes from the sequence $(x_{n,j-1})$, the terms $x_{j,j}$ and $x_{j-1,j-1}$ are contained in a ball of radius $\frac{1}{j}$. Since $x_{j+k,j+k}$ all come from the sequence $(x_{n,j})$, there also are in this ball of radius j. Thus, for any $k, l > 0$, we have $d(x_{j+k,j+k}, x_{j+l,j+l}) < \frac{2}{j}$. This holds for any j, so we've produced a Cauchy sequence, and by completeness, this sequence converges. Thus every sequence in M has a convergent subsequence, which means M is compact, as desired. ■

8.3.1. Topological Compactness for Metric Spaces

This is the first definition given for compactness in $\mathbb{R}$. In fact, it's equivalent to the sequential definition of compactness for metric spaces.

Theorem (sequential compactness implies topological compactness): Let (X, d) be a metric space, and let Y be a compact subset of X. Let $(V_\alpha)_{\alpha \in X}$ be a collection of open sets in X, and suppose that

$$Y \subseteq \bigcup_{\alpha \in A} V_\alpha.$$

Then there exists a finite subset F of A such that

$$Y \subseteq \bigcup_{\alpha \in F} V_\alpha.$$

Solution: Suppose for the sake of contradiction a finite subcover didn't exist. Pick $y \in Y$, and note that $B_{(X,d)}(y, r) \subseteq V_\alpha$ for some nonzero r from openness. Let

$$r(y) = \sup\left\{r : B_{(X,d)}(y, r) \subseteq V_\alpha \text{ for some } \alpha \in A\right\}.$$

for all $y \in Y$. Since r is nonzero, $r(y) > 0$. Now let

$$r_0 = \inf\{r(y) : y \in Y\}.$$

We have three cases: $r_0 = 0$, $r_0 \in (0, \infty)$, or $r_0 = \infty$.

- **Case 1:** $r_0 = 0$. We can thus pick a sequence $(y_n) \in Y$ such that $r(y_n) < \frac{1}{n}$, which implies $\lim_{n\to\infty} r(y_n) = 0$. Since Y is compact, there exists a subsequence of $\left(y_{n_i}\right)$ which converges to $y_0 \in Y$.

  From the open cover, we know $y_0 \in V_\alpha$ for some α. Thus for some ε, $B_\varepsilon(y_0) \subseteq V_\alpha$. Thus from the limit, for some N we have that $i \geq N \Rightarrow y_{n_i} \in B_{\varepsilon/2}(y_0)$. Then if we consider $B_{\varepsilon/2}\left(y_{n_i}\right)$, from the triangle inequality we can see that $B_{\varepsilon/2}\left(y_{n_i}\right) \subseteq B_\varepsilon(y_0) \subseteq V_\alpha$. Thus $r\left(y_{n_i}\right) \geq \frac{\varepsilon}{2}$. This holds for all $i \geq N$, but that contradicts $r(y_n) \to 0 \Rightarrow r\left(y_{n_i}\right) \to 0$.
- **Case 2:** $0 < r_0 < \infty$. Thus $r(y) > r_0/2$ for all $y \in Y$, and so for every $y \in Y$, there exists $\alpha \in A$ such that $B_{r_0/2}(y) \subseteq V_\alpha$.

  We construct a sequence with no Cauchy subsequences, which implies that no subsequence can converge, giving us the desired contradiction. Pick some $y_1 \in Y$. Since $B_{r_0/2}(y_1)$ is an open subset of one of the sets in the cover, it clearly can't cover Y (since it would be a finite subcover), so there exists $y_2 \in Y \setminus B_{r_0/2}(y_1)$, and thus $d(y_1, y_2) \geq r_0/2$. Through similar reasoning as before, $B_{r_0/2}(y_1) \cup B_{r_0/2}(y_2)$ can't cover Y, so again there must be some y_3 outisde the two balls for which $d(y_1, y_3), d(y_2, y_3) \geq r_0/2$. Continuing in this fashion, we obtain a sequence with $d\left(y_i, y_j\right) \geq r_0/2$ for any i, j, and thus no subsequence can be Cauchy, as desired.
- **Case 3:** $r_0 = \infty$. Same as the previous case, just replace $r_0/2$ with 1.

Theorem (topological compactness implies sequential compactness): Let (X, d) be a metric space, and let Y be a subset of X. If every open cover of Y has a finite subcover, then Y is compact.

Proof: Suppose for the sake of contradiction that Y is not compact. Thus there exists a sequence with no convergent subsequence, which is equivalent to the sequence having no limit points in Y. What this implies that for each $y \in Y$, there exists ε_y such that $B_{\varepsilon_y}(y)$ contains only finitely many terms of the sequence (if there didn't, then for arbitrarily small ε, a ball would contain infinitely many terms of the sequence, which would mean there's a limit point).

Clearly, all of these balls cover Y, so by hypothesis there exists some finite subcover. However, since each ball contains only finitely many terms, taken together only finitely many terms of the sequence are covered, which is a contradiction. ■

Corollary: Let (X, d) be a metric space, and let $K_1, K_2, \ldots$ be a sequence of nonempty compact subsets of X such that

$$K_1 \supseteq K_2 \supseteq K_3 \supseteq \cdots.$$

Then the intersection $\bigcap_{n=1}^{\infty} K_n$ is nonempty.

Solution: We present two proofs: one using sequential compactness, and one using topological compactness.

From each K_n, pick a number k_n. Thus we have a sequence (k_n), so by compactness, it has a convergent subsequence with limit L. Now consider K_i for some i. Clearly it must contain $(k_n)_{n \geq i}$, so the same subsequence must also be contained in K_i (minus finitely many initial terms). Thus by compactness, $L \in K_i$. This holds for all i, so indeed the intersection is nonempty.

8.4. Problems

Problem: Let (x_n) and (y_n) be two sequences in (M, d). Suppose $(x_n) \to x \in M$ and $(y_n) \to y \in M$. Show that $\lim_{n\to\infty} d(x_n, y_n) = d(x, y)$.

Solution: From the triangle inequality, we have

$$\begin{aligned} d(x_n, y_n) &\leq d(x_n, x) + d(x, y_n) \leq d(x_n, x) + d(y_n, y) + d(y, x), \\ d(x, y) &\leq d(x_n, x) + d(x_n, y) \leq d(x_n, x) + d(y_n, y) + d(y_n, x_n). \end{aligned}$$

Thus $|d(x_n, y_n) - d(x, y)| \leq d(x, x_n) + d(y, y_n)$. The right side gets arbitrarily close to zero for large n, so we're done.

Problem: Let M be a metric space, and let (K_n) be a nested decreasing sequence of compact sets in M. Since K_n is compact, $\ell_n := \operatorname{diam}(K_n) = \max_{x,y \in K_n} d(x, y)$ exists. Let $K = \bigcap_{n=1}^{\infty} K_n$. Then we have

$$\lim_{n\to\infty} \ell_n = \operatorname{diam}(K).$$

Proof: First we note that since $K_{n+1} \subseteq K_n$, $\ell_{n+1} \leq \ell_n$, so the sequence is decreasing. Since it's clearly bounded below by 0, we it converges to some limit ℓ by the monotone convergence theorem. We need to show that $\operatorname{diam}(K) = \ell$.

Since any $x, y \in K$ are automatically in K_n, we have $d(x, y) \leq \ell_n$ for all $n \in N$ and all $x, y \in K$. Thus $\operatorname{diam}(K) \leq \ell_n$ for all n, and taking the limit yields $\operatorname{diam}(K) \leq \ell$.

For each K_n, there exist $a_n, b_n \in K_n$ such that $d(a_n, b_n) = \ell_n$ by compactness. By compactness, there exists a subsequence of (a_n) that converges in K_1, say $\left(a_{n_i}\right)$. Again by compactness, there exists a subsequence of $\left(b_{n_i}\right)$ that converges in K_1. Thus, there exists a subsequence of (a_n) and (b_n) such that both contain the same indexed elements and both converge into K_1. Let the this common subsequence be $\left(a_{n_j}\right), \left(b_{n_j}\right)$, with limits a and b respectively. Then clearly $a, b \in K_{n_m}$ for any m, since $\left(a_{n_j}\right)_{n_j \geq n_m}, \left(b_{n_j}\right)_{n_j \geq n_m} \in K_{n_m}$, and the sequences are just missing finitely many starting terms. Since a and b are in infintely many K_n, they must be in K. Then from $d(a, b) = \lim_{j\to\infty} d\left(a_{n_j}, b_{n_j}\right) = \lim_{j\to\infty} \ell_{n_j} = \ell$, we see that $\operatorname{diam}(K) \geq \ell$ ∎

Problem: Suppose that $f : M \to N$ for metric spaces M, N satisfies two conditions:
a) For each compact $K \subseteq M$, $f(K)$ is compact.
b) For every nested decreasing sequence of compact sets $(K_n) \in M$,

$$f\left(\bigcap_{n=1}^{\infty} K_n\right) = \bigcap_{n=1}^{\infty} f(K_n).$$

Prove that f is continuous.

Solution: Pick $c \in M$, and let $K_n = \overline{B_{\frac{1}{n}}(c)}$. Clearly $\bigcap_{n=1}^{\infty} K_n = \{c\}$, so we have

$$\{f(c)\} = \bigcap_{n=1}^{\infty} f(K_n).$$

Since each K_n here is closed and bounded, they're clearly compact, and since $K_{n+1} \subseteq K_n$, we have $f(K_{n+1}) \subseteq f(K_n)$. Thus $(f(K_n))$ is a nested decreasing sequence of compact sets. Since $\operatorname{diam}(\{f(c)\}) = 0$, from the previous result we must have that $\operatorname{diam}(f(K_n)) \to 0$ as $n \to \infty$.

Now pick $\varepsilon > 0$. There exists N such that $n \to N$ implies $\operatorname{diam}(f(K_N)) < \varepsilon$. Thus the distance between any two points in K_n is less than ε. In particular, we have $x \in K_N \Rightarrow d_N(f(x), f(c)) < \varepsilon$. Since $K_N = \overline{B_{\frac{1}{N}}(c)}$, we have that

$$d_M(x, c) < \frac{1}{2N} \Rightarrow d_N(f(x), f(c)) < \varepsilon.$$

This works for arbitrary ε, so f is continuous at c. Since c was also arbitrary, f is indeed continuous.

9. Continuous Functions on Metric Spaces

9.1. Continuous Functions

Almost everything from $\mathbb{R}$ transfers over.

Definition (continuous): Let (X, d_X) and (Y, d_Y) be metric spaces, and let $f : X \to Y$ be a function. We say that f *is continuous at* x_0 if for every $\varepsilon > 0$, there exists δ such that $d_Y(f(x), f(x_0)) < \varepsilon$ whenever $d(x, x_0) < \delta$. We say f is *continous* if it's continuous at every point.

Proposition: Let (X, d_X) and (Y, d_Y) be metric spaces, let $f : X \to Y$ be a function, and let $x_0 \in X$. Then the following are equivalent:

a) f is continuous at x_0.
b) If $(x_n) \in X$ converges to x_0 with respect to d_X, then $(f(x_n)) \in Y$ converges to $f(x_0)$ with respect to d_Y.
c) For every open set $V \subseteq Y$ that contains $f(x_0)$, there exists an open set $U \subseteq X$ containing x_0 such that $f(U) \subseteq V$.

Proof: First suppose a) is true, and let $\varepsilon > 0$. Then by continuity, there exists $\delta > 0$ such that $d_X(x, x_0) < \delta \Rightarrow d_Y(f(x), f(x_0)) < \varepsilon$. Since $x_n \to x$, we know that there exists N such that $n \geq N \Rightarrow d_X(x_n, x_0) < \delta$, and thus for all $n \geq N$, we have $d_Y(f(x_n), f(x_0)) < \varepsilon$. This holds for arbitrary ε, so we indeed have $f(x_n) \to f(x_0)$.

We show b) $\Rightarrow$ c) through the contrapositive. Thus for some open set $V \subseteq Y$ that contains $f(x_0)$, every open set $U \subseteq X$ that contains x_0 has image not necessarily contained in V. Consider $B_X\left(x_0, \frac{1}{n}\right)$. By the hypothesis, there exists a point x_n in this ball such that $f(x_n) \notin V$. Thus we have a sequence (x_n) which clearly converges to x_0, but where its image has terms only outside V. Thus $(f(x_n))$ must converge to the exterior or boundary of V. However, since V is open, $f(x_0)$ must be in its interior, contradiction.

Pick $\varepsilon > 0$, and consider $B_Y(f(x_0), \varepsilon) \subseteq Y$. By hypothesis, there exists an open set $U \subseteq X$ that contains x_0 such that $f(U) \subseteq B_Y(f(x_0), \varepsilon)$. Since U is open, there exists some $\delta > 0$ such that $B_X(x_0, \delta) \subseteq U$. Thus we have $f(B_X(x_0, \delta)) \subseteq f(U) \subseteq B_Y(f(x_0, \varepsilon))$. Thus we have $d_X(x, x_0) < \delta \Rightarrow d_Y(f(x), f(x_0)) < \varepsilon$. This holds for every ε, so f is continuous at x_0. ∎

Proposition: Let (X, d_X) and (Y, d_Y) be metric spaces, and let $f : X \to Y$ be a function. Then the following are equivalent:

a) f is continuous.
b) Whenever $(x_n) \in X$ converges with respect to d_X, $(f(x_n)) \in Y$ converges with respect to d_Y.
c) Whenever V is an open set in Y, the set $f^{-1}(V)$ is an open set in X.
d) Whenever V is a closed set in Y, the set $f^{-1}(V)$ is a closed set in X.

Proof: a) and b) are equivalent easily by the last proposition. We can show a) implies c) by just taking unions of open sets, which will also be open. Similarly we can show that c) implies a) by applying the previous proposition to every point. For c) implies d), take the complement of a closed set, which is open, then apply c), and then take the inverse images complement, which must then be closed. Do the same thing by in reverse for d) implies c). ■

Proposition (composition preserves continuity): Let X, Y, and Z be metric spaces with their associated metrics.

a) If $f : X \to Y$ is continuous at $x_0 \in X$, and $g : Y \to Z$ is continuous at $f(x_0)$, then $g \circ f : X \to Z$ is continuous at x_0.
b) If f and g are continuous, then $g \circ f$ is continuous.

Proof: Suppose $(x_n) \in X$ converges to x_0. Then by continuity, $(f(x_n)) \in Y$ converges to $f(x_0)$, but again by continuity, $(g(f(x_n))) \in Z$ converges to $g(f(x_0))$, so we have the desired conclusion. b) then easily follows. ■

Proposition: If $f : \mathbb{R}^2 \to \mathbb{R}$ is continuous at (a, b), then

$$\begin{aligned} f(a,b) &= \lim_{x\to a} \limsup_{y\to b} f(x,y) = \lim_{y\to b} \limsup_{x\to a} f(x,y) \\ &= \lim_{x\to a} \liminf_{y\to b} f(x,y) = \lim_{y\to b} \liminf_{x\to a} f(x,y), \end{aligned}$$

where $\limsup_{x\to x_0} f(x) = \inf_{r>0} \sup_{|x-x_0|<r} f(x) = \lim_{r\to 0} \sup_{|x-x_0|<r} f(x)$ and similarly for lim inf.

Remark: The last equivalence for lim sup comes from noting that $\sup_{|x-x_0|<r} f(x)$ decreases as r decreases

Proof: We simply do the first equality, as the rest follow similarly. Pick $\varepsilon > 0$. From continuity, we have that for some δ, $\|(x, y) - (a, b)\| < \delta \Rightarrow |f(x, y) - f(a, b)| < \varepsilon$. Then for $x \in \left(a - \frac{\delta}{2}, a + \frac{\delta}{2}\right), y \in \left(b - \frac{\delta}{2}, b + \frac{\delta}{2}\right)$ (since then $\|(x, y) - (a, b)\| < \frac{\delta}{\sqrt{2}} < \delta$), we have $f(a, b) - \varepsilon < f(x, y) < f(a, b) + \varepsilon$. Thus, $f(a, b) - \varepsilon \leq \sup_{|y-b|<\frac{\delta}{2}} f(x, y) \leq f(a, b) + \varepsilon$, which then implies $f(a, b) - \varepsilon \leq \limsup_{y\to b} f(x, y) \leq f(a, b) + \varepsilon$.

Now note that for all $x \in \left(a - \frac{\delta}{2}, a + \frac{\delta}{2}\right)$, we have that $\left|\limsup_{y \to b} f(x, y) - f(a, b)\right| < \varepsilon$. Since this holds for arbitrary ε, we have the desired limit. ■

Corollary: If $f : \mathbb{R}^2 \to \mathbb{R}$ is continuous at (a, b) and the one sided limits both exist, then

$$\lim_{x \to a} \lim_{y \to b} f(x, y) = \lim_{y \to b} \lim_{x \to a} f(x, y) = f(a, b).$$

9.2. Continuity and Product Spaces

9.3. Compactness and Connectedness

Proposition: Let $f : X \to Y$ be a continuous function, and suppose $K \subseteq X$ is compact. Then $f(K)$ is compact.

Proof: If $(y_n) \in f(K)$, consider a sequence $(x_n) \in K$ such that $f(x_n) = y_n$. Since K is compact, some subsequence of (x_n) converges to $x_0 \in K$. Then by continuity, the image of this subsequence converges to $f(x_0) \in f(K)$. Thus (y_n) has a convergent subsequence, so $f(K)$ is compact. ■

Theorem (extreme value theorem on metric spaces): Suppose (X, d_X) is a compact metric space, and let $f : X \to \mathbb{R}$ be a continuous function. Then f is bounded and has a maximum and minimum.

Proof: Since X is compact, the image $f(X)$ is compact by the previous proposition, which then implies the image is closed and bounded. Consider $\inf f(X)$. There must be a sequence contained in $f(X)$ that converges to $\inf f(X)$, and thus by closedness, we must have $\inf f(X) \in f(X)$, so f attains a minimum. The maximum case follows similarly. ■

Definition (uniform continuity): Let $f : X \to Y$. We say f is *uniformly continuous* if for every $\varepsilon > 0$, there exists $\delta > 0$ such that $d_Y(f(a), f(b)) < \varepsilon$ whenever $d_X(a, b) < \delta$.

Proposition (sequential formulation of uniform continuity): Let $(a_n), (b_n) \in X$ such that $\lim_{n \to \infty} d(a_n, b_n) = 0$. Then $f : X \to Y$ is uniformly continuous if and only if $\lim_{n \to \infty} d(f(a_n), f(b_n)) = 0$.

Proof: Same as proof for $\mathbb{R}$. ■

Proposition: Let X and Y be metric spaces and suppose X is compact. Then $f : X \to Y$ is continuous if and only if it's uniformly continuous.

Proof: Again same as proof for $\mathbb{R}$. ∎

Definition (connected): Let (X, d) be a metric space. We say X is *disconnected* if there exist open sets $V, W \in X$ such that V and W are disjoint and $V \cup W = X$. We say X is *connected* if and only if it's nonempty and not disconnected. If Y is a subset of X, then Y if connected if $(Y, d|_{Y \times Y})$ is connected.

Proposition: Suppose $f : X \to Y$ is a continuous function, and let E be a connected subset of X. Then $f(E)$ is connected.

Proof: We prove the contrapositive. Suppose $f(E)$ is not connected. Then exist two open sets $V, W \in Y$ that are disjoint and such that $V \cup W = f(E)$. Then by continuity, the sets $f^{-1}(V)$ and $f^{-1}(W)$ are open in X. Since V and W are disjoint, these new sets are also disjoint. Furthermore, the union of the two must contain all points in X, since otherwise their images wouldn't jointly cover $f(E)$. Thus E is disconnected, as desired. ∎

9.4. Contraction Mapping Theorem

Definition (contraction): Let (X, d) be a metric space, and let $f : X \to X$ be a map. We say that f is a *contraction* if we have $d(f(x), f(y)) \leq d(x, y)$ for all $x, y \in X$. We say that f is a *strict contraction* if there exists $0 < c < 1$ such that $d(f(x), f(y)) \leq cd(x, y)$ for all $x, y \in X$.

The below theorem is also known as the Banach fixed point theorem.

Theorem (contraction mapping theorem): Let (X, d) be a metric space, and let $f : X \to d$ be a strict contraction. Then f can have at most one fixed point. Moreover, if X is nonempty and complete, then f has exactly one fixed point.

Proof: Suppose f has two fixed points $p, q \in X$. Then $d(p, q) = d(f(p), f(q)) \leq cd(p, q)$, which implies $d(p, q) = 0 \Rightarrow p = q$. Thus f can only have at most one fixed point.

Now suppose X is nonempty and complete. Pick $x \in X$, and let $x_0 = x, x_n = f(x_{n-1})$. We show that (x_n) is Cauchy, and since it's complete it has limit x. Then we have

$$f(x) = f\left(\lim_{n\to\infty} x_n\right) = \lim_{n\to\infty} f(x_n) = \lim_{n\to\infty} x_{n+1} = x,$$

where we could bring the limit out since f is a contraction, and thus continuous.

Note that we have $d(x_{n+1}, x_n) \le c^n d(x_1, x_0)$. Then for any $n \ge m \ge 1$, we have

$$\begin{aligned} d(x_n, x_m) \le d(x_n, x_{n-1}) + \dots + d(x_{m+1}, x_m) &\le d(x_1, x_0)(c^{n-1} + \dots + c^m) \\ &= d(x_1, x_0) \cdot c^m \frac{1 - c^n}{1 - c} \\ &\le d(x_1, x_0) \cdot \frac{c^m}{1 - c}. \end{aligned}$$

Since $c < 1$, the right side gets arbitrarily small for large m, so the sequence is indeed Cauchy. ■

9.5. Homeomorphisms

Definition (homeomorphic): Let (M, d_M) and (N, d_N) be metric spaces. Then M and N are *homeomorphic* if there exists a continuous bijection $f : M \to N$ with continuous inverse. If such a function exists, then it's called a *homeomorphism*.

Example: $(-1, 1)$ is homeomorphic to $\mathbb{R}$ via the homeomorphism $f(x) = \tan(\frac{\pi x}{2})$, which has continuous inverse $f^{-1}(x) = \frac{2}{\pi} \arctan(x)$.

Example: Being continuous doesn't gurantee that the inverse is continuous. Consider a function from $[0, 2\pi)$ to the circle, where f takes $\theta \in [0, 2\pi)$ and maps it to $e^{i\theta}$ on the unit circle. This is clearly a bijection, and continuous in one direction. However, the inverse function is not continuous, as if we approach 1 on the unit circle from below, the inverse functions output approached 2π, not 0.

Proposition: If M is compact, then a continuous bijection $f : M \to N$ is a homeomorphism.

Proof: We just need to show that the inverse is continuous. Suppose $q_n \to q$ in N. We need to show that $p_n = f^{-1}(q_n)$ converges to $p = f^{-1}(q)$ in M.

Suppose not for the sake of contradiction. Thus there's some subsequence $\left(p_{n_k}\right)$ such that $d_M\left(p_{n_k}, p\right) \ge \delta$ for some $\delta > 0$. Since M is compact, a subsequence of this subsequence, $\left(p_{n_{k(\ell)}}\right)$, converges to $p' \in M$. Clearly we have that $d_M(p, p') \ge \delta$, so $p \ne p'$.

Since f is continuous, we have

$$f\left(p_{n_{k(\ell)}}\right) \to f(p')$$

as $\ell \to \infty$. However, we also have

$$f\left(p_{n_{k(\ell)}}\right) = q_{n_{k(\ell)}} \to q = f(p).$$

Thus $f(p) = f(p')$, which contradicts f being a bijection. ■

9.6. Problems

Problem: Let (X, d) be a complete metric space, and let $f : X \to X$ and $g : X \to X$ be strict contractions with contractions coefficients c and c' respectively. By the fixed point theorem, f and g have unique fixed points x_0 and y_0 respectively. Suppose that $d(f(x), g(x)) \leq \varepsilon$ for all $x \in X$. Show that $d(x_0, y_0) \leq \frac{\varepsilon}{1-\min(c,c')}$.

Solution: For $x, y \in X$, we have

$$d(f(x), g(y)) \leq d(f(x), f(y)) + d(f(y), g(y)) \leq cd(x, y) + \varepsilon$$

and

$$d(f(x), g(y)) \leq d(f(x), g(x)) + d(g(x), g(y)) \leq \varepsilon + c'd(x, y).$$

Thus $d(f(x), g(y)) \leq \varepsilon + \min(c, c')d(x, y)$. Letting $x = x_0$ and $y = y_0$ yields

$$d(x_0, y_0) = d(f(x_0), g(y_0)) \leq \varepsilon + \min(c, c')d(x_0, y_0) \Rightarrow d(x_0, y_0) \leq \frac{\varepsilon}{1 - \min(c, c')},$$

as desired.

10. Lebesgue Integration

Before we can define Lebesgue integration, we need just a touch of measure theory.

10.1. Lebesgue Measure

Measure is a way to generalize volume of subsets of $\mathbb{R}^n$ to sets that aren't so simple to ascribe volume to. Ideally, for every $\Omega \subseteq \mathbb{R}^n$, we want to assign a value $m(\Omega) \in [0, +\infty]$ such that the following intuitive properties hold:

a) Empty set: $m(\emptyset) = 0$.
b) Monotonicity: If $A \subseteq B$, then $m(A) \leq m(B)$.
c) Countable sub-additivity: If $\left(A_j\right)_{j \in J}$ is a countable collection of sets, then $m\left(\bigcup_{j \in J} A_j\right) \leq \sum_{j \in J} m\left(A_j\right)$.
d) Countable additivity: If $\left(A_j\right)_{j \in J}$ is a countable collection of disjoint sets, then $m\left(\bigcup_{j \in J} A_j\right) = \sum_{j \in J} m\left(A_j\right)$.
e) Normalization: $m([0,1]^n) = 1$, where $[0,1]^n$ is the unit cube in $\mathbb{R}^n$.
f) Translation invariance. For any $\Omega \subseteq \mathbb{R}^n$ and $x \in \mathbb{R}^n$, we have $m(x + \Omega) = m(\Omega)$.

However, it turns out such a measure does not exist, and we will see an example of measure failing later. However, if we restrict our attention to a collection of subsets that we measurable based on a certain condition, then we can get of the properties above. The collection of measurable sets should also satisfy a few intuitive and useful properties that we hope would hold:

a) Complementarity: If Ω is measurable, then $\mathbb{R}^n \setminus \Omega$ is also measurable.
b) Borel property: If Ω is open in $\mathbb{R}^n$, then it's measurable (note that combining this with the above property also implies that every closed set is measurable).
c) σ-algebra property: If $\left(\Omega_j\right)_{j \in J}$ is a countable collection of measurable sets, then $\bigcup_{j \in J} \Omega_j$ and $\bigcap_{j \in J} \Omega_j$ are also measurable.

10.1.1. Outer Measure

Definition (open box): An *open box* B is $\mathbb{R}^n$ is any set of the form

$$B = (a_1, b_1) \times \cdots (a_n, b_n),$$

where $b_i \geq a_i$ are real numbers. We define the volume $\operatorname{vol}(B)$ to be

$$\operatorname{vol}(B) = \prod_{i=1}^{n} (b_i - a_i).$$

Definition (outer measure): If $\Omega \subseteq \mathbb{R}^n$, then the *outer measure* $m^*(\Omega)$ (sometimes denoted $m_n^*(\Omega)$ to emphasize we're in $\mathbb{R}^n$) is given by

$$m^*(\Omega) = \inf\left\{\sum_{j\in J} \mathrm{vol}(B_j) : (B_j)_{j\in J} \text{ covers } \Omega \text{ and is countable}\right\}.$$

Proposition: Outer measure satisfies the empty set property, positivity, monotonicity, countable sub-additivity, and translation invariance.

Proof: Consider $\left(-\sqrt[n]{\varepsilon}/2,\ \sqrt[n]{\varepsilon}/2\right)^n$. This clearly covers $\emptyset$, and has volume ε. Since ε is arbitrary, the infinum of the cover is 0, so $m^*(\emptyset) = 0$.

Positivity follows since $\mathrm{vol}(B) \geq 0$ for any box B, so the volume of a cover is at least 0.

Since any cover of B is a cover of $A \subseteq B$, the set of all covers of B is a subset of the set of all covers of A. Since $\inf Y \leq \inf X$ for $X \subseteq Y$, it follows that $m * (A) \leq m^*(B)$, so monotonicity holds.

Let (A_j) be a sequence of subsets of $\mathbb{R}^n$. By infinum properties, for each A_j there exists a covering $(B_k)_j$ with $\sum \mathrm{vol}(B_k) \leq m^*(A_j) + \frac{\varepsilon}{2^j}$. Since clearly the union of the covers will cover the union of (A_j), we have that $m^*\left(\cup_{j=1}^{\infty} A_j\right) \leq \sum_{j=1}^{\infty} m^*(A_j) + \varepsilon$. Taking $\varepsilon \to 0$ yields countable sub-additivity.

Pick $x \in \mathbb{R}^n$. For any cover $(B_j)_{j\in j}$ of Ω, note that $(x + B_j)_{j\in J}$ covers $x + \Omega$ and vice versa. Thus the set of volumes of the coverings of these sets is the same, so $m^*(\Omega) = m^*(x + \Omega)$. ■

Proposition: For any closed box

$$B = \prod_{i=1}^{n} [a_i, b_i],$$

we have

$$m^*(B) = \prod_{i=1}^{n} (b_i - a_i).$$

Proof: Note that clearly $\prod_{i=1}^n (a_i - \varepsilon, b_i + \varepsilon)$ covers B for every $\varepsilon > 0$. Thus $m^*(B) \leq \prod_{i=1}^n (b_i - a_i + 2\varepsilon)$. Letting $\varepsilon \to 0$ yields $m^*(B) \leq \prod_{i=1}^n (b_i - a_i)$.

Now we need to show that $m^*(B) \geq \prod_{i=1}^n (b_i - a_i)$. We proceed by induction on n.

First $n = 1$, so we're looking at the interval $[a, b]$ with $a \leq b$. Since this interval is compact, any covering will have a finite subcover, and since the finite cover will clearly have less volume than the original cover, we just need to

$$\sum_{j\in J} \mathrm{vol}(B_j) \geq b - a.$$

Let $B_j = (a_j, b_j)$, and let $f_j(x) = 1$ when $x \in B_j$ and 0 otherwise. Then we have

$$\int_{-\infty}^{\infty} f_j(x)\, dx = b_j - a_j = \operatorname{vol}(B_j).$$

Summing over all j, and since J is finite, we can swap the sum and integral to obtain

$$\int_{-\infty}^{\infty} \left(\sum_{j \in J} f_j(x)\right) dx = \sum_{j \in J} \operatorname{vol}(B_j).$$

Since (B_j) covers $[a, b]$, clearly $\sum_{j \in J} f_j(x) \geq 1$ for $x \in [a, b]$ for all j. Since the f_j are also nonnegative, we have

$$\int_{-\infty}^{\infty} \left(\sum_{j \in J} f_j(x)\right) dx \geq \int_a^b 1\, dx = b - a.$$

Thus we have the desired inequality.

Now suppose the claim holds up till $n - 1$. Again by compactness, we only need to prove that

$$\sum_{j \in J} \operatorname{vol}\left(B^{(j)}\right) \geq \prod_{i=1}^{n} (b_i - a_i)$$

holds for finite covers of $B \in \mathbb{R}^n$. Write each $B^{(j)}$ as $A^{(j)} \times \left(a_n^{(j)}, b_n^{(j)}\right)$, where $A^{(j)}$ is the box that's the projection of $B^{(j)}$ into $\mathbb{R}^{n-1}$. We have

$$\operatorname{vol}\left(B^{(j)}\right) = \operatorname{vol}_{n-1}\left(A^{(j)}\right)\left(b_n^{(j)} - a_n^{(j)}\right).$$

Similarly, we have that

$$\operatorname{vol}(B) = \operatorname{vol}_{n-1}(A)(b_n - a_n)$$

for a similarly defined A.

Similarly to before, for each $j \in J$ define $f^{(j)}(x) = \operatorname{vol}_{n-1}\left(A^{(j)}\right)$ for all $x \in \left(a_n^{(j)}, b_n^{(j)}\right)$ and 0 otherwise. Then

$$\int_{-\infty}^{\infty} f^{(j)}(x)\, dx = \operatorname{vol}_{n-1}\left(A^{(j)}\right)\left(b_n^{(j)} - a_n^{(j)}\right) = \operatorname{vol}\left(B^{(j)}\right).$$

Then summing over j and swapping yields

$$\sum_{j \in J} \operatorname{vol}\left(B^{(j)}\right) = \int_{-\infty}^{\infty} \left(\sum_{j \in J} f^{(j)}(x)\right) dx.$$

Consider $(x_1, ..., x_{n-1}) \in A$. Then for any $x_n \in [a_n, b_n]$, clearly $(x_1, ..., x_n) \in B^{(j)}$ for some j, which implies that $(x_1, ..., x_{n-1}) \in A^{(j)}$. In particular, the collection $\left(A^{(j)}\right)$ covers A. Thus by the inductive hypothesis, we have

$$\sum_{j \in J} \operatorname{vol}_{n-1}\left(A^{(j)}\right) \geq \operatorname{vol}_{n-1}(A).$$

Thus again similarly to before, we have

$$\int_{-\infty}^{\infty}\left(\sum_{j\in J} f^{(j)}(x)\right)dx \geq \int_{a_n}^{b_n}\left(\sum_{j\in J}\mathrm{vol}_{n-1}\left(A^{(j)}\right)\right)dx \geq \int_{a_n}^{b_n}\mathrm{vol}_{n-1}(A)\,dx = \mathrm{vol}_{n-1}(A)(b_n - a_n)$$
$$= \mathrm{vol}(B).$$

Thus we have our desired inequality. ■

Corollary: For any open box, we have

$$m^*(B) = \prod_{i=1}^{n}(b_i - a_i).$$

Proof: We have

$$\prod_{i=1}^{n}[a_i + \varepsilon, b_i - \varepsilon] \subseteq \prod_{i=1}^{n}(a_i, b_i) \subseteq \prod_{i=1}^{n}[a_i, b_i].$$

Thus the previous result and monotonicity yields

$$\prod_{i=1}^{n}(b_i - a_i - 2\varepsilon) \leq m^*(B) \leq \prod_{i=1}^{n}(b_i - a_i).$$

Letting $\varepsilon \to 0$ and using the squeeze theorem yields the desired result. ■

10.1.2. Failure of Outer Measure

The reason outer measure doesn't work is that if we assume it has countable additivity. In fact, the following proposition shows that for any measure that satisfies all the properties listed above, we obtain a contradiction.

Proposition: There exists a countable collection $\left(A_j\right)$ of disjoint subsets of $\mathbb{R}$ such that $m^*\left(\bigcup_{j=1}^{\infty} A_j\right) \neq \sum_{j=1}^{\infty} m^*\left(A_j\right)$.

Proof:

■

Proposition: There exists a finite collection of disjoint sets for which additivity fails.

Proof:

■

10.1.3. Measurable Sets

Definition (Lebesgue measurability): Let E be a subset of $\mathbb{R}^n$. Then E is *Lebesgue measurable* if for every $A \subseteq \mathbb{R}^n$, we have

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

If E is measurable, then we define the Lebesgue measure of E to be $m(E) = m^*(E)$.

Essentially what this definition does it force the collection of measurable sets to satisfy additivity. Now we show that the Lebesgue measure satisfies all the properties we want, plus a few other useful ones.

Proposition: The empty set, $\mathbb{R}^n$, and the upper (and lower) half space are measurable.

Proof: For any $A \subseteq \mathbb{R}^n$, we have

$$m^*(A \cap \emptyset) + m^*(A \backslash \emptyset) = m^*(\emptyset) + m^*(A) = m^*(A)$$

and

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

Thus $\emptyset$ and $\mathbb{R}^n$ are measurable.

To prove the the upper half space is measurable, we need the following lemma:

Lemma: If A is an open box in $\mathbb{R}^n$, and E is the upper half space $\{(x_1, ..., x_n) \in \mathbb{R}^n : x_n > 0\}$, then $m^*(A) = m^*(A \cap E) + m^*(A \backslash E)$.

Proof: Let $A = A' \times (a_n, b_n)$, where A' is an open box in $\mathbb{R}^{n-1}$. If $b_n \leq 0$ or $a_n > 0$, then the claim obviously follows, so suppose $a_n \leq 0 < b_n$. Then $A \cap E = A' \times [0, b_n)$ and $A \backslash E = A' \times (a_n, 0]$. We have

$$m^*(A' \times [0, b_n)) \leq m^*(A' \times \{0\}) + m^*(A' \times (0, b_n))$$

by sub-additivity. Note that the open box $A' \times \left(-\frac{\varepsilon}{\mathrm{vol}_{n-1}(A')}, \frac{\varepsilon}{\mathrm{vol}_{n-1}(A')}\right)$ covers $A' \times \{0\}$ and has volume 2ε. Since ε is arbitrary, $m^*(A' \times \{0\}) = 0$. Thus

$$m^*(A' \times [0, b_n)) \leq m^*(A' \times \{0\}) + m^*(A' \times (0, b_n)) = \mathrm{vol}_{n-1}(A')b_n.$$

Similarly, we have

$$m^*(A' \times (a_n, 0]) \leq -\mathrm{vol}_{n-1}(A')a_n.$$

Thus we have

$$m^*(A \cap E) + m^*(A \backslash E) \leq \mathrm{vol}_{n-1}(A')b_n - \mathrm{vol}_{n-1}(A')a_n = \mathrm{vol}(A) = m^*(A).$$

From sub-additivity, we also obtain

$$m^*(A) \leq m^*(A \cap E) + m^*(A \backslash E).$$

Thus we have the desired claim. ■

Now let A be an arbitrary subset of $\mathbb{R}^n$. Then there exists a cover of open boxes (B_j) such that $\sum m^*(B_j) = \sum \operatorname{vol}(B_j) \leq m^*(A) + \varepsilon$. Then clearly $(B_j \cap E)$ covers $A \cap E$ and $(B_j \backslash E)$ covers $A \backslash E$. Thus we have

$$m^*(A \cap E) + m^*(A \backslash E) \leq \sum m^*(B_j \cap E) + \sum m^*(B_j \backslash E) = \sum m^*(B_j) \leq m^*(A) + \varepsilon,$$

where the equality comes from the lemma, the first inequality follows from the definition of outer measure. Letting $\varepsilon \to 0$ yields $m^*(A \cap E) + m^*(A \backslash E) \leq m^*(A)$. Combining that with $m^*(A) \leq m^*(A \cap E) + m^*(A \backslash E)$, which follows from sub-additivity, yields the desired result.

Proposition (some properties of measurable sets):
a) If E is measure, then $\mathbb{R}^n \backslash E$ is also measurable.
b) Measurable sets are translation invariant.
c) The finite union and intersection of measurable sets are measurable, and disjoint sets are finitely additive.
d) Every open and closed box is measurable.
e) Any set E of outer measure zero is measurable.

Proof: If E is measurable, then for every $A \subseteq \mathbb{R}^n$, we have $m^*(A) = m^*(A \cap E) + m^*(A \backslash E)$. Note that $A \cup (\mathbb{R}^n \backslash E) = A \backslash E$ and $A \backslash (\mathbb{R}^n \backslash E) = A \cup E$. Thus replacing the sets in the equation yields that $\mathbb{R}^n \backslash E$ is measurable.

Translation invariance follows easily by just shifting A by the same value.

If we can show that the union and intersection of two measurable sets is measurable, then finite unions and intersections follow by induction. Let E_1, E_2 be measurable, and let A be arbitrary. By the measurability of E_1, we have

$$\begin{aligned} m^*(A \backslash (E_1 \cap E_2)) &= m^*((A \backslash (E_1 \cap E_2)) \cap E_1) + m^*((A \backslash (E_1 \cap E_2)) \backslash E_1) \\ &= m^*(A \cap E_1 \backslash E_2) + m^*(A \backslash E_1), \end{aligned}$$

where the equality comes from the corresponding sets being equal. Adding $m^*(A \cap E_1 \cap E_2)$ yields

$$\begin{aligned} m^*(A \cap E_1 \cap E_2) + m^*(A \backslash (E_1 \cap E_2)) &= m^*(A \cap E_1 \cap E_2) + m^*(A \cap E_1 \backslash E_2) + m^*(A \backslash E_1) \\ &= m^*(A \cap E_1) + m^*(A \backslash E_1) \\ &= m^*(A), \end{aligned}$$

where the two equalities follow from the measurability of E_2 and E_1 respectively. Thus $E_1 \cap E_2$ is measurable.

Now we show the union is measurable. Since E_1 and E_2 are measurable, by complementarity $\mathbb{R}^n \backslash E_1$ and $\mathbb{R}^n \backslash E_2$ are also measurable. Then $(\mathbb{R}^n \backslash E_1) \cap (\mathbb{R}^n \backslash E_2) = \mathbb{R}^n \backslash (E_1 \cup E_2)$ is measurable. Then by complementarity again, $\mathbb{R}^n \backslash (\mathbb{R}^n \backslash (E_1 \cup E_2)) = E_1 \cup E_2$ is measurable.

Now we show finite additivity. We show it for two sets, and then finite additivity follows from induction. Let E_1, E_2 be measurable and disjoint, and let $A \subseteq \mathbb{R}^n$ be arbitrary. Then by measurability of E_1, we have

$$\begin{aligned} m^*(A \cap (E_1 \cup E_2)) &= m^*(A \cap (E_1 \cup E_2) \cap E_1) + m^*(A \cap (E_1 \cup E_2) \backslash E_1) \\ &= m^*(A \cap E_1) + m^*(A \cap E_2), \end{aligned}$$

where the equality follows from disjointness. Then let $A = \mathbb{R}^n$. Since E_1 and E_2 are measurable, then union is as well, so m^* can be replaced with m, yielding $m(E_1 \cup E_2) = m(E_1) + m(E_2)$, as desired.

Suppose $\prod_{i=1}^n (a_i, b_i)$ is an open box. We know that the any half space of $\mathbb{R}^n$ is measurable, and then by translation invariance, the set $\{(x_1 ..., x_n) \in \mathbb{R}^n : x_i > a_i\}$ is also measurable. Similarly, $\{(x_1, ..., x_n) \in \mathbb{R}^n : x_i < b_i\}$ is also measurable. Taking the intersection over all i yields $\prod_{i=1}^n (a_i, b_i)$, and thus is measurable. For the closed box, we do the same procedure except with $\{(x_1 ..., x_n) \in \mathbb{R}^n : x_i < a_i\}$ and $\{(x_1, ..., x_n) \in \mathbb{R}^n : x_i > b_i\}$, take unions, then take the complement.

Suppose $m^*(E) = 0$. Then we have

$$m^*(A) \leq m^*(A \cup E) + m^*(A \backslash E) \leq m^*(E) + m^*(A) = m^*(A),$$

where the first inequality follows from finite sub-additivity and second follows from monotonicity. Thus we must have equality, so E is measurable. ■

Proposition (countable additivity): If (E_j) is a countable collection of disjoint measurable sets, then $\bigcup_{j=1}^\infty E_j$ is measurable, and $m\left(\bigcup_{j=1}^\infty E_j\right) = \sum_{j=1}^\infty m(E_j)$.

Proof: Let $E = \bigcup_{j=1}^\infty$ and let $A \subseteq \mathbb{R}^n$ be arbitrary. We need to show that

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

Note we have that

$$A \cap E = \bigcup_{j=1}^\infty (A \cap E_j),$$

so by countable sub-additivity, we have

$$m^*(A \cap E) \leq \sum_{j=1}^\infty m^*(A \cap E_j),$$

which we rewrite as

$$m^*(A \cap E) \leq \sup_{N \geq 1} \sum_{j=1}^N m^*(A \cap E_j).$$

Let $F_N = \bigcup_{j=1}^N E_j$. By finite additivity, we have

$$\sum_{j=1}^N m^*(A \cap E_j) = m^*(A \cap F_N).$$

Combining this with the previous inequality yields

$$m^*(A \cap E) \leq \sup_{N \geq 1} m^*(A \cap F_N).$$

Since $F_N \subseteq E$, we have $A \backslash E \subseteq A \backslash F_N$. Thus monotonicity implies $m^*(A \backslash E) \leq m^*(A \backslash F_N)$. Then we have

$$m^*(A \cap E) + m^*(A \backslash E) \leq \sup_{N \geq 1}(m^*(A \cap F_N) + m^*(A \backslash E)) \leq \sup_{N \geq 1}(m^*(A \cap F_N) + m^*(A \backslash F_N)).$$

By finite unions, we know that F_N is measurable for each N. Thus right side is just $\sup_{N \geq 1} m^*(A) = m^*(A)$. Combining this with monotonicity ($m^*(A) \leq m^*(A \cap E) + m^*(A \backslash E)$) yields the desired result. ■

Proposition: If $A \subseteq B$ are measurable, then $B \backslash A$ is also measurable, and

$$m(B \backslash A) + m(A) = m(B).$$

Proof: Let $C = \mathbb{R}^n \setminus B$, which is measurable by complementarity. Then $A \cup C$ is measurable, and thus $\mathbb{R}^n \backslash (A \cup C) = B \backslash A$ is also measurable. Then the equality follows by disjoint union. ■

Proposition: If (E_j) is a countable collection of measurable sets, then $\bigcup_{j=1}^{\infty} E_j$ and $\bigcap_{j=1}^{\infty}$ are also measurable.

Proof: Let $F_N = \bigcup_{j=1}^{N} E_j$, and let $G_N = F_N \backslash F_{N-1}$ for $N \geq 1$, where $F_0 = \emptyset$. Note that G_N contains every element in E_N that isn't in a E_j with smaller E_j. Thus (G_j) is a collection of disjoint sets. By finite unions, we know that F_N is measurable for each N, and from the previous proposition, we know that G_N is measurable. Then applying countable unions implies that $\bigcup_{j=1}^{\infty} G_j = \bigcup_{j=1}^{\infty} E_j$ is measurable. Countable intersection follows by taking the complement, taking countable unions, and taking the complement again. ■

Proposition: Every open and closed set is measurable.

Proof: Closed sets follow by complementarity, so we prove this for open sets. We claim that all open sets can be written as a countable union of open boxes. The result then follows from the fact that open boxes are measurable and the previous proposition.

Call a box rational if all its component intervals have rational endpoints. Note there are $\mathbb{Q}^{2n}$ such boxes, and thus countably many. Suppose $B_r(x)$ is an open ball, where $x = (x_1, ..., x_n)$. Then there exists rationals a_i, b_i such that

$$x_i - \frac{r}{n} < a_i < x_i < b_i < x_i + \frac{r}{n}.$$

Then the box $\prod_{i=1}^{n}(a_i, b_i)$ contains x. Note that the longest diagonal has length less than $\frac{2r}{\sqrt{n}}$, which follows from the Pythagorean theorem, so the box is contained in $B_r(x)$. Thus every open ball contains a rational box that contains the balls center.

Now let E be an open set, and let S be the set of all rational boxes that are subsets of E, and consider the union $E' = \bigcup_{B \in S} B$ of all of them. Clearly $E' \subseteq E$. Since E is open, for every $x \in E$, there exists an open ball $B_r(x)$ that is contained in E. By the previous paragraph, this ball contains a rational box that contains x, which implies $x \in E'$. Since this holds for all $x \in E$, we have $E \subseteq E'$, and thus $E = E'$, as desired. ■

10.2. Measurable Functions

Definition (measurable function): Let Ω be a measurable subset of $\mathbb{R}^n$, and let $f : \Omega \to \mathbb{R}^m$ be a function. A function f is *measurable* if and only if $f^{-1}(V)$ is measurable for every open set $V \subseteq \mathbb{R}^m$.

Proposition: Let Ω be a measurable subset of $\mathbb{R}^m$, and let $f : \Omega \to \mathbb{R}^m$ be continuous. Then f is measurable.

Proof: Suppose V is an open subset of $\mathbb{R}^m$. Since f is continuous, we know that $f^{-1}(V)$ is open relative to Ω, which is the same as $f^{-1}(V) = W \cap \Omega$ for some open W. Since W is open, it's measurable, so the union of W and Ω, which is $f^{-1}(V)$, is also measurable. ■

Proposition: Let Ω be a measurable subset of $\mathbb{R}^m$, and let $f : \Omega \to \mathbb{R}^m$ be a function. Then f is measurable if and only if $f^{-1}(B)$ is measurable for every open box B.

Proof: The if direction follows by definition. Now suppose V is open in $\mathbb{R}^m$. Then we know that we can write it as the countable union of boxes (B_j). Note that we have $f^{-1}(V) = f^{-1}\left(\bigcup_{j=1}^{\infty} B_j\right) = \bigcup_{j=1}^{\infty} f^{-1}(B_j)$. By the hypothesis we know that $f^{-1}(B_j)$ is measurable for each j, so the union is as well, which implies $f^{-1}(V)$ is measurable, as desired. ■

Proposition: Let Ω be a measurable subset of $\mathbb{R}^n$, and let $f : \Omega \to \mathbb{R}^m$ be a function. Suppose that $f = (f_1, ..., f_m)$, where $f_j : \Omega \to \mathbb{R}^m$. Then f is measurable if and only if each f_j is measurable.

Proof: First suppose f is measurable. We show that f_1 is measurable, and the rest follows similarly. Let U be an open set in $\mathbb{R}$. Note that $U \times \mathbb{R}^{m-1} \subseteq \mathbb{R}^m$ is open, so $f^{-1}(U \times \mathbb{R}^{m-1})$ is open by the measurability of f. We claim that $f_1^{-1}(U) = f^{-1}(U \times \mathbb{R}^{m-1})$, and then the proposition follows. Suppose $x \in f_1^{-1}(U)$. Then $f(x) = (u, y_2, ..., y_n)$ for $u \in U$. Since $(u, y_2, ..., y_n) \in U \times \mathbb{R}^{m-1}$, $x \in f^{-1}(U \times \mathbb{R}^{m-1})$, so $f_1^{-1}(U) \subseteq f^{-1}(U \times \mathbb{R}^{m-1})$. Now suppose $x \in f^{-1}(U \times \mathbb{R}^{m-1})$. Then $f(x) = (u, y_2, ..., y_n)$ for $u \in U$, which implies that $f_1(x) = u$. Thus $x \in f_1^{-1}(U)$, so we have the reverse inclusion as well. Thus both sets are equal, as desired.

Now suppose each component of f is measurable. We show f is measurable for every open box, and then the previous proposition implies that f is measurable. Let B be an open box. When can thus write it as $(a_1, b_1) \times \cdots \times (a_m, b_m)$. Since each component of f is open, $f_i^{-1}((a_i, b_i))$ is open in Ω. We claim that $\bigcap_{i=1}^{m} f_i^{-1}((a_i, b_i)) = f^{-1}((a_1, b_1) \times \cdots (a_n, b_n))$, and then the proposition follows by finite intersections. If $x \in f^{-1}(B)$, then clearly $f_i(x) \in (a_i, b_i)$, so $x \in f_i^{-1}((a_i, b_i))$

for each i, and thus is in the set on the left hand side. If x is in the set on the left hand side, then $f_i(x) \in (a_i, b_i)$ for each i, which implies that $f(x) \in B$, and thus f is in $f^{-1}(B)$. Since we have inclusions in both directions, the sets are equal, as desired. ■

Proposition: Let Ω be a measurable subset of $\mathbb{R}^n$, and let W be an open subset of $\mathbb{R}^m$. If $f : \Omega \to W$ is measurable, and $g : W \to \mathbb{R}^p$ is continuous, then $g \circ f : \Omega \to \mathbb{R}^p$ is measurable.

Proof: Let V be open in $\mathbb{R}^p$. We need to show that $(g \circ f)^{-1}(V) = f^{-1}(g^{-1}(V))$ is measurable. Since g is continuous, $g^{-1}(V) = U \cap W$ for some open U. Thus we have $f^{-1}(U \cap W) = f^{-1}(U) \cap f^{-1}(W)$. Since f is measurable and U, W are open, their inverse images are measurable, and thus their intersection is measurable, as desired. ■

Remark: Unfortunately, there exist measurable functions for which the composition is not measurable.

Corollary: Let Ω be a measurable subset of $\mathbb{R}^n$. If $f : \Omega \to \mathbb{R}$ and $g : \Omega \to \mathbb{R}$ are measurable functions, then so is $f + g, f - g, fg, \max(f, g)$, and $\min(f, g)$. If $g \neq 0$ on Ω, then f/g is measurable.

Proof: We show this for $f + g$, and the rest follow similarly. Let $h : \Omega \to \mathbb{R}^2$ be $h(x) = (f(x), g(x))$, and let $k : \mathbb{R}^2 \to R$ be $k(a, b) = a + b$. Since the components are measurable, h is measurable, and since k is continuous, $k \circ h = f + g$ is measurable. ■

Proposition: Let Ω be a measurable subset of $\mathbb{R}^n$, and let $f : \Omega \to \mathbb{R}$ be a function. Then f is measurable if and only if $f^{-1}((a, \infty))$ is measurable for every real number a.

Proof: Suppose f is measurable. Since (a, ∞) is open, its invsere image will be measurable. Now suppose $f^{-1}((a, \infty))$ is measurable for all a. Then $f^{-1}\left(\left(a - \frac{1}{n}, \infty\right)\right)$ is measurable as well, so $f^{-1}([a, \infty)) = f^{-1}\left(\bigcap_{n=1}^{\infty} \left(a - \frac{1}{n}, \infty\right)\right) = \bigcap_{n=1}^{\infty} f^{-1}\left(\left(a - \frac{1}{n}, \infty\right)\right)$ is measurable as well. Then for any open box (a, b), we have that $f^{-1}((a, b)) = f^{-1}((a, \infty) \setminus [b, \infty)) = f^{-1}((a, \infty)) \setminus f^{-1}([b, \infty))$ is measurable as well. Since the inverse image of every open box is measurable, f is measurable as well. ■

Proposition: Let Ω be a measurable subset of $\mathbb{R}^n$, and let $f : \Omega \to \mathbb{R}$ be measurable. Suppose $g : \Omega \to \mathbb{R}$ agrees with f except for on a set A with measure zero. Then g is measurable.

Proof: Suppose $f(x) \neq g(x)$ for $x \in A$. We need to show that $g^{-1}((a, \infty))$ is measurable for all a. Note that $f^{-1}((a, \infty))$ is measurable, since f is measurable. Let B_a be the subset of A for which $g(B) \subseteq (a, \infty)$. Note that B also has measure zero. Then $x \in A \setminus B_a \Rightarrow g(x) \leq$

a. Thus $g^{-1}((a, \infty)) = f^{-1}((a, \infty)) \backslash (A \backslash B_a)$. Since the complements of measurable sets are measurable, $g^{-1}((a, \infty))$ is measurable, as desired. ■

Definition: Let Ω be a measurable subset of $\mathbb{R}^n$. A function $f : \Omega \to \mathbb{R}^*$ is measurable if and only if $f^{-1}((a, +\infty])$ is measurable for every real a.

Proposition: Let Ω be a measurable subset of $\mathbb{R}^n$. For each positive integer n, let $f_n : \Omega \to \mathbb{R}^*$ be a measurable function. Then the functions $\sup_{n \geq 1} f_n$, $\inf_{n \geq 1} f_n$, $\limsup_{n \to \infty} f_n$, $\liminf_{n \to \infty}$ are measurable.

Proof: We first show the claim for $\sup_{n \geq 1} f_n = g$. We need to show that $g^{-1}((a, +\infty])$ is measurable for every a. We claim that

$$g^{-1}((a, +\infty]) = \bigcup_{n=1}^{\infty} f_n^{-1}((a, +\infty]).$$

Then the claim follows by countable unions. Suppose $x \in g^{-1}((a, +\infty])$. Thus $g(x) \in (a, +\infty]$. Pick ε such that $g(x) - \varepsilon > a$ (which must exist since the interval is open ignoring $+\infty$). By supremum properties, there exists some N for which $a < g(x) - \varepsilon < f_N(x)$. Thus $x \in f_N^{-1}((a, +\infty])$, and so is in the set on the right side of the equation. Now suppose x is in the set on the right. Thus $f_N(x) \in (a, +\infty]$ for some N. By the definition of the supremum, we have that $a < f_N(x) \leq g(x)$. Thus $x \in g^{-1}((a, +\infty])$. Since we have inclusion in both directions, the sets are equal, as desired.

For the inf case, we can show that if $f^{-1}((a, +\infty])$ is measurable, then $f^{-1}([-\infty, a))$ is measurable as well (this follows by showing that $f^{-1}([a, +\infty])$ is measurable, which we can do using a similar method to what we did in the previous proposition, and then taking complements). Then the same method used above works.

The lim inf and lim sup cases follow by definition, since $\limsup_{n \to \infty} f_n = \inf_{N \geq 1} \sup_{n \geq N} f_n$ and similarly for lim inf. ■

10.3. Simple Functions

Similar to how Riemann integration can be formulated using piecewise constant functions as building block, Lebesgue integration is formulated using simple functions.

Definition (simple function): Let Ω be a measurable subset of $\mathbb{R}^n$, and let $f : \Omega \to \mathbb{R}$ be a measurable function. Then f is a *simple function* if the image $f(\Omega)$ is finite.

Definition (characteristic/indicator function): Suppose Ω be a subset of $\mathbb{R}^n$, and let E be a subest of Ω. Then we define the *characteristic (indicator) function* $\chi_E : \Omega \to \mathbb{R}$ by $\chi_E(x) = 1$ if $x \in E$ and $\chi_E(x) = 0$ otherwise. Sometimes this function is also denoted at 1_E.

Proposition: Let Ω be a measurable subset of $\mathbb{R}^n$, and let $f : \Omega \to \mathbb{R}$ and $g : \Omega \to \mathbb{R}$ be simple functions. Then $f + g$ is also a simple function. The function cf is also simple for any $c \in \mathbb{R}$.

Proof: The image of of $f + g$ is $f(\Omega) + g(\Omega)$, and since each of these is finite, the sum is also finite. Similarly, the image of cf is $cf(\Omega)$, which has the same size as $f(\Omega)$, which is finite. ■

Proposition: Let Ω be a measurable set of $\mathbb{R}^n$, and let $f : \Omega \to \mathbb{R}$ be a simple function. Then there exists a finite number of real numbers $c_1, ..., c_N$, and finite number of disjoint measurable sets $E_1, ..., E_N$ in Ω such that $f = \sum_{i=1}^{N} c_i \chi_{E_i}$.

Proof: Suppose $f(\Omega) = \{c_1, ..., c_N\}$. Define $E_i := f^{-1}(\{c_i\})$. Since f is simple it's measurable, so the inverse image of any open interval is measurable. Since the image of f is finite, there exists some open interval about c_i that contains no other output of f. If this invterval is I, then $f^{-1}(I) = f^{-1}(\{c_i\})$ is measurable, and thus E_i is measurable. Clearly each of the E_i are disjoint, and the representation of f as the linear combination of indicator functions follows. ■

Proposition: Let Ω be a measurable subset of $\mathbb{R}^n$, and let $f : \Omega \to [0, +\infty]$ be a measurable function. Then there exists a sequence $f_1, f_2, ...$ of simple functions from Ω to $\mathbb{R}$ such that

$$0 \leq f_1(x) \leq f_2(x) \leq \cdots$$

for all x and such that f_n converges pointwise to f.

Proof: We claim that the $f_n(x) := \max\{j/2^n : j \in \mathbb{Z}, j/2^n \leq \min(f(x), 2^n)\}$ is such a sequence. Note that each of these functions has finite image, since the set of all possible outputs is a subset of all nonnegative multiples of 2^{-n} up until 2^n. Fix $x \in \Omega$, and first suppose that $f(x) = +\infty$. Then the term on the right side of the inequality is just 2^n, so $f_n(x) = 2^n$, which is clearly nonnegative, increasing, and converging to $+\infty$. Now suppose $f(x)$ is finite. If $\lfloor \log_2(f(x)) \rfloor = N$, then $f_n(x) = 2^n$ for all $n \leq N$. For $n > N$, the minimum then becomes $f(x)$. Then $f_n(x) = \lfloor 2^n f(x) \rfloor / 2^n$. Note that this is increasing (which follows from $2\lfloor x \rfloor \leq \lfloor 2x \rfloor$) bounded above by $f(x)$, and bounded below by $(2^n f(x) - 1)/2^n = f(x) - 1/2^n$. Thus $f_n(x) \to f(x)$.

Now we just need to show that f_n is measurable, and then we can conclude that each of them is simple. Note that for any open V, $f_n^{-1}(V)$ is the just union of $f_n^{-1}(\{j/2^n\})$ for $0 \leq j \leq 4^n$. Thus we just need to show that $f_n^{-1}(\{j/2^n\})$ is measurable for some fixed j. Consider the intervals $(j/2^n - 1/k, (j+1)/2^n)$ with $k \geq 1$. Then

$$f^{-1}\left(\bigcap_{k=1}^{\infty}\left(\frac{j}{2^n}-\frac{1}{k},\frac{j+1}{2^n}\right)\right)=f^{-1}\left(\left[\frac{j}{2^n},\frac{j+1}{2^n}\right)\right)$$

is measurable, since f is measurable. The right side is the set of all $x \in \Omega$ such that $\frac{j}{2^n} \leq f(x) < \frac{j+1}{2^n}$. By definition, for each of these x we have $f_n(x) = \frac{j}{2^n}$. Similarly, for all x such that $f_n(x) = \frac{j}{2^n}$, we clearly have $x \in f^{-1}\left(\left[\frac{j}{2^n}, \frac{j+1}{2^n}\right)\right)$. Since we have inclusions both ways, $f_n^{-1}\left(\left\{\frac{j}{2^n}\right\}\right) = f^{-1}\left(\left[\frac{j}{2^n}, \frac{j+1}{2^n}\right)\right)$, and thus is measurable. ■

Definition (Lebesgue integral of simple functions): Let Ω be a measurable subset of $\mathbb{R}^n$, and let $f : \Omega \to \mathbb{R}$ be a simple nonnegative simple function. Then the *Lebesgue integral* of f on Ω is

$$\int_{\Omega} f := \sum_{\lambda \in f(\Omega)} \lambda m(f^{-1}(\{\lambda\})).$$

Remark: Note that since f is simple, the inverse of image of $\{\lambda\}$ is indeed measurable. The integral is sometimes denoted as $\int_{\Omega} f\, dm$, and sometimes denoted using a dummy variable just like a Riemann integral.

Proposition (swapping integrals and indicator functions): Let Ω be a measurable subset of $\mathbb{R}^n$ and let $E_1, ..., E_N$ be a disjoint measurable subsets of Ω. Let $c_1, ..., c_N$ be nonnegative numbers. Then

$$\int_{\Omega} \sum_{j=1}^{N} c_j \chi_{E_j} = \sum_{j=1}^{N} c_j m(E_j).$$

Proof: We can assume that none of the c_i are 0, since they can just be removed from both side. Clearly the image $f = \sum_{j=1}^{N} c_j \chi_{E_j}$ is finite (namely $\{0, c_1, ..., c_N\}$), and since the E_i are measurable, each open set has measurable inverse image. Thus f is simple Let $d_1, ..., d_K$ be the distinct values taken on by f (not including 0), and let $F_1, ..., F_K$ denote the sets on which f takes these values. Note that if $c_{i_1} = ... = c_{i_\ell} = d_j$, then $F_j = E_{i_1} \cup ... \cup E_{i_\ell}$. Thus

$$\int_{\Omega} f = \sum_{i=1}^{K} d_i m(f^{-1}(\{d_i\})) = \sum_{i=1}^{K} d_i m(F_j).$$

Since the E_i are disjoint, $m(F_j) = m\left(E_{i_1}\right) + \cdots + m\left(E_{i_\ell}\right)$. Since $c_{i_1} = \cdots = c_{i_\ell} = d_j$, the sum on the right can be written as $\sum_{j=1}^{N} c_j m(E_j)$, as desired. ■

In line with measure theory, we say a property P holds almost everywhere if the set on which is doesn't hold (with respect to some domain) has measure 0. So for example, $\chi_{\mathbb{Q}}$ is 0 almost everywhere.

Proposition: Let Ω be a measurable set, and let $f, g : \Omega \to \mathbb{R}$ be nonnegative simple functions.

a) $0 \leq \int_\Omega f \leq +\infty$. Furthermore, $\int_\Omega f = 0$ if and only if f is zero almost everywhere.
b) $\int_\Omega (f + g) = \int_\Omega f + \int_\Omega g$.
c) $\int_\Omega af = a \int_\Omega f$ for any positive c.
d) If $f(x) \leq g(x)$ for all $x \in \Omega$, then $\int_\Omega f \leq \int_\Omega g$.
e) If $f(x) = g(x)$ for almost every $x \in \Omega$, then $\int_\Omega f = \int_\Omega g$.

Proof:

a) The inequality follows easily from the definition of the integral of simple functions, since each value is nonnegative. Now suppose f is zero almost everywhere. Then for any nonzero c_i that f takes on, its inverse image has measure zero (since it's a subset of measure zero), and thus the integral is 0. Now suppose the integral is 0. Since f is simple, we can write it as $\sum_{i=1}^N c_i \chi_{E_i}$ for distinct nonnegative c_i and disjoint E_i. Then from the previous proposition, the integral of f is equal to $\sum_{i=1}^N c_i m(E_i) = 0$. Assuming $c_1 = 0$, we see that we must have $m(E_i) = 0$ for $2 \leq i \leq N$. Thus f is nonzero on $E_2 \cup \cdots \cup E_N$, which has measure zero by additivity.

b) Suppose $f = \sum_{j=1}^N c_j \chi_{E_j}$ for positive c_j and disjoint E_j. Let $E_0 := \Omega \backslash \bigcup_{j=1}^N E_j$ and $c_0 := 0$. Similarly write $g = \sum_{k=1}^M d_k \chi_{F_k}$, and define F_0 and d_0 similarly. Then

$$f = \sum_{j=0}^N c_j \chi_{E_j} \quad \text{and} \quad g = \sum_{k=0}^M d_k \chi_{F_k}.$$

Since $\Omega = E_0 \cup \cdots \cup E_N = F_0 \cup \cdots \cup F_M$, we have

$$f = \sum_{j=0}^N \sum_{k=0}^M c_j \chi_{E_j \cap F_k} \quad \text{and} \quad g = \sum_{k=0}^M \sum_{j=0}^N d_k \chi_{E_j \cap F_k}.$$

Thus

$$f + g = \sum_{j=0}^N \sum_{k=0}^M (c_j + d_k) \chi_{E_j \cap F_k}.$$

Then by previous proposition, we have

$$\begin{aligned} \int_\Omega (f+g) = \sum_{j=0}^N \sum_{k=0}^M (c_j + d_k) m(E_j \cap F_k) &= \sum_{j=0}^N \sum_{k=0}^M c_j m(E_j \cap F_k) + \sum_{j=0}^N \sum_{k=0}^M d_k m(E_j \cap F_k) \\ &= \int_\Omega f + \int_\Omega g. \end{aligned}$$

c) Writing $af = \sum_{j=1}^N ac_j \chi_{E_j}$, we obtain

$$\int_\Omega af = \sum_{j=1}^N ac_j m(E_j) = a \sum_{j=1}^N c_j m(E_j) = a \int_\Omega f.$$

d) Let $h := g - f$. Then h is simple and nonngative, and so by b), we have $\int_\Omega g = \int_\Omega f + \int_\Omega h$. By a) we know that $\int_\Omega h \geq 0$, so the desired inequality follows.

e) First suppose $f(x) \leq g(x)$ for all $x \in \Omega$. Then we can write $f + h = g$ for a nonnegative simple h. Since $g - f$ is zero almost everywhere, g is zero almost everywhere, so $\int_\Omega h = 0$ by a). Thus $\int_\Omega f = \int_\Omega g$ by b).

Now if f is not always less than g, we can define an intermediate function f' by taking points in f that lie above g and simply moving them down onto g. Since we're editing a subset of set that f and g differ on, f' only differs from f by a set of measure 0. Thus from an earlier result, we know it's measurable. Since it's also clear that the image of f' is finite (its a subset of the union of the image of f and g), f' is simple. Since $f' \leq f$, $f' \leq g$, and since all three are pairwise equal almost everywhere, from the first paragraph, we have that $\int_\Omega f = \int_\Omega f' = \int_\Omega g$, as desired.

■

10.4. Integration of Nonngeative Measurable Functions

Essentially we just take better and better approximations of a function using simple functions.

Definition (majorizes/minorizes): Let $f, g : \Omega \to \mathbb{R}$ be functions. Then f *majorize* g, or g *minorizes* f, if $f(x) \geq g(x)$ for all $x \in \Omega$.

Definition (Lebesgue integral for nonngetive functions): Let Ω be a measurable subset of $\mathbb{R}^n$, and let $f : \Omega \to [0, \infty]$ be measurable and nonnegative. Then the *Lebesgue integral* of f on Ω is

$$\int_\Omega f := \sup\left\{\int_\Omega s : s \text{ is simple, nonnegative, and minorizes } f\right\}.$$

Remark: Note that this definition is consistent if f is a nonnegative simple function, since f clearly minorizes itself, and any other simple function that minorizes f will have smaller integral by the last part of the previous proposition.

Proposition: Let Ω be a measurable set, and let $f, g : \Omega \to [0, +\infty]$ be nonnegative measurable functions.

a) $0 \leq \int_\Omega f \leq +\infty$. Furthermore, we have $\int_\Omega f = 0$ if and only if $f(x) = 0$ for almost every $x \in \Omega$.
b) For any positive c, we have $\int_\Omega cf = c \int_\Omega f$.
c) If $f(x) \leq g(x)$ for all $x \in \Omega$, then $\int_\Omega f \leq \int_\Omega g$.
d) If $f(x) = g(x)$ for almost every $x \in \Omega$, then $\int_\Omega f = \int_\Omega g$.
e) If $\Omega' \subseteq \Omega$ is measurable, then $\int_{\Omega'} f = \int_\Omega f \chi_{\Omega'} \leq \int_\Omega f$.

Proof:

a) The inequality follows by taking the supremum of $0 \leq \int_\Omega s \leq +\infty$ for any nonnegative simple s that minorizes f, which follows from the previous proposition.

   First suppose $f(x) = 0$ for almost all $x \in \Omega$, and let s be a minorizing nonnegative simple function. Then clearly $s(x) = 0$ for almost all $x \in \Omega$, and by the previous proposition, $\int_\Omega s = 0$. Taking the supremum yields $\int_\Omega f = 0$. Now suppose $\int_\Omega f = 0$. Then for any nonnegative simple s that minorizes f, we must have $\int_\Omega s = 0$.

   We know from an earlier result that there exists a sequence $0 \leq s_1(x) \leq s_2(x) \leq \cdots$ of increasing nonnegative simple functions that converge pointwise to f for all $x \in \Omega$. Let E_n denote the set of points $x \in \Omega$ for which $s_n(x) > 0$. Then from the previous paragraph, we know that $m(E_n) = 0$ for all n. Note also that $E_n \subseteq E_{n+1}$. Let $E = \bigcup_{n=1}^\infty E_n$. Since E is the countable union of sets of measure zero, it also has measure zero. Suppose $x \in E$. Then it must lie in E_N for some N, and by the increasing subset condition, it will lie in E_n for all $n \geq N$. Thus $s_n(x) > 0$ for all $n \geq N$, and thus $f(x) = \lim_{n \to \infty} s_n(x) > 0$. Now suppose $x \notin E$. Then it doesn't lie in any E_n. In particular, $s_n(x) = 0$ for all n, and thus $f(x) = 0$. Thus f is nonzero if and only if $x \in E$. Since E is the countable union of sets of measure zero, it also has measure zero, so f is zero almost everywhere, as desired.

b) Let s be a nonnegative simple function that minorize f. By the previous proposition, we have $\int_\Omega cs = c \int_\Omega s$. Clearly cs minorizes cf, so taking supremums on both sides yields the desired equality.

c) Suppose s is a nonnegative simple function that minorizes f. Note that $g(x) - f(x) \geq 0$ for all $x \in \Omega$, so we know there exists a nonnegative simple function u that minorizes $g - f$. Then it's clear that $u + s$ is a nonnegative simple function that minorizes g. In particular, $\int_\Omega g \geq \int_\Omega (u + s) = \int_\Omega u + \int_\Omega s \geq \int_\Omega s$. Taking the supremum on the right side over all nonnegative simple s that minorize f yields $\int_\Omega g \geq \int_\Omega f$.

d) Suppose f and g are unequal on E, which we know has $m(E) = 0$. Suppose s is a nonngative simple function that minorizes f. Then define $t : \Omega \to \mathbb{R}$ to be equal to 0 if $x \in E$ and equal to $s(x)$ otherwise. Clearly t has finite image, since it's just the image of s together with 0. Since t differs from s on a set of measure zero, it is also measurable, and thus simple. Then from part e) of the last proposition, we know that $\int_\Omega s = \int_\Omega t$. Since it's clear that t minorizes g, we see that $\int_\Omega s \leq \int_\Omega g$. Taking the supremum yields $\int_\Omega f \leq \int_\Omega g$. Doing the same thing with f and g swapped yields the reverse inequality, and thus the integrals are equal.

e) Note the inequality follows from c), so we just need to prove the equality. We first prove the result for nonnegative simple s. We know that we can write $s = \sum_{j=1}^{N} c_j \chi_{E_j}$ for disjoint $E_j \in \Omega$. Then $s\chi_{\Omega'} = \sum_{j=1}^{N} c_j \chi_{E_j \cap \Omega'}$. Since $s = s\chi_{\Omega'}$ for all $x \in \Omega'$, we see that $s|_{\Omega'}$ is also equal to the sum. Since E_j is disjoint, $E_j \cap \Omega$ is also disjoint. Then

$$\int_{\Omega'} s = \sum_{j=1}^{N} c_j m(E_j \cap \Omega') = \int_{\Omega} s\chi_{\Omega'}.$$

Now suppose s minorizes f. Then clearly $s\chi_{\Omega'}$ minorizes $f\chi_{\Omega'}$, so we have $\int_{\Omega'} s = \int_{\Omega} s\chi_{\Omega'} \leq \int_{\Omega} f\chi_{\Omega'}$. Then the taking the supremum of the left yields $\int_{\Omega'} f \leq \int_{\Omega} f\chi_{\Omega'}$. Now suppose s minorizes $f\chi_{\Omega'}$. Then we can write it as $s\chi_{\Omega'}$ without changing anything, and clearly s will minorize f, so we obtain $\int_{\Omega} s = \int_{\Omega} s\chi_{\Omega'} = \int_{\Omega'} s \leq \int_{\Omega'} f$. Taking the supremum on the left yields $\int_{\Omega} f \leq \int_{\Omega'} f$. We have the inequality in both direction, so we must have equality, as desired.

■

Unlike the Riemann integral, the Lebesgue integral interacts well with limits.

Theorem (montone convergence theorem): Let Ω be a measurable subset of $\mathbb{R}^n$, and let $f_n : \Omega \to [0, +\infty]$ be a sequence of nonnegative measurable functions that are increasing. Then

$$0 \leq \int_{\Omega} f_1 \leq \int_{\Omega} f_2 \leq \cdots$$

and

$$\int_{\Omega} \sup_{n \geq 1} f_n = \sup_{n \geq 1} \int_{\Omega} f_n.$$

Remark: If we allow limits to take on values of infinity, then the sup can be replaced with lim, since monotone sequences either converge or diverge.

Proof: The inequality follows from c) of the proposition above. Since by definition we have $\sup_{m \geq 1} f_m \geq f_n$ for any n, again by c) we obtain

$$\int_{\Omega} \sup_{m \geq 1} f_m \geq \int_{\Omega} f_n.$$

Taking the supremum on the right yields

$$\int_{\Omega} \sup_{m \geq 1} f_m \geq \sup_{n \geq 1} \int_{\Omega} f_n.$$

Now we just need to show the other direction. To do this, we show that

$$\int_{\Omega} s \leq \sup_{n \geq 1} \int_{\Omega} f_n$$

for all s which minorize $\sup_{n\geq 1} f_n$, as the result then follows from taking the supremum of all such s. Fix such an s. We show

$$(1-\varepsilon)\int_\Omega s \leq \sup_{n\geq 1}\int_\Omega f_n$$

for all $0 < \varepsilon < 1$, and then letting $\varepsilon \to 0$ yields the desired result.

Fix ε. Since $s \leq \sup_{n\geq 1} f_n$, for each $x \in \Omega$, there exists N that depends on x for which $f_N(x) \geq (1-\varepsilon)s(x)$. Since f_n is increasing, the inequality $f_n(x) \geq (1-\varepsilon)s(x)$ holds for all $n \geq N$. Now define

$$E_n := \{x \in \Omega : f_n(x) \geq (1-\varepsilon)s(x)\}.$$

Then we have $E_1 \subseteq E_2 \subseteq \cdots$ and $\bigcup_{n=1}^{\infty} E_n = \Omega$.

Since s is simple, we can write it as $\sum_{j=1}^{N} c_j \chi_{F_j}$ for distinct nonnegative c_j and disjoint measurable F_j combine to make Ω. Then we can easily show that $E_n = \bigcup_{j=1}^{N} f_n^{-1}\left(\left[(1-\varepsilon)c_j, +\infty\right]\right) \cap F_j$. Since f_n is measurable, the inverse image of an interval is measurable, so E_n is measurable.

Now from the parts of the previous proposition, we have

$$(1-\varepsilon)\int_{E_n} s = \int_{E_n} (1-\varepsilon)s \leq \int_{E_n} f_n \leq \int_\Omega f_n.$$

If we can show that $\sup_{n\geq 1}\int_{E_n} s = \int_\Omega s$, then taking the supremum of both sides yields the desired result. Using the representation of s as indicator functions earlier, we have

$$\int_{E_n} s = \sum_{j=1}^{N} c_j m(F_j \cap E_n) \quad \text{and} \quad \int_\Omega s = \sum_{j=1}^{N} c_j m(F_j).$$

Thus, if we show $\sup_{n\geq 1} m(F_j \cap E_n) = m(F_j)$, we're done. Clearly we have $m(F_j \cap E_n) \leq m(F_j)$, so taking the supremum of the left yields one direction of the inequality. Now fix $\varepsilon > 0$. Since $\bigcup_{n=1}^{\infty} E_n = \Omega$, there exists N such that $n \geq N \Rightarrow m(\Omega \setminus E_n) < \varepsilon$ (since otherwise arbitrarily many E_n would exist for which $m(\Omega\setminus E_n) \geq \varepsilon$, and since they're nested increasing, some point must always lie outside of the E_n's). We can replace Ω with F_j and E_n with $F_j \cap E_n$ to then obtain that $\varepsilon > m(F_j \setminus (F_j \cap E_n)) = m(F_j) - m(F_j \cap E_n)$ for all $n \geq N$. Rearranging yields $m(F_j) - \varepsilon < m(F_j \cap E_n)$. Since ε was arbitrary, we have $m(F_j) \leq m(F_j \cap E_n)$. Taking the supremum of the right yields the other direction of the inequality, so we have equality, as desired. ■

Proposition: Let Ω be a measurable subset of $\mathbb{R}^n$, and let $f, g : \Omega \to [0, +\infty]$ be measurable. Then $\int_\Omega (f+g) = \int_\Omega f + \int_\Omega g$.

Proof: We know there exist increasing sequences of nonnegative simple functions s_n, t_n such that $\sup_n s_n = f, \sup_n t_n = g$. Clearly $\sup_n (s_n + t_n) = f + g$, so by the monotone convergence theorem, we have

$$\int_\Omega (f+g) = \sup_n \int_\Omega (s_n + t_n) = \sup_n \left(\int_\Omega s_n + \int_\Omega t_n\right) = \sup_n \int_\Omega s_n + \sup_n \int_\Omega t_n = \int_\Omega f + \int_\Omega .$$

We were able to split the sup because the sequence of functions is increasing. ■

Corollary: If Ω is a measurable subset of $\mathbb{R}^n$, and $g_1, g_2, \ldots$ are a sequence of nonnegative measurable functions from Ω to $[0, +\infty]$, then

$$\int_\Omega \sum_{n=1}^{\infty} g_n = \sum_{n=1}^{\infty} \int_\Omega g_n.$$

Proof: Apply the monotone convergence theorem to the partial sums, for which we have $\int_\Omega \sum_{n=1}^N g_n = \sum_{n=1}^N \int_\Omega g_n$ by the previous proposition. ■

Lemma (Fatou's lemma): Let Ω be a measurable subset of $\mathbb{R}^n$, and let $f_1, f_2, \ldots$ be a sequence of nonnegative measurable functions from Ω to $[0, +\infty]$. Then

$$\int_\Omega \liminf_{n\to\infty} f_n \leq \liminf_{n\to\infty} \int_\Omega f_n.$$

Proof: Since $\liminf_{n\to\infty} f_n = \sup_n (\inf_{m\geq n} f_m)$, by the monotone convergence theorem, we have

$$\int_\Omega \liminf_{n\to\infty} f_n = \sup_n \int_\Omega \inf_{m\geq n} f_m.$$

Since the integral on the right is less than f_j for all $j \geq n$, we have

$$\int_\Omega \inf_{m\geq n} f_m \leq \int_\Omega f_j.$$

Taking the inf yields

$$\int_\Omega \inf_{m\geq n} f_m \leq \inf_{j\geq n} \int_\Omega f_j.$$

Taking the supremum in n of both sides and using the first equality yields

$$\int_\Omega \liminf_{n\to\infty} f_n = \sup_n \int_\Omega \inf_{m\geq n} f_m \leq \sup_n \left(\inf_{j\geq n} \int_\Omega f_j \right) = \liminf_{n\to\infty} \int_\Omega f_n.$$

■

Proposition: Let Ω be a measurable subset of $\mathbb{R}^n$, and let $f : \Omega \to [0, +\infty]$ be a nonnegative measurable function such that $\int_\Omega f$ is finite. Then f is finite almost everywhere.

Proof: Suppose otherwise, so there exists $E \subsetneq \Omega$ with positive measure on which f is equal to $+\infty$. Define g to be be equal to f is f is infinite and 0 otherwise. Then g is simple and minorizes f, so $\int_\Omega f \geq \int_\Omega g = +\infty \cdot m(E) = +\infty$, since $m(E) > 0$, which is a contradiction. ■

Lemma (Borel-Cantelli lemma): Let $\Omega_1, \Omega_2, \ldots$ be measurable subset of $\mathbb{R}^n$ such that $\sum_{n=1}^{\infty} m(\Omega_n)$ is finite. Then the set

$$\{x \in \mathbb{R}^n : x \in \Omega_n \text{ for infinitely many } n\}$$

is a set of measure zero.

Proof: Consider $f = \sum_{n=1}^{\infty} \chi_{\Omega n}$. We have

$$\int_{\Omega} f = \sum_{n=1}^{\infty} m(\Omega_n) < +\infty,$$

so by the previous proposition, f is finite almost everywhere. Since $f(x)$ is only infinite if x is contained in infintely many Ω_n, we have our desired result. ■

10.5. Integration of Absolutely Integrable Functions

Definition (absolutely integral function): Let Ω be a measurable subset of $\mathbb{R}^n$. A measurable function $f : \Omega \to \mathbb{R}^*$ is said to be *absolutely integrable* if the integral $\int_{\Omega} |f|$ si finite.

Definition (positive/negative part): If $f : \Omega \to \mathbb{R}^*$ is a function, then the *positive part* $f^+ : \Omega \to [0, +\infty]$ and the *negative part* $f^- : \Omega \to [0, +\infty]$ are given by

$$f^+ = \max(f, 0) \quad \text{and} \quad f^- = -\min(f, 0).$$

If f is measurable, then both the positive and negative parts are also measurable, since max and min are continuous. Thus the next definition makes sense.

Definition (Lebesgue integral): Let $f : \Omega \to \mathbb{R}^*$ be an absolutely integrable function. Then the *Lebesgue integral* of f is given by

$$\int_{\Omega} f := \int_{\Omega} f^+ - \int_{\Omega} f^-.$$

Proposition: Let Ω be a measurable set, and let $f, g : \Omega \to \mathbb{R}$ be absolutely integrable functions.

a) For any real c, $\int_\Omega cf = c \int_\Omega f$.
b) $\left|\int_\Omega f\right| \leq \int_\Omega |f|$.
c) The function $f + g$ is absolutely integrable, and $\int_\Omega (f + g) = \int_\Omega f + \int_\Omega g$.
d) If $f(x) \leq g(x)$ for all $x \in \Omega$, then we have $\int_\Omega f \leq \int_\Omega g$.
e) If $f(x) = g(x)$ for almost every $x \in \Omega$, then $\int_\Omega f = \int_\Omega g$.

Proof:

a) If $c = 0$ then the equality is obvious. If $c > 0$, then

$$\int_\Omega cf = \int_\Omega cf^+ - \int_\Omega cf^- = c\left(\int_\Omega f^+ - \int_\Omega f^-\right) = c\int_\Omega f.$$

If $c < 0$, then

$$\int_\Omega cf = \int_\Omega -cf^- - \int_\Omega -cf^+ = -c\left(\int_\Omega f^- - \int_\Omega f^+\right) = c\int_\Omega f.$$

b)
$$\left|\int_\Omega f\right| = \left|\int_\Omega f^+ - \int_\Omega f^-\right| \leq \left|\int_\Omega f^+\right| + \left|\int_\Omega f^-\right| = \int_\Omega f^+ + \int_\Omega f^- = \int_\Omega |f|.$$

c) Note that $|f + g| \leq |f| + |g|$, so $\int_\Omega |f + g| \leq \int_\Omega |f| + \int_\Omega |g| < +\infty$, so $f + g$ is absolutely integrable. Then, doing casework on the sign and magnitude of each function, we can show that $(f + g)^+ + f^- + g^- = (f + g)^- + f^+ + g^+$. Since all of these functions are positive, we can integrate and split the sums to get, then rearrange to get

$$\int_\Omega (f + g)^+ - \int_\Omega (f + g)^- = \int_\Omega f^+ - \int_\Omega f^- + \int_\Omega g^+ - \int_\Omega g^- \Rightarrow \int_\Omega (f + g) = \int_\Omega f + \int_\Omega g.$$

d) The inequality implies that $f^+ \leq g^+$ and $f^- \geq g^-$. Integrating both inequalities and subtracting the second from the first yields the desired inequality.

e) Clearly if $f = g$ almost everywhere, then $f^+ = g^+$ and $f^- = g^-$ almost everywhere.

■

Theorem (dominated convergence theorem): Let Ω be a measurable subset of $\mathbb{R}^n$, and let $f_1, f_2, \ldots$ be sequence of measurable functions from Ω to $\mathbb{R}^*$ that converge pointwise. Suppose also that there exists an absolutely integrable $F : \Omega \to [0, +\infty]$ such that $|f_n(x)| \leq F(x)$ for all $x \in \Omega, n \in \mathbb{N}$. Then

$$\int_\Omega \lim_{n\to\infty} f_n = \lim_{n\to\infty} \int_\Omega f_n.$$

Proof: Note that F must be finite almost everywhere to be absolutely integrable, so we can delete the set for which F is infinite (since it has measure zero, removing it doesn't effect the integrals). Thus, (f_n) converges pointwise to f with $|f(x)| \leq F(x)$.

Since $F + f_n$ is nonnegative for all n by the condition in the statement, using Fatou's lemma yields

$$\int_\Omega F + f = \int_\Omega \liminf_{n\to\infty}(F + f_n) \leq \liminf_{n\to\infty} \int_\Omega F + f_n,$$

where the equality came from the fact that $f_n \to f$. We can split the integral on the right and then split the liminf, yielding

$$\liminf_{n\to\infty} \int_\Omega F + f_n \leq \liminf_{n\to\infty} \int_\Omega F + \liminf_{n\to\infty} \int_\Omega f_n = \int_\Omega F + \liminf_{n\to\infty} \int_\Omega f_n.$$

Thus $\int_\Omega f \leq \liminf_{n\to\infty} \int_\Omega f_n$.

The condition also implies that $F - f_n$ is nonnegative for all n. By Fatou's lemma again, we obtain

$$\int_\Omega F - f \leq \liminf_{n\to\infty} \int_\Omega F - f_n \leq \int_\Omega F + \liminf_{n\to\infty} \int_\Omega -f_n = \int_\Omega F - \limsup_{n\to\infty} \int_\Omega f_n,$$

where the liminf changes into a limsup because we moved the sign. Thus $\int_\Omega f \geq \limsup_{n\to\infty} \int_\Omega f_n$. This and the inequality above with liminf imply the desired equality. ■

Definition (upper/lower Lebesgue integral): Let Ω be a measurable subset of $\mathbb{R}^n$, and let $f : \Omega \to \mathbb{R}$ be a function (not necessarily measurable). Then the *upper Lebesgue integral* $U(f, \Omega)$ is

$$U(f, \Omega) := \inf\left\{\int_\Omega g : g \text{ is an absolutely integrable function from } \Omega \text{ to } \mathbb{R} \text{ that majorizes } f\right\},$$

and the *lower Lebesgue integral* $L(f, \Omega)$ is

$$L(f, \Omega) := \sup\left\{\int_\Omega g : g \text{ is an absolutely integrable function from } \Omega \text{ to } \mathbb{R} \text{ that minorizes } f\right\}.$$

Clearly $L(f, \Omega) \leq U(f, \Omega)$, and when f is absolutely integrable, these two must be equal (since we can just use simple functions). The converse is also true, similar to Riemann integrals.

Proposition: Let Ω be a measurable subset of $\mathbb{R}^n$, and let $f : \Omega \to \mathbb{R}$ be a function (not necessarily measurable). let A be a real number, and suppose $L(f, \Omega) = U(f, \Omega)$. Then f is absolutely integrable, and $\int_\Omega f = A$.

Proof: For every n, there exists absolutely integrable function f_n^+, f_n^- that majorizes, minorizes f for which $A - \frac{1}{n} \leq \int_\Omega f_n^- \leq \int_\Omega f_n^+ \leq \Omega \leq A + \frac{1}{n}$. Let $F^+ = \inf_n f_n^-, F^- = \sup_n f_n^+$. Since each f_n is measurable, these two functions are also measurable. Note they're also absolutely integrable, since they're between two absolutely integrable functions (f_1^+ and f_1^-). Note that

F^+ majorizes F^-, since $f_n^+ \geq f_n^-$. From the inequalities on the integrals of the sequence of functions, we can take limits to obtain $A \leq \int_\Omega F^- \leq \int_\Omega F^+ \leq A$. Thus $\int_\Omega F^- = \int_\Omega F^+$. Since $F^+ \geq F^-$, we conclude that $F - = F^+$ almost everywhere (this follows by subtracting the two integrals and then noting the integran must be zero almost everywhere, since the integrand is nonnegative). Since f is between $F -$ and F^+, its equal to them almost everywhere, and thus measurable. Then $\int_\Omega F^+ = \int_\Omega f = A$, as desired. ■

10.5.1. Riemann Integrable Functions Are Lebesgue Integrable

Here we show that if a function is Riemann integrable, then it's also Lebesgue integrable, and show that they have the same value. Since $\chi_{[0,1]\setminus\mathbb{Q}}$ is not Riemann integrable but is Lebesgue integral, the latter is strictly stronger than the former.

Let $I \subseteq \mathbb{R}$ be a bounded interval, and let A denote the value of the Riemann integral of f over I. Then there exists a partition $P_\varepsilon = \{x_0, ..., x_N\}$ of I for which

$$A - \varepsilon \leq \sum_{i=0}^{N-1} m_i(x_{i+1} - x_i) \leq A \leq \sum_{i=0}^{N-1} M_i(x_{i+1} - x_i) \leq A + \varepsilon,$$

where m_i and M_i denote the inf, sup of f over that subinterval. Define $f_\varepsilon^+ = \sum_{i=0}^{N-1} M_i \chi_{[x_i, x_{i+1}]}$ and define f_ε^- similarly. Clearly this is simple, we know the Lebesgue integral of this over I is just

$$\sum_{i=0}^{N-1} \int_I m_i \chi_{[x_i, x_{i+1}]} = \sum_{i=1}^{N-1} M_i(x_{i+1} - x_i),$$

and similarly for $\int_I f_\varepsilon^-$. Thus we can replace these in the inequality to obtain

$$A - \varepsilon \leq \int_I f_\varepsilon^- \leq A \leq \int_I f_\varepsilon^+ \leq A + \varepsilon.$$

Since $f_\varepsilon^- \leq f \leq f_\varepsilon^+$, we have $A - \varepsilon \leq L(f, I) \leq A \leq U(f, I) \leq A - \varepsilon$, where these are the lower and upper Lebesgue integral. Since this holds for all ε, we have $L(f, I) = U(f, I) = A$, and thus the previous proposition, f is absolutely integrable with $\int_I f = A$.

10.6. Fubini's Theorem (INCOMPLETE)

Theorem (Fubini's theorem): Let $f : \mathbb{R}^2 \to \mathbb{R}$ be an absolutely integrable function. Then there exists absolutely integrable functions $F, G : \mathbb{R} \to \mathbb{R}$ such that for almost every x, $f(x, y)$ is absolutely integrable in y with

$$F(x) = \int_{\mathbb{R}} f(x, y)\, dy,$$

and for almost every y, $f(x, y)$ is absolutely integrable in x with

$$G(y) = \int_{\mathbb{R}} f(x, y)\, dx.$$

We also have

$$\int_{\mathbb{R}} F(x)\, dx = \int_{\mathbb{R}^2} f = \int_{\mathbb{R}} G(y)\, dy.$$

Remark: Once we have Fubini's theorem for $\mathbb{R}^2$, we can get it for any measurable $\Omega \subseteq \mathbb{R}^2$ by using $f\chi_\Omega$.

Proof: We only prove the case with $F(x)$, as $G(y)$ follows similarly. We start with a few simplifications.

First suppose Fubini's theorem holds for nonnegative functions. That is, if $f : \mathbb{R}^2 \to [0, +\infty)$ has finite integral, then $f(x, y)$ is absolutely integrable in y for almost every x, and $F(x)$ is also absolutely integrable with integral equal to $\int_{\mathbb{R}^2} f$. Then for general f that's absolutely integrable, we write $f = f^+ - f^-$. Since f is absolutely integrable, both the positive and negative part are as well ($\int f^+ + \int f^- = \int f^+ + f^- = \int |f| < +\infty$). Thus we can apply Fubini's theorem to each part. In particular, $f^+(x, y)$ and $f^-(x, y)$ are both absolutely integrable in y for almost every x (unions of sets of measure zero are zero, so we can ignore the union of the sets for which one or the other is not absolutely integrable). Thus

$$F^+(x) := \int_{\mathbb{R}} f^+(x, y)\, dy \quad \text{and} \quad F^-(x) := \int_{\mathbb{R}} f^-(x, y)\, dy$$

are both absolutely integrable, and

$$\int_{\mathbb{R}} F^+(x)\, dx = \int_{\mathbb{R}^2} f^+ \quad \text{and} \quad \int_{\mathbb{R}} F^-(x)\, dx = \int_{\mathbb{R}^2} f^-.$$

Since f^+ and f^- are both absolutely integrable for almost all x, $f = f^+ - f^-$ will also be absolutely integrable for almost all x (in particular, if f^+ is not absolutely integrable on E_1, and f^- is not absolutely integrable on E_2, then f is not absolutely integrable on $E_1 \cup E_2$, which has measure zero). Then

$$F(x) := F^+(x) - F^-(x) = \int_{\mathbb{R}} f(x, y)\, dy$$

will also be absolutely integrable, since both F^+ and F^- are absolutely integrable. Integrating both sides then yields

$$\int_{\mathbb{R}} F(x)\,dx = \int_{\mathbb{R}} F^+(x)\,dx - \int_{\mathbb{R}} F^-(x)\,dx = \int_{\mathbb{R}^2} f^+ - \int_{\mathbb{R}^2} f^- = \int_{\mathbb{R}^2} f,$$

so Fubini's theorem holds for general absolutely integrable f.

Now suppose Fubini's theorem holds for $f\chi_{[-N,N]^2}$ for nonnegative absolutely integrable f and integers $N \geq 1$. Then for general absolutely integrable nonnegative f, we have $f = \sup_{N\geq 1} f\chi_{[-N,N]^2}$. Since f is absolutely integrable, $f\chi_{[-N,N]^2} \leq f$ is clearly absolutely integrable, so Fubini's theorem applies. Integrating both sides of the equation and applying the monotone convergence theorem (since $f\chi_{[-N,N]^2}$ is clearly increasing) yields

$$\sup_{N\geq 1} \int_{\mathbb{R}} \left(f\chi_{[-N,N]^2}\right)(x,y)\,dy = \int_{\mathbb{R}} f(x,y)\,dy.$$

From Fubini's theorem, we know that each $f\chi_{[-N,N]^2}$ has measure zero set E_N on which it's not absolutely integrable. Then $E = \bigcup_{i=1}^{\infty} E_i$ has measure zero. For $x \notin E$, $\left(f\chi_{[-N,N]^2}\right)(x,y)$ is absolutely integrable in y. Let E' denote the set of x for which that sequence is unbounded. (FIND WAY TO SHOW THAT f IS ABSOLUTELY INTEGRABLE BECAUSE RIGHT NOW I HAVE NO CLUE) Thus $f(x,y)$ is absolutely integrable in y for almost all x. Let $F_N(x)$ equal the integral on the left of the equation, and let $F(x)$ equal the integral on the right. Then integrating and applying the monotone convergence theorem again yields

$$\sup_{N\geq 1} \int_{\mathbb{R}} F_N(x)\,dx = \int_{\mathbb{R}} F(x)\,dx.$$

From Fubini's theorem, we know that $\int_{\mathbb{R}} F_N(x)\,dx = \int_{\mathbb{R}^2} f\chi_{[-N,N]^2}$, so we can bring the supremum back inside by the monotone convergence theorem and obtain

$$\int_{\mathbb{R}^2} f = \int_{\mathbb{R}^2} \sup_{N\geq 1} f\chi_{[-N,N]^2} = \int_{\mathbb{R}} F(x)\,dx,$$

so Fubini's theorem holds for general nonnegative absolutely integrable f ($F(x)$ is absolutely integrable since its integral is equal to the finite integral $\int_{\mathbb{R}^2} f$).

Let $f_N := f\chi_{[-N,N]^2}$. Now suppose Fubini's theorem holds for nonnegative simple functions that are zero outside $[-N,N]^2$. Then for general nonnegative absolutely integrable f_N, we can write is as the supremum of a sequence increasing simple functions $0 \leq s_1 \leq s_2 \leq \cdots \leq f_N$. Then we can do the same thing we did in the previous case to show that Fubini's theorem holds for f_N.

Let s_N denote a nonnegative simple absolutely integrable function that's zero outside $[-N,N]^2$. Now suppose Fubini's theorem holds for characteristic functions of sets that are contained in $[-N,N]^2$. We can write $s_N = \sum_{j=1}^{M} c_j\chi_{E_j}$ for nonnegative c_j and disjoint E_j. Clearly χ_{E_j} is absolutely integrable, since $\int_{\mathbb{R}^2} \chi_{E_j} \leq \int_{\mathbb{R}^2} \chi_{[-N,N]^2} = 4N^2$, so Fubini's theorem holds. Then integrating yields

$$\int_{\mathbb{R}} s_N(x,y)\,dy = \sum_{j=1}^{M} c_j \int_{\mathbb{R}} \chi_{E_j}(x,y)\,dy.$$

Since $\chi_{E_j}(x,y)$ is absolutely integrable for every x ($\int_{\mathbb{R}} \chi_{E_j}(x,y)\,dy \leq \int_{\mathbb{R}} \chi_{[-N,N]^2}(x,y)\,dy = 2N$), $s_n(x,y)$ is absolutely integrable for all x as well. Letting $S_N(x) := \int_{\mathbb{R}} s_N(x,y)\,dy$ and $\psi_{E_j}(x) := \int_{\mathbb{R}} \chi_{E_j}(x,y)\,dy$ and integrating both sides yields

$$\int_{\mathbb{R}} S_N(x)\,dx = \sum_{j=1}^{M} c_j \int_{\mathbb{R}} \psi_{E_j}(x)\,dx = \sum_{j=1}^{M} c_j \int_{\mathbb{R}^2} \chi_{E_j} = \sum_{j=1}^{M} c_j m(E_j) = \int_{\mathbb{R}^2} s_N.$$

Thus Fubini's theorem holds for s_N.

■

10.7. Problems

11. Measures

11.1. Measurable Spaces and Functions

Definition (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 three conditions are satisfied:

- $\emptyset \in \mathcal{S}$,
- if $E \in \mathcal{S}$, then $X \backslash E \in \mathcal{S}$,
- if $E_1, E_2, \ldots$ is a sequence of elements in $\mathcal{S}$, then $\bigcup_{k=1}^{\infty} E_k \in \mathcal{S}$.

Example: Suppose X is a set. Then $\{\emptyset, X\}$ is a σ-algebra on X, and so is $\mathcal{P}(X)$.

Proposition (properties of σ-algebras): Suppose $\mathcal{S}$ is a σ-algebra on a set 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 $D \backslash E \in \mathcal{S}$,
c) if $E_1, E_2, \ldots$ is a sequence of elements of $\mathcal{S}$, then $\bigcap_{k=1}^{\infty} E_k \in \mathcal{S}$.

Proof: Follows from definition of σ-algebra and DeMorgan's laws. ■

Definition (mesaurable space/set): 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 *measurable set* if $\mathcal{S}$ is clear from context.

Proposition: 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.

Proof: Let $\mathcal{B}$ be the intersection of all such σ-algebras. Note that there must at least one since $\mathcal{P}(X)$ contains $\mathcal{A}$. Note also that each must contain $\emptyset$ by definition, so $\emptyset \in \mathcal{B}$.

Now suppose $E \in \mathcal{B}$. Thus E is contained in all σ-algebras that contain $\mathcal{A}$. Thus all σ-algebras that contain $\mathcal{A}$ contain $X \backslash E$, and thus $X \backslash E \in \mathcal{B}$. Similar logic works to show that $\mathcal{B}$ is closed under countable unions, so $\mathcal{B}$ is a σ-algebra on X, as desired. ■

Remark: The result basically says that there exists a smallest σ-algebra that contains $\mathcal{A}$.

Definition (Borel set): The smallest σ-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 σ-algebra is called a *Borel set*.

Definition (measurable function): Suppose $(X, \mathcal{S})$ is a measurable space. A function $f : X \to \mathbb{R}$ is called *$\mathcal{S}$-measurable* if $f^{-1}(B) \in \mathcal{S}$ for every Borel set $B \subseteq \mathbb{R}$.

Definition (characteristic function): Suppose E is a subset of X. The *characteristic function* of E is the function $\chi_E : X \to \mathbb{R}$ defined by

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

Proposition: Suppose $(X, \mathcal{S})$ is a measurable space and $f : X \to \mathbb{R}$ is a function such that $f^{-1}((a, \infty)) \in \mathcal{S}$ for all $a \in \mathbb{R}$. Then f is an $\mathcal{S}$-measurable function.

Proof: Let $\mathcal{T} = \{A \subseteq \mathbb{R} : f^{-1}(A) \in \mathcal{S}\}$. We show that every Borel subset of $\mathbb{R}$ is in $\mathcal{T}$. First we show that $\mathcal{T}$ is a σ-algebra on $\mathbb{R}$. ■

12. Operator Swapping

A chapter that acts as a compendium of rules for when you can swap operators. Results with swapping sups, infs, limsups, and liminfs with other operators will be included in the section with the other operator and limit (or limit limit if both operators are the above operations).

12.1. Limit Limit

Theorem: Let (X, d_X) and (Y, d_Y) be metric spaces with Y complete, and let E be a subset of E. Let (f_n) be a sequence of functions from E to Y that converges uniformly in E to $f : E \to Y$. Suppose $x_0 \in X$ is an adherent point of E, and suppose $\lim_{x \to x_0} f_n(x) = L_n$ exists for all n. Then $\lim_{x \to x_0} f(x)$ exists and is equal to the limit of the sequence (L_n). In other words,

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

Proof: First we show that (L_n) is Cauchy, and since Y is complete, this implies that $L_n \to L$ for some $L \in Y$. Pick $\varepsilon > 0$. We have

$$d_Y(L_n, L_m) \leq d_Y(L_n, f_n(x)) + d_Y(f_n(x), f_m(x)) + d_Y(f_m(x), L_m)$$

for all $x \in E$ and $n, m \in \mathbb{N}$s. Since $f_n \to f$ uniformly, there exists N such that $n, m \geq N$ implies that $d_Y(f_n(x), f_m(x)) < \frac{\varepsilon}{3}$ for all $x \in E$. Since we know each of the limits exist, for a fixed pair n, m there exist $\delta_{n,m} > 0$ such that $0 < d_X(x, x_0) < \delta_{n,m} \Rightarrow d_Y(f_n(x), L_N), d_Y(f_m(x), L_m) < \frac{\varepsilon}{3}$. Note that N does not depend on δ, however, so δ's dependence on n, m is not an issue. Thus $d_Y(L_n, L_m) < \frac{\varepsilon}{3} + \frac{\varepsilon}{3} + \frac{\varepsilon}{3} = \varepsilon$, so (L_n) is Cauchy, as desired.

Now we show that $\lim_{x \to x_0} f(x) = L$. Pick $\varepsilon > 0$. We have

$$d_Y(f(x), L) \leq d_Y(f(x), f_n(x)) + d_Y(f_n(x), L_n) + d_Y(L_n, L)$$

for all $x \in E$ and $n \in \mathbb{N}$. We know from uniform convergence and the previous paragraph that there exists N such that $n \geq N \Rightarrow d_Y(f(x), f_n(x)), d_Y(L_n, L) < \frac{\varepsilon}{3}$ for all $x \in E$. Fix $n = N$. Thus we have

$$d_Y(f(x), L) < \frac{\varepsilon}{3} + d_Y(f_N(x), L_N) + \frac{\varepsilon}{3}.$$

Then from the limits, we know there exists $\delta > 0$ such that $0 < d_X(x, x_0) < \delta \Rightarrow d_Y(f_N(x), L_N) < \frac{\varepsilon}{3}$. Since δ doesn't depend on anything, the limit exists and is equal to L, as desired. ■

Proposition: If $f : \mathbb{R}^2 \to \mathbb{R}$ is continuous at (a, b), then

$$\begin{aligned} f(a,b) &= \lim_{x\to a} \limsup_{y\to b} f(x,y) = \lim_{y\to b} \limsup_{x\to a} f(x,y) \\ &= \lim_{x\to a} \liminf_{y\to b} f(x,y) = \lim_{y\to b} \liminf_{x\to a} f(x,y), \end{aligned}$$

where $\limsup_{x\to x_0} f(x) = \inf_{r>0} \sup_{|x-x_0|<r} f(x) = \lim_{r\to 0} \sup_{|x-x_0|<r} f(x)$ and similarly for $\liminf$.

Remark: The last equivalence for $\limsup$ comes from noting that $\sup_{|x-x_0|<r} f(x)$ decreases as r decreases

Proof: We simply do the first equality, as the rest follow similarly. Pick $\varepsilon > 0$. From continuity, we have that for some δ, $\|(x,y)-(a,b)\| < \delta \Rightarrow |f(x,y)-f(a,b)| < \varepsilon$. Then for $x \in \left(a - \frac{\delta}{2}, a + \frac{\delta}{2}\right), y \in \left(b - \frac{\delta}{2}, b + \frac{\delta}{2}\right)$ (since then $\|(x,y)-(a,b)\| < \frac{\delta}{\sqrt{2}} < \delta$), we have $f(a,b) - \varepsilon < f(x,y) < f(a,b) + \varepsilon$. Thus, $f(a,b) - \varepsilon \leq \sup_{|y-b|<\frac{\delta}{2}} f(x,y) \leq f(a,b) + \varepsilon$, which then implies $f(a,b) - \varepsilon \leq \limsup_{y\to b} f(x,y) \leq f(a,b) + \varepsilon$.

Now note that for all $x \in \left(a - \frac{\delta}{2}, a + \frac{\delta}{2}\right)$, we have that $\left|\limsup_{y\to b} f(x,y) - f(a,b)\right| < \varepsilon$. Since this holds for arbitrary ε, we have the desired limit. ■

Corollary: If $f : \mathbb{R}^2 \to \mathbb{R}$ is continuous at (a, b) and the one sided limits both exist, then

$$\lim_{x\to a} \lim_{y\to b} f(x,y) = \lim_{y\to b} \lim_{x\to a} f(x,y) = f(a,b).$$

12.2. Derivative Derivative

Theorem (Clairaut's theorem): Let E be an open subset of $\mathbb{R}^n$, let $x_0 \in E$, and let $f : E \to \mathbb{R}^m$ be twice continuously differentiable on E. Then

$$\frac{\partial}{\partial x_j} \frac{\partial f}{\partial x_i}(x_0) = \frac{\partial}{\partial x_i} \frac{\partial f}{\partial x_j}(x_0)$$

for all $1 \leq i, j \leq n$.

Proof: We work with one component of f at a time, so we can assume $m = 1$. The theorem is obvious for $i = j$, so suppose $i \neq j$. Without loss of generality, assume $x_0 = 0$. Let $\frac{\partial}{\partial x_j} \frac{\partial f}{\partial x_i}(0) = a_1$ and $\frac{\partial}{\partial x_i} \frac{\partial f}{\partial x_j}(0) = a_2$. We need to show that $a_1 = a_2$.

Pick $\varepsilon > 0$. From the continuity of the double derivatives, there exists $\delta > 0$ such that if $\|x\| < 2\delta$, we have

$$\left|\frac{\partial}{\partial x_j}\frac{\partial f}{\partial x_i}(x) - a_1\right| < \varepsilon \quad \text{and} \quad \left|\frac{\partial}{\partial x_i}\frac{\partial f}{\partial x_j}(x) - a_2\right|.$$

Define

$$X = f\left(\delta e_i + \delta e_j\right) - f(\delta e_i) - f\left(\delta e_j\right) + f(0).$$

From the fundamental theorem of calculus in x_i, we have

$$f\left(\delta e_i + \delta e_j\right) - f(\delta e_i) = \int_0^\delta \frac{\partial f}{\partial x_i}\left(x_i e_i + \delta e_j\right) dx_i \quad \text{and} \quad f\left(\delta e_j\right) - f(0) = \int_0^\delta \frac{\partial f}{\partial x_i}(x_i e_i)\, dx_i,$$

so

$$X = \int_0^\delta \left(\frac{\partial f}{\partial x_i}\left(x_i e_i + \delta e_j\right) - \frac{\partial f}{\partial x_i}(x_i e_i)\right) dx_i.$$

From the mean value theorem in the x_j variable, for each $x_i \in [0, \delta]$, there exists $t_{x_i} \in (0, \delta)$ such

$$\frac{\partial f}{\partial x_i}\left(x_i e_i + \delta e_j\right) - \frac{\partial f}{\partial x_i}(x_i e_i) = \delta \frac{\partial}{\partial x_j}\frac{\partial f}{\partial x_i}\left(x_i e_i + t_{x_i} e_j\right).$$

Thus by construction we have

$$\left|\frac{\partial f}{\partial x_i}\left(x_i e_i + \delta e_j\right) - \frac{\partial f}{\partial x_i}(x_i e_i) - \delta a_1\right| < \varepsilon\delta.$$

Integrating both sides yields

$$\begin{aligned}\left|X - \delta^2 a_1\right| &= \left|\int_0^\delta \left(\frac{\partial f}{\partial x_i}\left(x_i e_i + \delta e_j\right) - \frac{\partial f}{\partial x_i}(x_i e_i) - \delta a_1\right) dx_i\right| \\ &\leq \int_0^\delta \left|\frac{\partial f}{\partial x_i}\left(x_i e_i + \delta e_j\right) - \frac{\partial f}{\partial x_i}(x_i e_i) - \delta a_1\right| dx_i < \varepsilon\delta^2.\end{aligned}$$

Swapping the roles of i and j, we similarly obtain $\left|X - \delta^2 a_2\right| < \varepsilon\delta^2$. Applying the triangle inequality yields $\left|\delta^2 a_1 - \delta^2 a_2\right| < 2\varepsilon\delta^2 \Rightarrow |a_1 - a_2| < 2\varepsilon$. Since ε is arbitrary, we have $a_1 = a_2$, as desired. ■

Here's a slick proof for functions from $\mathbb{R}^2$ to $\mathbb{R}$ that uses Fubini's theorem.

Proof: Suppose $[a, b] \times [c, d]$ is a box in E. Since f is twice continuously differentiable on E, the mixed partials will be continuous on the both. Since the box is compact, f is absolutely integrable (since its bounded by the extreme value theorem), so Fubini's theorem applies. Thus

$$\begin{aligned}\int_{[a,b]\times[c,d]} \frac{\partial}{\partial x_1}\frac{\partial f}{\partial x_2} &= \int_c^d \int_a^b \frac{\partial}{\partial x_1}\frac{\partial f}{\partial x_2}(x, y)\, dx\, dy \\ &= \int_c^d \frac{\partial f}{\partial x_2}(b, y) - \frac{\partial f}{\partial x_2}(a, y)\, dy \\ &= f(b, d) - f(b, c) - f(a, c) + f(a, b).\end{aligned}$$

Doing the same for the other mixed partial yields the same value. Thus

$$\int_{[a,b]\times[c,d]} \left(\frac{\partial}{\partial x_1} \frac{\partial f}{\partial x_2} - \frac{\partial}{\partial x_2} \frac{\partial f}{\partial x_1} \right) = 0.$$

Now suppose the mixed partials weren't equal at some point. Since the integrand above is continuous, there exists a box around at point where the integrand has the same sign, and thus the integral over that box would be nonzero. However, the above equation applies to any box in E, so we have a contradiction. ■

12.3. Integral Integral

Theorem (Fubini's Theorem): Let $Q = A \times B$, where A is a rectangle in $\mathbb{R}^k$ and B is a rectangle in $\mathbb{R}^n$. Let $f : Q \to \mathbb{R}$ be a bounded function and write in the form $f(x, y)$ for $x \in A$ and $y \in B$. For each $x \in A$, consider the lower and upper integrals

$$\underline{\int_B} f(x,y) \quad \text{and} \quad \overline{\int_B} f(x,y).$$

If f is integrable over Q, then these two functions of x are integrable over A, and

$$\int_Q f = \int_A \underline{\int_B} f(x,y) = \int_A \overline{\int_B} f(x,y).$$

Proof: Define

$$\underline{I}(x) = \underline{\int_B} f(x,y) \quad \text{and} \quad \overline{I}(x) = \overline{\int_B} f(x,y)$$

for $x \in A$. Assuming $\int_Q f$ exists, we show that $\overline{I}$ and $\underline{I}$ are integrable over A.

Let P be partition of Q. Then P consists of a partition P_A of A, and a partition P_B of B. If R_A is a subrectangle of A induced by P_A, and similarly for R_B, then $R_A \times R_B$ is a subrectangle of Q induced by P.

We show that $L(f, P) \leq L(\underline{I}, P_A)$. Pick some rectangle R in Q induced by P. Then from above, we can write $R = R_A \times R_B$. Fix some $x_0 \in R_A$. Then

$$m_R(f) \leq f(x_0, y)$$

for all $y \in R_B$, so taking the infinum over all $y \in R_B$ yields

$$m_R(f) \leq m_{R_B}(f(x_0, y)).$$

Now multiply both sides by $v(R_B)$ and sum over all R_B in B. We then have

$$\sum_{R_B} m_R(f) v(R_B) \leq \sum_{R_B} m_{R_B}(f(x_0, y)) v(R_B) = L(f(x_0, y), R_B) \leq \underline{I}(x_0).$$

This holds for all $x_0 \in R_A$, so $\sum_{R_B} m_R(f) v(R_B) \leq m_{R_A}(\underline{I})$. Then multiplying by $v(R_A)$ and summing over all of them yields

$$L(f,P)=\sum_R m_R(f)v(R)=\sum_{R_A}\sum_{R_B}m_R(f)v(R_A)v(R_B)\leq\sum_{R_A}m_{R_A}(\underline{I})v(R_A)=L(\underline{I},P_A).$$

The same method shows that $U(f,P)\geq U\left(\overline{I},P_A\right)$.

Combinding everything we have

$$L(f,P)\leq L(\underline{I},P_A)\leq L\left(\overline{I},P_A\right),U(\underline{I},P_A)\leq U\left(\overline{I},P_A\right)\leq U(f,P).$$

Since f is integrable, there exists a partition P_ε for which the extreme bounds are ε apart. That implies that $L(\underline{I},P_A)$ and $U(\underline{I},P_A)$ are within ε of each other, and similarly for $\overline{I}$. Thus they are integrable over A. Note that since P is arbitrary, we must have that the two extreme ends must be equal, so combining everything, we obtain

$$\int_A\underline{I}=\int_A\overline{I}=\int_Q f.$$

■

Corollary: Let $Q=A\times B$, where A is a rectangle in $\mathbb{R}^k$ and B is a rectangle in $\mathbb{R}^n$. Let $f:Q\to\mathbb{R}$ be a bounded function. If $\int_Q f$ exists, and if $\int_{y\in B}f(x,y)$ exists for each $x\in A$, then

$$\int_Q f=\int_{x\in A}\int_{y\in B}f(x,y).$$

Corollary: Let $Q=I_1\times\cdots\times I_n$, where I_j is a closed interval in $\mathbb{R}$. If $f:Q\to\mathbb{R}$ is continuous, then

$$\int_Q f=\int_{x_1\in I_1}\cdots\int_{x_n\in I_n}f(x_1,\ldots,x_n).$$

12.4. Sum Sum

Lemma: Supose the iterated series

$$\sum_{i=1}^\infty\sum_{j=1}^\infty|a_{ij}|$$

converges. Then

$$\sum_{j=1}^\infty\sum_{i=1}^\infty|a_{ij}|$$

converges to the same value.

Proof: First we show that the second series converges. For the sake of contradiction, suppose it doesn't. Since all the terms are positive, there are two cases in which the double doesn't converge: for some j, the single sum in i doesn't converge, or the sum over j of the single sums doesn't converge.

Suppose for some j, the single sum $\sum_{i=1}^{\infty} |a_{ij}|$ doesn't converge. Then note

$$+\infty = \sum_{i=1}^{\infty} |a_{ij}| \leq \sum_{i=1}^{\infty} \left(\sum_{j=1}^{\infty} |a_{ij}| \right),$$

but this contradicts the first double sum converging.

For the second case, let $p_j := \sum_{i=1}^{\infty} |a_{ij}|$. Then $\sum_{j=1}^{\infty} p_j$ doesn't converge. Fix large $M \geq 0$. Then by definition, there exists N_M for which

$$\sum_{j=1}^{N_M} p_j \geq M.$$

Since there are finitely many p_j in this sum, there exists J for which

$$\sum_{i=1}^{J} |a_{ij}| > p_j - \frac{\varepsilon}{N_M}$$

for all $1 \leq j \leq N_M$. Thus we have

$$M \leq \sum_{j=1}^{N_M} \sum_{i=1}^{J} |a_{ij}| + \varepsilon = \sum_{i=1}^{J} \sum_{j=1}^{N_M} |a_{ij}| + \varepsilon.$$

Since the first iterated series converges, the inner sums are bounded by their infinite sum value, so the right side is at most $\sum_{i=1}^{J} \sum_{j=1}^{\infty} |a_{ij}| + \varepsilon$. This implies that the first double sum gets arbitrarily large (we can pick $\varepsilon = \frac{1}{2}$ for concreteness), since M was arbitrary, so the first double sum cannot converge, contradiction.

Now we prove they converge to the same value. Define $b_i := \sum_{j=1}^{\infty} |a_{ij}|$ and $c_j := \sum_{i=1}^{\infty} |a_{ij}|$. Then

$$\sum_{i=1}^{\infty} \sum_{j=1}^{\infty} |a_{ij}| = \sum_{i=1}^{\infty} b_i = S_1 \quad \text{and} \quad \sum_{j=1}^{\infty} \sum_{i=1}^{\infty} |a_{ij}| = \sum_{i=1}^{\infty} c_j = S_2.$$

Pick $\varepsilon > 0$. Then there exists N for which $S_1 + \varepsilon > \sum_{i=1}^{N} b_i > S_1 - \varepsilon$ and $S_2 + \varepsilon > \sum_{j=1}^{N} c_j > S_2 - \varepsilon$. Then, since we're only dealing with a finite amount of infinite sums (namely b_i, c_j for $i, j \leq N$), there exists M for which $b_i + \frac{\varepsilon}{N} > \sum_{i=1}^{M} |a_{ij}| > b_i - \frac{\varepsilon}{N}$ and $c_j + \frac{\varepsilon}{N} > \sum_{j=1}^{M} |a_{ij}| > c_j - \frac{\varepsilon}{N}$. Plugging these in to the first inequalities yields

$$S_1 + 2\varepsilon > \sum_{i=1}^{N} \sum_{j=1}^{M} |a_{ij}| > S_1 - 2\varepsilon \quad \text{and} \quad S_2 + 2\varepsilon > \sum_{j=1}^{N} \sum_{i=1}^{M} |a_{ij}| > S_2 - 2\varepsilon.$$

Now let $P = \max\{M, N\}$. Note that each double sum is bounded above by their corresponding value, so increasing both upper indices to P will still keep both double sums bounded, yielding

$$S_1 \geq \sum_{i=1}^{P} \sum_{j=1}^{P} |a_{ij}| > S_1 - 2\varepsilon \quad \text{and} \quad S_2 \geq \sum_{j=1}^{P} \sum_{i=1}^{P} |a_{ij}| > S_2 - 2\varepsilon.$$

Now suppose for the sake of contradiction that $S_1 \neq S_2$. Then letting $\varepsilon = |S_1 - S_2|/2$ would yield a contradiction, since it would implie $\sum_{i=1}^{P} \sum_{j=1}^{P} |a_{ij}| > \sum_{j=1}^{P} \sum_{i=1}^{P} |a_{ij}|$ or vice versa.■

Theorem (Fubini's theorem for sums): Suppose

$$\sum_{i=1}^{\infty} \sum_{j=1}^{\infty} |a_{ij}|$$

converges. Then both $\sum_{i=1}^{\infty} \sum_{j=1}^{\infty} a_{ij}$ and $\sum_{j=1}^{\infty} \sum_{i=1}^{\infty} a_{ij}$ converge, and

$$\lim_{n\to\infty} s_{nn} = \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} a_{ij} = \sum_{j=1}^{\infty} \sum_{i=1}^{\infty} a_{ij},$$

where $s_{nn} = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij}$.

Proof: From the previous lemma, we know that both double sums converge, so we just need to prove the second equation. First we show the limit exists.

Let $t_{mn} = \sum_{i=1}^{m} \sum_{j=1}^{n} |a_{ij}|$. Note that (t_{nn}) is increasing and bounded by $\sum_{i=1}^{\infty} \sum_{j=1}^{\infty} |a_{ij}|$, and so converges. Thus the sequence is Cauchy. Then for $n \geq m$, we have

$$|s_{nn} - s_{mm}| \leq \sum_{i=m+1}^{n} \sum_{j=1}^{n} |a_{ij}| + \sum_{i=1}^{n} \sum_{j=m+1}^{n} |a_{ij}| + \sum_{i=m+1}^{n} \sum_{j=m+1}^{n} |a_{ij}| = |t_{nn} - t_{mm}|.$$

Since (t_{nn}) is Cauchy, we can make the right arbitrarily small for large enough n, m, so (s_{nn}) is also Cauchy.

Now let $\lim_{n\to\infty} s_{nn} = S$. We need to show that S equals the double sums. We only show it's equal to the first, as the second follows similarly. We have

$$|s_{mn} - S| \leq |s_{mn} - s_{nn}| + |s_{nn} - S|.$$

For the first term, assuming without loss of generality that $n \geq m$, we have

$$\begin{aligned} |s_{mn} - s_{nn}| \leq \sum_{i=m+1}^{n} \sum_{j=1}^{n} |a_{ij}| &\leq \sum_{i=m+1}^{n} \sum_{j=1}^{n} |a_{ij}| + \sum_{i=1}^{n} \sum_{j=m+1}^{n} |a_{ij}| + \sum_{i=m+1}^{n} \sum_{j=m+1}^{n} |a_{ij}| \\ &= |t_{nn} - t_{mm}|. \end{aligned}$$

Thus

$$|s_{mn} - S| \leq |t_{nn} - t_{mm}| + |s_{nn} - S|.$$

Since (t_{nn}) is Cauchy, and since $s_{nn} \to S$, there exists N for which $n, m \geq N$ implies both terms are less than $\frac{\varepsilon}{2}$. Thus

$$|s_{mn} - S| < \varepsilon$$

for all $n, m \geq N$. Letting $n \to \infty$, then $m \to \infty$ (which we can do since we know the iterated series converges) yields

$$\left| \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} a_{ij} - S \right| < \varepsilon.$$

Since ε is arbitrary, the two are equal. ■

12.5. Limit Derivative

Theorem: Suppose $f_k : [a, b] \to \mathbb{R}$ and assume each f_k is differentiable. If (f_n') converges uniformly to g, and there exists some $x_0 \in [a, b]$ such that $(f_k(x_0))$ converges, then (f_k) converges uniformly to some f with $f' = g$.

Remark: The condition on (f_n) converging at some point is needed so that the sequence of functions doesn't blow up to infinity because of some increasing constant that disappears under differentiation.

Proof: First we show that f uniformly converges. We have

$$\begin{aligned}|f_n(x) - f_m(x)| &\leq |(f_n(x) - f_m(x)) - (f_n(x_0) - f_m(x_0))| + |f_n(x_0) - f_m(x_0)| \\ &= |(x - x_0)(f_n'(c) - f_m'(c))| + |f_n(x_0) - f_m(x_0)| \\ &\leq |a - b||f_n'(c) - f_m'(c)| + |f_n(x_0) - f_m(x_0)|,\end{aligned}$$

where the equality came from using the mean value theorem on $f_n - f_m$ with c between x and x_0. Since (f_n') converges uniformly, the sequence is uniformly Cauchy, so there exists N_1 such that $n, m \geq N_1 \Rightarrow |f_n'(c) - f_m'(c)| < \frac{\varepsilon}{2|a-b|}$. Since $(f_k(x_0))$ converges, it's also Cauchy, so there exists N_2 such that $n, m \geq N_2 \Rightarrow |f_n(x_0) - f_m(x_0)| < \frac{\varepsilon}{2}$. Letting $N = \max\{N_1, N_2\}$, we have for any $n, m \geq N$ that

$$|a - b||f_n'(c) - f_m'(c)| + |f_n(x_0) - f_m(x_0)| < |a - b| \cdot \frac{\varepsilon}{2|a - b|} + \frac{\varepsilon}{2} = \varepsilon.$$

Thus (f_k) is uniformly Cauchy, and so uniformly converges to some function f.

Next we show that $f' = g$. We have

$$\left|\frac{f(x) - f(c)}{x - c} - g(c)\right| \leq \left|\frac{f(x) - f(c)}{x - c} - \frac{f_n(x) - f_n(c)}{x - c}\right| + \left|\frac{f_n(x) - f_n(c)}{x - c} - f_n'(c)\right| + |f_n'(c) - g(c)|.$$

Consider

$$\frac{f_m(x) - f_m(c)}{x - c} - \frac{f_n(x) - f_n(c)}{x - c}.$$

As we did in the first part, we can use the mean value theorem on $f_m(x) - f_n(x)$ to obtain

$$\frac{f_m(x) - f_m(c)}{x - c} - \frac{f_n(x) - f_n(c)}{x - c} = f_m'(y) - f_n'(y)$$

for some y in between x and c. Since (f_n') converges uniformly, there exists N_1 such that $n, m \geq N_1$ implies

$$\left|\frac{f_m(x) - f_m(c)}{x - c} - \frac{f_n(x) - f_n(c)}{x - c}\right| = |f_m'(y) - f_n'(y)| < \frac{\varepsilon}{3}.$$

Letting $m \to \infty$ yields that for any $n \geq N_1$, we have

$$\left| \frac{f(x) - f(c)}{x - c} - \frac{f_n(x) - f_n(c)}{x - c} \right| < \frac{\varepsilon}{3}.$$

Since (f_n') converges uniformly to g, there exists N_2 such that $n \geq N_2 \Rightarrow |f'(c) - g(c)| < \frac{\varepsilon}{3}$. Let $N = \max\{N_1, N_2\}$. Since f_N is differentiable, there exists δ such that $0 < |x - c| < \delta$ implies

$$\left| \frac{f_N(x) - f_N(c)}{x - c} - f_N'(c) \right| < \frac{\varepsilon}{3}.$$

Combining everything with the initial inequality yields

$$\begin{aligned} \left| \frac{f(x) - f(c)}{x - c} - g(c) \right| &\leq \left| \frac{f(x) - f(c)}{x - c} - \frac{f_N(x) - f_N(c)}{x - c} \right| + \left| \frac{f_N(x) - f_N(c)}{x - c} - f_N'(c) \right| + |f_N'(c) - g(c)| \\ &< \frac{\varepsilon}{3} + \frac{\varepsilon}{3} + \frac{\varepsilon}{3} = \varepsilon. \end{aligned}$$

Thus f is differentiable at c with derivative $g(c)$, as desired. ■

12.6. Limit Integral

Theorem: Suppose each $f_k : [a, b] \to \mathbb{R}$ is integrable. If (f_k) converges uniformly to f, then f is integrable, and

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

Proof:

Lemma: For bounded f and g on $[a, b]$, we have

$$U(f + g) \leq U(f) + U(g)$$

and

$$L(f + g) \geq L(f) + L(g).$$

Proof: We prove the upper sum case, as the lower sum case follows similarly. For any partition P, we have

$$U(f + g) \leq U(f + g, P) \leq U(f, P) + U(g, P).$$

Since $U(f)$ is an infinum, there exists a sequence of partitions such that $U(f, P_n)$ approaches $U(f)$. Similarly, there exists such a sequence of partitions for $U(g)$. Taking the union of each term in the sequence of partitions gives a sequence for which both terms converge to their upper sums. Since the inequality above holds for all partitions, we obtain $U(f + g) \leq U(f) + U(g)$, as desired. ■

Since each f_k is integrable, each is bounded, which implies f is bounded by our boundedness results.

Now we can prove that $L(f) = U(f)$. By uniform convergence, there exists N such that $k \geq N$ implies

$$|f_k(x) - f(x)| < \frac{\varepsilon}{2(b-a)}$$

for all $x \in [a, b]$.

Then we have

$$\begin{aligned} U(f) - L(f) &= U(f - f_N + f_N) - L(f - f_N + f_N) \\ &\leq U(f - f_N) + U(f_N) - L(f - f_N) - L(f_N), \end{aligned}$$

where the inequality comes from the previous proposition. Since f_N is integrable, $U(f_N) = L(f_N)$, we get $U(f) - L(f) \leq U(f - f_N) - L(f - f_N)$. From uniform convergence, we have $-\frac{\varepsilon}{2(b-a)} < f_N(x) - f(x) < \frac{\varepsilon}{2(b-a)}$. Then we get

$$U(f) - L(f) \leq U(f - f_N) - L(f - f_N) < U\left(\frac{\varepsilon}{2(b-a)}\right) - L\left(-\frac{\varepsilon}{2(b-a)}\right) = \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

Thus $0 \leq U(f) - L(f) < \varepsilon$, and so $U(f) - L(f) = 0$. Thus f is integrable.

Now we prove the integral converges to the integral of the convergent function. By uniform convergence, there exists N such that $k \geq N \Rightarrow |f_k(x) - f(x)| < \frac{\varepsilon}{b-a}$ for all $x \in [a, b]$. Thus

$$f_k(x) - \frac{\varepsilon}{b-a} < f(x) < f_k(x) + \frac{\varepsilon}{b-a}$$

for all $k \geq N$ and $x \in [a, b]$. Integrating both sides yields

$$\int_a^b f_k(x) - \frac{\varepsilon}{b-a}\, dx = \int_a^b f_k(x)\, dx - \varepsilon < \int_a^b f(x)\, dx < \int_a^b f_k(x) + \frac{\varepsilon}{b-a}\, dx = \int_a^b f_k(x)\, dx + \varepsilon$$

for all $k \geq N$, which implies

$$\left| \int_a^b f_k(x)\, dx - \int_a^b f(x)\, dx \right| < \varepsilon,$$

and so the sequence does converge to $\int_a^b f(x)\, dx$.

12.7. Limit Sum

12.8. Derivative Integral

Theorem (Leibniz integral rule): Let $f(x, y)$ be a function such that both f and $\frac{\partial f}{\partial x}$ are continuous in some region of the xy-plane, including $a(x) \leq y \leq b(x)$, $x_0 \leq x \leq x_1$. Also suppose that both $a(x)$ and $b(x)$ are differentiable on $x_0 < x < x_1$. Then for x in this range, we have

$$\frac{d}{dx}\left(\int_{a(x)}^{b(x)} f(x, y)\, dy\right) = f(x, b(x))b'(x) - f(x, a(x))a'(x) + \int_{a(x)}^{b(x)} \frac{\partial f}{\partial x}(x, y)\, dy.$$

Proof: We split the integral as $\int_0^{b(x)} f(x, y)\, dy - \int_0^{a(x)} f(x, y)\, dy$ and show the result for the lower bound being 0 (we can assume without loss of generality that $a(x) = 0$ in the problem statement to ensure continuity in the region of integration). We want to compute

$$\begin{aligned} &\lim_{h \to 0} \frac{1}{h}\left(\int_0^{b(x+h)} f(x+h, y)\, dy - \int_0^{b(x)} f(x, y)\, dy\right) \\ &= \lim_{h \to 0} \frac{1}{h}\left(\int_{b(x)}^{b(x+h)} f(x+h, y)\, dy + \int_0^{b(x)} f(x+h, y) - f(x, y)\, dy\right). \end{aligned}$$

Let I_1 be the first integral and I_2 be the second integral. We show that the limits of I_1 and I_2 as $h \to 0$ exist, and thus we can split the limit and obtain the desired formula.

First we compute I_2. Pick $\varepsilon > 0$. Note that since $\frac{\partial f}{\partial x}$ is continuous on a compact set (since the x and y regions of continuity are compact, their product must be as well), it must be uniformly continuous. Thus, there exists $\delta > 0$ for which

$$\|(x_1, y_1) - (x_2, y_2)\| < \delta \Rightarrow \left|\frac{\partial f}{\partial x}(x_1, y_1) - \frac{\partial f}{\partial x}(x_2, y_2)\right| < \varepsilon.$$

If we let y_1 and y_2 be an arbitrary y, then we have

$$|x_1 - x_2| < \delta \Rightarrow \left|\frac{\partial f}{\partial x}(x_1, y) - \frac{\partial f}{\partial x}(x_2, y)\right| < \varepsilon.$$

Let $0 < |h| < \delta$. Then by the mean value theorem,

$$\frac{f(x+h, y) - f(x, y)}{h} = \frac{\partial f}{\partial x}(x + c, y)$$

for some c in between 0 and h (whether h is positive or negative) and for all y. Since clearly $0 < |c| < \delta$, we have

$$\left|\frac{f(x+h, y) - f(x, y)}{h} - \frac{\partial f}{\partial x}(x, y)\right| = \left|\frac{\partial f}{\partial x}(x + c, y) - \frac{\partial f}{\partial x}(x, y)\right| < \varepsilon$$

for all $0 < |h| < \delta$ and for all y.

Since f and $\frac{\partial f}{\partial x}$ are continuous, they are both integrable, and since the above equation holds for all y, we can integrate with respect to y to obtain

$$\begin{aligned}\left|\int_0^{b(x)} \frac{f(x+h,y)-f(x,y)}{h}\,dy - \int_0^{b(x)} \frac{\partial f}{\partial x}(x,y)\,dy\right| &\leq \int_0^{b(x)} \left|\frac{f(x+h,y)-f(x,y)}{h} - \frac{\partial f}{\partial x}(x,y)\right| dy \\ &< \varepsilon b(x).\end{aligned}$$

Since $b(x)$ is constant with respect to y, we do indeed have

$$\lim_{h\to 0}\int_0^{b(x)} \frac{f(x+h,y)-f(x,y)}{h}\,dy = \int_0^{b(x)} \frac{\partial f}{\partial x}(x,y)\,dy.$$

Now we compute I_1. We have two cases: $b'(x) = 0$ and $b'(x) \neq 0$.

First suppose $b'(x) = 0$. Since f is continuous, on the region we're interested in, f is bounded by some M. Thus we have

$$\left|\frac{1}{h}\int_{b(x)}^{b(x+h)} f(x+h,y)\,dy\right| \leq M\left|\frac{b(x+h)-b(x)}{h}.\right|$$

Taking the limit as $h \to 0$, the squeeze theorem yields that

$$\lim_{h\to 0}\frac{1}{h}\int_{b(x)}^{b(x+h)} f(x+h,y)\,dy = 0 = f(x,b(x))b'(x).$$

Now suppose $b'(x) \neq 0$. Thus there exists δ such that $0 < |h| < \delta \Rightarrow b(x+h) \neq b(x)$. Then we have

$$I_1 = \frac{b(x+h)-b(x)}{h}\cdot\frac{1}{b(x+h)-b(x)}\int_{b(x)}^{b(x+h)} f(x+h,y)\,dy.$$

Since $b(x+h) \neq b(x)$ is a neighborhood of x, the second fraction is well defined. Then from the mean value theorem, there exists $t(h)$ between $b(x)$ and $b(x+h))$ such that the integral is equal to $f(x+h,t(h))$. Thus

$$I_1 = \frac{b(x+h)-b(x)}{h}f(x+h,t(h)).$$

Note that $\lim_{h\to 0} t(h) = b(x)$ (since by cotinuity $b(x+h) \to b(x)$ as $h \to 0$). Thus we have

$$\lim_{h\to 0} I_1 = \lim_{h\to 0}\left(\frac{b(x+h)-b(x)}{h}\right)\lim_{h\to 0} f(x+h,t(h)) = b'(x)f(x,b(x)),$$

where the last limit exists by continuity of f. Thus we have our desired formula. ■

12.9. Derivative Sum

12.10. Integral Sum

The following section of the notes dives into analysis on manifolds. Multivariable calculus stuff will be found here.

MAKE BIG TITLE HERE FOR THAT

13. Multivariable Differential Calculus

Our goal is to find a definition of differentiabilty for functions from $\mathbb{R}^n$ to $\mathbb{R}^m$. Add some more motivating stuff here.

13.1. Derivative

Definition (differentiabilty): Let E be a subset of $\mathbb{R}^n$, let $f : E \to \mathbb{R}^m$ be a function, let $x_0 \in E$ be a limit point of E, and let $L : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation. Then f is *differentiable* at x_0 with derivative L if

$$\lim_{x \to x_0} \frac{\|f(x) - (f(x_0) + L(x - x_0))\|}{\|x - x_0\|} = 0,$$

where $\|\cdot\|$ denotes the l^2 metric. This if often referred to as the *total derivative* of f.

Example:

Proposition: Suppose $f : E \to \mathbb{R}^n$, where $E \subseteq \mathbb{R}^m$, is differentiable at x_0. Then the derivative at x_0 is unique.

Proof: Suppose there exist two linear tranforms, $L_1, L_2 : \mathbb{R}^m \to \mathbb{R}^n$, that are derivatives of f at x_0. Then there exists v such that $L_1 v \neq L_2 v$. From the definition of a limit, there exists $B_r(x_0)$ such that

$$x \in B_r(x_0) \Rightarrow \|f(x) - f(x_0) - L_i(x - x_0)\| < \varepsilon \|x - x_0\|$$

for both $i = 1, 2$. Let $x = x_0 + tv$, where t is an arbitrary scalar for which $x \in B_r(x_0)$. Adding the two inequalities and using the triangle inequality on the left yields

$$\|L_1(tv) - L_2(tv)\| < 2\varepsilon \|tv\| \Rightarrow \|L_1 v - L_2 v\| < 2\varepsilon \|v\|.$$

Note that $\|v\|$ is fixed, and ε is arbitrary, so the rigt side gets arbitrarily small. However, the left side is a nonzero constant, so we have a contradiction. ■

Remark: Since we've established uniqueness, we often denote the derivative at x_0 as $f'(x_0)$, but be warned that this denotes a linear transformation, not a scalar.

Proposition: Let $L : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation. Then L is uniformly continuous.

Proof: Let A be the matrix representation of L with respect to the standard basis, and let S be the sum of the squares of the entries of M. Letting $x = (x_1, ..., x_n)^t$, we have

$$\begin{aligned}\|Lx\|^2 &= \sum_{r=1}^{m} \left(a_{r,1}x_1 + a_{r,2}x_2 + \cdots + a_{r,n}x_n\right)^2 \\ &\leq \sum_{r=1}^{m} \left(a_{r,1}^2 + \cdots + a_{r,n}^2\right)\left(x_1^2 + \cdots + x_n^2\right) = S\|x\|^2,\end{aligned}$$

where the inequality follows by Cauchy-Schwarz. Plugging in $x - y$ yields

$$\|Lx - Ly\| \leq \sqrt{S}\|x - y\|.$$

Thus L is Lipschitz, and so it uniformly continuous, as desired. ∎

Proposition: If $f : E \to \mathbb{R}^n$ is differentiable at x_0, then it's continuous at x_0.

Proof: We want to show that $\lim_{x \to x_0} \|f(x) - f(x_0)\| = 0$. Set $f'(x_0) = L$. We have

$$\begin{aligned}\|f(x) - f(x_0)\| &\leq \|f(x) - f(x_0) - L(x - x_0)\| + \|L(x - x_0)\| \\ &= \frac{\|f(x) - f(x_0) - L(x - x_0)\|}{\|x - x_0\|} \cdot \|x - x_0\| + \|L(x - x_0)\|.\end{aligned}$$

Taking limits on the right yields

$$\lim_{x \to x_0} \left(\frac{\|f(x) - f(x_0) - L(x - x_0)\|}{\|x - x_0\|} \cdot \|x - x_0\| + \|L(x - x_0)\| \right) =$$

$$\left(\lim_{x \to x_0} \frac{\|f(x) - f(x_0) - L(x - x_0)\|}{\|x - x_0\|} \right) \left(\lim_{x \to x_0} \|x - x_0\| \right) + \lim_{x \to x_0} \|L(x - x_0)\| = 0,$$

where the last limit comes from the fact that linear maps are continuous. Thus we have $\lim_{x \to x_0} \|f(x) - f(x_0)\| = 0$, as desired. ∎

Theorem (chain rule): Let E be a subset of $\mathbb{R}^n$ and F be a subset of $\mathbb{R}^m$. Let $g : E \to F$ and $f : F \to \mathbb{R}^p$. Let c be an interior point of E. If g is differentiable at c, $g(c)$ is an interior point of F, and f is differentiable at $g(c)$, then $f \circ g : E \to \mathbb{R}^p$ is differentiable at c with derivative

$$(f \circ g)'(c) = f'(g(c))g'(c).$$

Proof: Let M be the Lipschitz constant of the linear transformation $f'(g(c))$ (which we know exists from the proof that linear transformations are uniformly continuous), and let S be the Lipschitz constant of $g'(c)$.

Let $(x_n) \in E$ be an arbitrary sequence such that $x_n \to c$. Since $g'(c)$ exists, we know that

$$\lim_{n \to \infty} \frac{\|g(x_n) - g(c) - g'(c)(x_n - c)\|}{\|x_n - c\|} = 0.$$

Thus, for $\varepsilon > 0$, there exists N_1 such that $n \geq N_1$ implies

$$\|g(x_n) - g(c) - g'(c)(x_n - c)\| < \varepsilon\|x_n - c\|.$$

Since g is differentiable at c, it's also continuous there, so $g(x_n) \to g(c)$. Now we split into two cases:

- **Case 1:** $g(x_n) = g(c)$ finitely many times.
  Since $g(x_n) \to g(c)$, since f is differentiable at $g(c)$, and since after a certain point, $g(x_n) \neq g(c)$, we have

  $$\lim_{n\to\infty} \frac{\|f(g(x_n)) - f(g(c)) - f'(g(c))(g(x_n) - g(c))\|}{\|g(x_n) - g(c)\|} = 0.$$

  This follows from the fact that if a sequence $(y_n) \in F$ such that $y_n \to g(c)$, then

  $$\lim_{n\to\infty} \frac{\|f(y_n) - f(g(c)) - f'(g(c))(y_n - g(c))\|}{\|y_n - g(c)\|} = 0,$$

  and since we eventually don't have any divide by zero issues, we can replace y_n with $g(x_n)$. Thus, there exists N_2 such that $n \geq N_2$ implies

  $$\|f(g(x_n)) - f(g(c)) - f'(g(c))(g(x_n) - g(c))\| < \varepsilon\|g(x_n) - g(c)\|.$$

  From the linear transformation proposition, we know that

  $$\|f'(g(c))x\| \leq M\|x\|.$$

  Thus

  $$\begin{aligned}\|f'(g(c))(g(x_n) - g(c)) - f'(g(c))g'(c)(x_n - c)\| &\leq M\|g(x_n) - g(c) - g'(c)(x_n - c)\| \\ &< M\varepsilon\|x_n - c\|.\end{aligned}$$

  Let $N = \max\{N_1, N_2\}$. Applying the triangle inequality, we have for $n \geq N$ that

  $$\begin{aligned}\|f(g(x_n)) - f(g(c)) - f'(g(c)g'(c))(x_n - c)\| &\leq \|f(g(x_n)) - f(g(c)) - f'(g(c))(g(x_n) - g(c))\| \\ &\quad + \|f'(g(c))(g(x_n) - g(c)) - f'(g(c))g'(c)(x_n - c)\| \\ &< \varepsilon\|g(x_n) - g(c)\| + M\varepsilon\|x_n - c\|.\end{aligned}$$

  Now we just need to bound by right side. Since the second term is already in that form, we focus on the first. We have

  $$\begin{aligned}\|g(x_n) - g(c)\| &\leq \|g(x_n) - g(c) - g'(c)(x_n - c)\| + \|g'(c)(x_n - c)\| \\ &< \varepsilon\|x_n - c\| + S\|x_n - c\|.\end{aligned}$$

  Thus the right side is bounded by

  $$\|x_n - c\|(\varepsilon^2 + S\varepsilon + M\varepsilon).$$

  Thus, for $n \geq N$, we have

  $$\frac{\|f(g(x_n)) - f(g(c)) - f'(g(c)g'(c))(x_n - c)\|}{\|x_n - c\|} < \varepsilon^2 + (S + M)\varepsilon,$$

  where S and M are independent of ε. Clearly the right side gets arbitrarily small, so this case is done.

- **Case 2:** $g(x_n) = g(c)$ infinitely often.
  We split the sequence into two subsequences such that one subsequence contains all terms such that $g(x_n) = g(c)$, and the other subsequence contains every other term. From case 1,

we know that limit we're looking for is equal to 0 for the non constant sequence, so we just need to show that for the constant sequence, the limit is also 0, after which it's easy to see that the combined original sequence will have limit 0, and then the proof will be complete, since the limit will be 0 for any arbitrary sequence.
Since g is differentiable at c, we have

$$\lim_{n\to\infty} \frac{\|g(x_n) - g(c) - g'(c)(x_n - c)\|}{\|x_n - c\|} = \lim_{n\to\infty} \frac{\|g'(c)(x_n - c)\|}{\|x_n - c\|} = 0.$$

We also have

$$\begin{aligned} \frac{\|f(g(x_n)) - f(g(c)) - f'(g(c))g'(c)(x_n - c)\|}{\|x_n - c\|} &= \frac{\|f'(g(x))g'(c)(x_n - c)\|}{\|x_n - c\|} \\ &\leq \frac{M\|g'(c)(x_n - c)\|}{\|x_n - c\|}. \end{aligned}$$

Thus by the squeeze theorem, the limit of the left side as $n \to \infty$ is 0, as desired.

■

Corollary: Let E be an open subset of $\mathbb{R}^n$, let $f : E \to \mathbb{R}^m$ be a differentiable function at x_0, and let $T : \mathbb{R}^m \to \mathbb{R}^p$ be a linear transformation. Then T is continuously differentiable everywhere, has derivative T, and

$$(T \circ f)'(x_0) = T'(f(x_0))f'(x_0) = Tf'(x_0).$$

Proof: The formula follows easily via the chain rule, so we just need to show that T is continuously differentiable. We have

$$\frac{\partial T}{\partial e_i}(x) = \lim_{t\to 0} \frac{T(x + te_i) - T(x)}{t} = \lim_{t\to 0} Te_i = Te_i.$$

Since the partials are constant, they're clearly continuous everywhere, so T is continuously differentiable everywhere. We also have

$$\lim_{x\to x_0} \frac{\|T(x) - T(x_0) - T(x - x_0)\|}{\|x - x_0\|} = \lim_{x\to x_0} 0 = 0,$$

so clearly $T'(x_0) = T$, as desired. ■

13.2. Partial and Directional Derivatives

Definition (directional derivative): Let E be a subset of $\mathbb{R}^n$, let $f : E \to \mathbb{R}^m$ be a function, let x_0 be an interior point of E, and let v be a vector in $\mathbb{R}^n$. If

$$\lim_{t\to 0^+} \frac{f(x_0 + tv) - f(x_0)}{t}$$

exists, then f is *differentiable in the direction v at x_0*, and is denoted with $D_v f(x_0)$.

Example: Let $f : \mathbb{R}^2 \to \mathbb{R}^3$ be defined by $f(x, y) = (x^2, y^2, x^2y^2)$. Then we have

$$D_{(3,4)}f(1, 2) = \lim_{t\to 0^+} \frac{((1 + 3t)^2, (2 + 4t)^2, (1 + 3t)^2(2 + 4t)^2) - (1, 4, 4)}{t} = (6, 16, 40).$$

Proposition: Let E be a subset of $\mathbb{R}^n$, let $f : E \to \mathbb{R}^m$ be a function, let x_0 be an interior point of E, and let v be a vector in $\mathbb{R}^n$. If f is differentiable at x_0, then f is also differentiable in the direction v at x_0, and

$$D_v f(x_0) = f'(x_0)v.$$

Proof: From the defintion of the derivative, we know there exists δ such that $x \in B_\delta(x_0) \setminus \{x_0\}$ implies

$$\frac{\|f(x) - f(x_0) - f'(x_0)(x - x_0)\|}{\|x - x_0\|} < \frac{\varepsilon}{\|v\|}.$$

Thus, for $0 < t < \frac{\delta}{\|v\|}$, we have

$$\frac{\|f(x_0 + tv) - f(x_0) - tf'(x_0)v\|}{t\|v\|} < \frac{\varepsilon}{\|v\|} \Rightarrow \left\| \frac{f(x_0 + tv) - f(x_0)}{t} - f'(x_0)v \right\| < \varepsilon,$$

as desired. ∎

Definition (partial derivative): Let E be a subset of $\mathbb{R}^n$, let $f : E \to \mathbb{R}^m$ be a function, let x_0 be an interior point of E, and let $1 \leq j \leq n$. Then the *partial derivative* of f with respect to the x_j variable, denoted with $\frac{\partial f}{\partial x_j}(x_0)$, is defined by

$$\lim_{t\to 0} \frac{f(x_0 + te_j) - f(x_0)}{t},$$

provided it exists. Here e_j is a standard basis vector in $\mathbb{R}^n$.

Definition (continuously differentiable): If E is a subset of $\mathbb{R}^n$, the function $f : E \to \mathbb{R}^m$ is *continuously differentiable* on E if the partial derivatives for each of the x_n variables exist and are continuous. Furthermore, we say that f is n *times continuously differentiable* if each partial derivative of f is $n - 1$ times continuously differentiable.

Lemma: Suppose $f : E \to \mathbb{R}^m$, where $E \subseteq \mathbb{R}^n$, is continuously differentiable. Then each component of $f = (f_1, ..., f_m)$ is also continuously differentiable.

Proof: Continuity of the partial derivatives follows from the fact that each component has to be continuous as well, since otherwise $\frac{\partial f}{\partial x_j}$ wouldn't be continuous.

Now we show differentiabilty of the components. By definition, we have

$$\left\| \frac{f(x_0 + te_j) - f(x_0)}{t} - \frac{\partial f}{\partial x_j}(x_0) \right\| < \varepsilon$$

for $0 < t < \delta$ for some $\delta > 0$. Let c_i be the i-th component of $\frac{\partial f}{\partial x_j}(x_0)$. Then we clearly have

$$\left| \frac{f_i(x_0 + te_j) - f_i(x_0)}{t} - c_i \right| \leq \left\| \frac{f(x_0 + te_j) - f(x_0)}{t} - \frac{\partial f}{\partial x_j}(x_0) \right\| < \varepsilon.$$

Thus $\frac{\partial f_i}{\partial x_j}(x_0) = c_j$, as desired. ■

Proposition: Let E be a subset of $\mathbb{R}^n$, let $f : E \to \mathbb{R}^m$ be a function, let F be a subset of E, and let x_0 be an interior point of F. If all partial derivatives $\frac{\partial f}{\partial x_j}$ exist on F and are continuous at x_0, then f is differentiable at x_0, and the derivative $f'(x_0) : \mathbb{R}^n \to \mathbb{R}^m$ is given by

$$f'(x_0)v = \sum_{j=1}^{n} a_j \frac{\partial f}{\partial x_j}(x_0),$$

where $v = a_1 e_1 + \dots + a_n e_n$.

Remark: The expression for $f'(x_0)$ comes from

$$D_v f(x_0) = f'(x_0)v = a_1 f'(x_0)e_1 + \dots + a_n f'(x_0)e_n = \sum_{j=1}^{n} a_j \frac{\partial f}{\partial x_j}(x_0).$$

Proof: Let $L : \mathbb{R}^n \to \mathbb{R}^m$ be the linear transformation in the statement, and pick $\varepsilon > 0$. To show that f is differentiable with derivative L at x_0, we need find $\delta > 0$ such that $x \in B_\delta(x_0) \setminus \{x_0\}$ implies

$$\|f(x) - f(x_0) - L(x - x_0)\| < \varepsilon \|x - x_0\|.$$

We note that for any vector $a = (a_1, \dots, a_k) \in \mathbb{R}^k$, we have $|a_i| \leq \|a\| \leq \sum_{i=1}^{k} |a_i|$, where the second inequality follows from the triangle inequality.

Write $f = (f_1, \dots, f_m)$, where $f_i : E \to \mathbb{R}$. From the previous lemma, we know that each f_i has partial derivatives on F that are continuous at x_0.

From the continuity of the partial derivatives, we know there exists $\delta_j > 0$ such that $\left\| \frac{\partial f}{\partial x_j}(x) - \frac{\partial f}{\partial x_j}(x_0) \right\| < \frac{\varepsilon}{mn}$ for each $\|x - x_0\| < \delta_j$. Let $\delta = \min(\delta_1, \dots, \delta_n)$, and let $x \in B_\delta(x_0) \setminus \{x_0\} = B$. Write $x = x_0 + a_1 e_1 + \dots + a_n e_n$. We now need to show that

$$\left\| f(x_0 + a_1 e_1 + \dots + a_n e_n) - f(x_0) - \sum_{j=1}^{n} a_j \frac{\partial f}{\partial x_j}(x_0) \right\| < \varepsilon \|x - x_0\|.$$

From the mean value theorem in the x_1 variable, we have

$$f_i(x_0 + a_1 e_1) - f_i(x_0) = a_1 \frac{\partial f_i}{\partial x_1}(x_0 + t_1 e_1)$$

for some $0 < t_1 < a_1$. Since clearly $x_0 + t_1 e_1 \in B$, we have

$$\left| \frac{\partial f_i}{\partial x_1}(x_0 + t_1 e_1) - \frac{\partial f_i}{\partial x_1}(x_0) \right| \le \left\| \frac{\partial f}{\partial x_1}(x_0 + t_1 e_1) - \frac{\partial f}{\partial x_1}(x_0) \right\| < \frac{\varepsilon}{mn}.$$

This implies

$$\left| f_i(x_0 + a_1 e_1) - f_i(x_0) - a_1 \frac{\partial f_i}{\partial x_1}(x_0) \right| < \frac{\varepsilon |a_1|}{mn} < \frac{\varepsilon \|x - x_0\|}{mn}.$$

Summing over all $1 \le i \le m$, and using the inequality at the beginning, we obtain

$$\left\| f(x_0 + a_1 e_1) - f(x_0) - a_1 \frac{\partial f}{\partial x_1}(x_0) \right\| < \frac{\varepsilon \|x - x_0\|}{n}.$$

Applying the same method, we obtain

$$\left\| f(x_0 + a_1 e_1 + \dots + a_j e_j) - f(x_0 + a_1 e_1 + \dots + a_{j-1} e_{j-1}) - a_j \frac{\partial f}{\partial x_j}(x_0) \right\| < \frac{\varepsilon \|x - x_0\|}{n}.$$

Summing over $1 \le j \le n$, applying the triangle inequality, and telescoping yields

$$\left\| f(x_0 + a_1 e_1 + \dots + a_n e_n) - f(x_0) - \sum_{j=1}^{n} a_j \frac{\partial f}{\partial x_j}(x_0) \right\| < \varepsilon \|x - x_0\|,$$

as desired. ■

Definition (derivative matrix): Let $E \subseteq \mathbb{R}^n$, and let $f : E \to \mathbb{R}^m$, and write $f = (f_1, \dots, f_m)$. If the partial derivatives of f exist on E, then the *derivative matrix* is the matrix given by

$$Df(x_0) = \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(x_0) & \frac{\partial f_1}{\partial x_2}(x_0) & \cdots & \frac{\partial f_1}{\partial x_n}(x_0) \\ \frac{\partial f_2}{\partial x_1}(x_0) & \frac{\partial f_2}{\partial x_2}(x_0) & \cdots & \frac{\partial f_2}{\partial x_n}(x_0) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1}(x_0) & \frac{\partial f_m}{\partial x_2}(x_0) & \cdots & \frac{\partial f_m}{\partial x_n}(x_0) \end{pmatrix}.$$

Remark: If the partial derivatives are continuous then f is differentiable, and its easy to see that Df is the matrix representation derivative of f with respect to the standard basis.

Theorem (Clairaut's theorem): Let E be an open subset of $\mathbb{R}^n$, let $x_0 \in E$, and let $f : E \to \mathbb{R}^m$ be twice continuously differentiable on E. Then

$$\frac{\partial}{\partial x_j} \frac{\partial f}{\partial x_i}(x_0) = \frac{\partial}{\partial x_i} \frac{\partial f}{\partial x_j}(x_0)$$

for all $1 \le i, j \le n$.

Proof: We work with one component of f at a time, so we can assume $m = 1$. The theorem is obvious for $i = j$, so suppose $i \neq j$. Without loss of generality, assume $x_0 = 0$. Let $\frac{\partial}{\partial x_j}\frac{\partial f}{\partial x_i}(0) = a_1$ and $\frac{\partial}{\partial x_i}\frac{\partial f}{\partial x_j}(0) = a_2$. We need to show that $a_1 = a_2$.

Pick $\varepsilon > 0$. From the continuity of the double derivatives, there exists $\delta > 0$ such that if $\|x\| < 2\delta$, we have

$$\left|\frac{\partial}{\partial x_j}\frac{\partial f}{\partial x_i}(x) - a_1\right| < \varepsilon \quad \text{and} \quad \left|\frac{\partial}{\partial x_i}\frac{\partial f}{\partial x_j}(x) - a_2\right|.$$

Define

$$X = f(\delta e_i + \delta e_j) - f(\delta e_i) - f(\delta e_j) + f(0).$$

From the fundamental theorem of calculus in x_i, we have

$$f(\delta e_i + \delta e_j) - f(\delta e_i) = \int_0^\delta \frac{\partial f}{\partial x_i}(x_i e_i + \delta e_j)\, dx_i \quad \text{and} \quad f(\delta e_j) - f(0) = \int_0^\delta \frac{\partial f}{\partial x_i}(x_i e_i)\, dx_i,$$

so

$$X = \int_0^\delta \left(\frac{\partial f}{\partial x_i}(x_i e_i + \delta e_j) - \frac{\partial f}{\partial x_i}(x_i e_i)\right) dx_i.$$

From the mean value theorem in the x_j variable, for each $x_i \in [0, \delta]$, there exists $t_{x_i} \in (0, \delta)$ such

$$\frac{\partial f}{\partial x_i}(x_i e_i + \delta e_j) - \frac{\partial f}{\partial x_i}(x_i e_i) = \delta \frac{\partial}{\partial x_j}\frac{\partial f}{\partial x_i}(x_i e_i + t_{x_i} e_j).$$

Thus by construction we have

$$\left|\frac{\partial f}{\partial x_i}(x_i e_i + \delta e_j) - \frac{\partial f}{\partial x_i}(x_i e_i) - \delta a_1\right| < \varepsilon\delta.$$

Integrating both sides yields

$$\begin{aligned} |X - \delta^2 a_1| &= \left|\int_0^\delta \left(\frac{\partial f}{\partial x_i}(x_i e_i + \delta e_j) - \frac{\partial f}{\partial x_i}(x_i e_i) - \delta a_1\right) dx_i\right| \\ &\leq \int_0^\delta \left|\frac{\partial f}{\partial x_i}(x_i e_i + \delta e_j) - \frac{\partial f}{\partial x_i}(x_i e_i) - \delta a_1\right| dx_i < \varepsilon\delta^2. \end{aligned}$$

Swapping the roles of i and j, we similarly obtain $|X - \delta^2 a_2| < \varepsilon\delta^2$. Applying the triangle inequality yields $|\delta^2 a_1 - \delta^2 a_2| < 2\varepsilon\delta^2 \Rightarrow |a_1 - a_2| < 2\varepsilon$. Since ε is arbitrary, we have $a_1 = a_2$, as desired. ■

13.3. Inverse Function Theorem

Lemma: Let $B_r(0) \in \mathbb{R}^n$ and let $g : B_r(0) \to \mathbb{R}^n$ be a map such that $g(0) = 0$ and

$$\|g(x) - g(y)\| \leq \frac{1}{2}\|x - y\|$$

for all $x, y \in B_r(0)$. Then $f : B_r(0) \to \mathbb{R}^n$ defined by $f(x) = x + g(x)$ is injective, and the image $f(B_r(0))$ contains the ball $B_{r/2}(0)$.

Proof: First we show f is injective. If $f(x) = f(y)$, then $x + g(x) = y + g(y) \Rightarrow \|x - y\| = \|g(x) - g(y)\| \leq \frac{1}{2}\|x - y\|$, which is only possible if $x = y$.

Now we show that second claim. Pick $y \in B_{r/2}(0)$. We need to find $x \in B_r(0)$ such that $f(x) = y \Rightarrow x = y - g(x)$. Thus, if we let $F(x) : B_r(0) \to \mathbb{R}^n$ deonte the function $F(x) = y - g(x)$, we want to find a fixed point of F. We do this using the contraction mapping theorem, so we need to show that some closed subset of $B_r(0)$ (and thus complete) maps into itself.

Since $B_{r/2}(0)$ is open, some $\varepsilon/2$ neighborhood centered at y lies entirely within the ball. Then, if $x \in \overline{B_{r-\varepsilon}(0)}$, we have

$$\|F(x)\| \leq \|y\| + \|g(x)\| \leq \frac{r-\varepsilon}{2} + \|g(x) - g(0)\| \leq \frac{r-\varepsilon}{2} + \frac{1}{2}\|x - 0\| \leq \frac{r-\varepsilon}{2} + \frac{r-\varepsilon}{2} = r - \varepsilon.$$

Thus $F\left(\overline{B_{r-\varepsilon}(0)}\right) \subseteq \overline{B_{r-\varepsilon}(0)}$. Furthermore, for any $x, x' \in B_r(0)$, we have

$$\|F(x) - F(x')\| = \|g(x) - g(x')\| \leq \frac{1}{2}\|x' - x\|.$$

Thus F is a strict contraction on $B_r(0)$, and therefore clearly a strict contraction on $\overline{B_{r-\varepsilon}(0)}$. Thus by the contraction mapping theorem, F has some fixed point $x \in B_r(0)$, and thus $F(x) = x = y - g(x) \Rightarrow f(x) = y$, as desired. ■

Remark: This lemma essentially says that small perturbations of the identity function remain injective and cannot create any holes in the ball.

Theorem (inverse function theorem): Let E be an open subset of $\mathbb{R}^n$, and let $f : E \to \mathbb{R}^n$ be a function which is continuously differentiable on E. Suppose there exists $x_0 \in E$ such that the linear transformation $f'(x_0) : \mathbb{R}^n \to \mathbb{R}^n$ is invertible. Then there exists an open set U in E containing x_0 and an open set V in $\mathbb{R}^n$ containing $f(x_0)$ such that f is a bijection from U to V. In particular, there is an inverse map $f^{-1} : V \to U$. Furthermore, this inverse map is differentiable at $f(x_0)$ with derivative

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

Proof: If f^{-1} is differentiable, then the formula follows easily by the chain rule. Since $f^{-1} \circ f = I$, where I is the identity map, differentiating yields $\left(f^{-1}\right)'(f(x_0))f'(x_0) = I'(x_0) = I$, and multiplying by the inverse of $f'(x_0)$ on both sides yields the desired formula.

We perform a series of simplifications on the conditions on f. First, it's enough to show the theorem in the special case where $f(x_0) = 0$. The general case follows by applying the special case to the new function $\tilde{f}(x) = f(x) - f(x_0)$: If f is continuously differentiable, then clearly so is $\tilde{f}$, and if $f'(x_0)$ is invertible, then $\tilde{f}'(x_0) = f'(x_0)$ is invertible, so there exist open sets U containing x_0 and V containing $\tilde{f}(x_0) = f(x_0) - f(x_0) = 0$ for which $\tilde{f} : U \to V$ is a bijection and for which the inverse map is differentiable at $\tilde{f}(x_0) = 0$. Thus $f(x) = \tilde{f}(x) + f(x_0) : U \to V + f(x_0)$ is a bijection as well (and clearly $V + f(x_0)$ is open), so an inverse map $f^{-1} : V + f(x_0) \to U$ exists and is given by $f^{-1}(y) = \tilde{f}^{-1}(y - f(x_0))$. In particular, $(f^{-1})'(f(x_0)) = \left(\tilde{f}^{-1}\right)'(f(x_0) - f(x_0)) = \left(\tilde{f}^{-1}\right)'(0)$, so f^{-1} is indeed differentiable at $f(x_0)$.

Next, it's enough to show the theorem in the special case where $x_0 = 0$. The general case follows by applying the special case to the new function $\tilde{f}(x) = f(x + x_0)$: If f is continuously differentiable, then clearly so is $\tilde{f}$, and if $f'(x_0)$ is invertible, then $\tilde{f}'(0) = f'(0 + x_0)$ is invertible, so there exists open sets U containing 0 and V containing $\tilde{f}(0) = f(x_0) = 0$ for which $\tilde{f} : U \to V$ is a bijection and for which the inverse map is differentiable at $\tilde{f}(0) = 0$. Thus $f(x) = \tilde{f}(x - x_0) : U + x_0 \to V$ is a bijection as well (and clearly $U + x_0$ is open), so an inverse map $f^{-1} : V \to U + x_0$ exists and is given by $f^{-1}(y) = \tilde{f}^{-1}(y) + x_0$. In particular, $(f^{-1})'(0) = \left(\tilde{f}^{-1}\right)'(0)$, so f^{-1} is indeed differentiable at 0.

Finally, it's enough to show the theorem in the special case where $f'(0) = I$, where $I : \mathbb{R}^n \to \mathbb{R}^n$ is the identity map. The general case follows by applying the special case to the new function $\tilde{f}(x) = f'(0)^{-1} f(x)$: If f is continuously differentiable, then clearly so if $\tilde{f}$ (since $f'(0)^{-1}$ is a linear map, its continuously differentiable, so clearly the composition is as well), and clearly $\tilde{f}'(0) = \frac{d}{dx}(f'(0)^{-1} f(x))|_{x=0} = f'(0)^{-1} f'(0) = I$ is invertible, so there exists open sets U containing 0 and V containing 0 for which $\tilde{f} : U \to V$ is a bijection and for which the inverse map is differentiable at 0. Now consider $f(x) = f'(0)\tilde{f}(x) : U \to f'(0)(V)$. Note that f is a bijection, since $f'(0)$ is an invertible linear map, which means it's a bijection, and $\tilde{f}$ is a bijection. Note also that $f'(0)(V)$ is open, as since $f'(0)^{-1}$ is a linear map, it's continuous, so the inverse image of a set in its codomain will be open in its domain, and the inverse image will be given by $f'(0)$ (since again both maps are invertible and thus bijections). Finally note that $0 \in f'(0)V$, since $f'(0)$ is a linear map, and $0 \in V$. Thus an inverse map $f^{-1} : f'(0)(V) \to U$ exists and is given by $f^{-1}(y) = \tilde{f}^{-1}(f'(0)^{-1} y)$. In particular, $(f^{-1})'(0) = \left(\tilde{f}^{-1}\right)'(f'(0)^{-1} 0) f'(0)^{-1} = \left(\tilde{f}^{-1}\right)'(0) f'(0)^{-1}$ is indeed differentiable at 0.

Thus, we only need to prove the theorem in the case where $x_0 = 0$, $f(x_0) = 0$, and $f'(x_0) = I$. Let $g : E \to \mathbb{R}^n$ denote the function $g(x) = f(x) - x$. Then $g(0) = 0$ and $g'(0) = 0$. Thus $\frac{\partial g}{\partial x_j}(0) = 0$ for all $1 \leq j \leq n$. Since g is continuously differentiable, there exists a ball $B_r(0)$ in E such that

$$\left\| \frac{\partial g}{\partial x_j}(x) \right\| \leq \frac{1}{2n^2}$$

for $x \in B_r(0)$. Thus for all $x \in B_r(0)$ and $v = (v_1, ..., v_n)$, we have

$$\begin{aligned}\|D_v g(x)\| &= \left\| \sum_{j=1}^{n} v_j \frac{\partial g}{\partial x_j}(x) \right\| \\ &\leq \sum_{j=1}^{n} |v_j| \left\| \frac{\partial g}{\partial x_j}(x) \right\| \\ &\leq \sum_{j=1}^{n} \frac{\|v\|}{2n^2} = \frac{1}{2n}\|v\|.\end{aligned}$$

Now, for any $x, y \in B_r(0)$ and for any component g_j of f, we have by the fundamental theorem of calculus

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

From the chain rule, we have that the integrand is equal to $g_j'(x + t(y-x))(y-x) = D_{y-x} g_j(x + t(y-x))$. Note that this is a component of $D_{y-x} g(x + t(y-x))$, so we have

$$\left|D_{y-x} g_j(x + t(y-x))\right| \leq \left\|D_{y-x} g(x + t(y-x))\right\| \leq \frac{1}{2n}\|y-x\|,$$

since $x + t(y-x) \in B_r(0)$ for $t \in [0,1]$. Thus $\left|g_j(y) - g_j(x)\right| \leq \frac{1}{2n}\|y-x\|$ for all $1 \leq j \leq n$, which then implies

$$\|g(y) - g(x)\| \leq \sum_{j=1}^{n} \left|g_j(y) - g_j(x)\right| \leq \sum_{j=1}^{n} \frac{1}{2n}\|y-x\| = \frac{1}{2}\|y-x\|.$$

Thus g is a strict contraction with contraction constant $\frac{1}{2}$. Letting $y = 0$ in the contraction bound, we have

$$\|g(x)\| \leq \frac{1}{2}\|x\| \Rightarrow \|f(x) - x\| \leq \frac{1}{2}\|x\|.$$

Applying the reverse triangle inequality to the left side, unraveling the absolute value, and adding $\|x\|$ to both sides yields

$$\frac{1}{2}\|x\| \leq \|f(x)\| \leq \frac{3}{2}\|x\|.$$

Now we find U and V. Set $V = B_{r/2}(0)$ and $U = f^{-1}(V) \cap B_r(0)$, where f^{-1} denotes the invere image. Since f is continuous and V is open, clearly the inverse image is also open, so both U and V are open. Note the from the lemma that $f = g + I$ is injective on $B_r(0)$, so clearly it will be injective from $U \subseteq B_r(0)$ to V. From the lemma we also know that $B_{r/2}(0) \subseteq f(B_r(0))$, any $y \in V$ will be the image of some $x \in U$. Thus f is surjective as well, so f is a bijection. Thus, there is a well defined inverse $f^{-1} : V \to U$.

Now we just need to show that f^{-1} is differentiable at 0 with derivative $I^{-1} = I$. Thus we need to show that

$$\lim_{y \to 0} \frac{\left\|f^{-1}(y) - f^{-1}(0) - I(y - 0)\right\|}{\|y\|} = 0.$$

Simplifying, we need to show that

$$\lim_{y \to 0} \frac{\|f^{-1}(y) - y\|}{\|y\|} = 0.$$

Let $(y_n) \in V$ be a sequence that converges to 0. Thus we want to show

$$\lim_{n \to \infty} \frac{\|f^{-1}(y_n) - y_n\|}{\|y_n\|} = 0.$$

Now let $x_n = f^{-1}(y_n) \in U$. Note that from our earlier bound on $f(x)$, we have $\frac{1}{2}\|x_n\| \leq \|y_n\| \leq \frac{3}{2}\|x_n\|$. Thus (x_n) also converges to 0. Rewriting the function in the limit with x_n's, we need to show that

$$\lim_{n \to \infty} \frac{\|x_n - f(x_n)\|}{\|f(x_n)\|} = 0.$$

Note that again from the bound on $f(x)$, we have

$$\frac{2}{3} \cdot \frac{\|x_n - f(x_n)\|}{\|x_n\|} \leq \frac{\|x_n - f(x_n)\|}{\|f(x_n)\|} \leq 2 \cdot \frac{\|x_n - f(x_n)\|}{\|x_n\|}.$$

Thus, if we show the limit of the right side is 0, we're done. Since f is differentiable with derivative I, we have

$$\lim_{n \to \infty} \frac{\|f(x_n) - f(0) - I(x_n - 0)\|}{\|x_n\|} = 0.$$

Simplifying the inside yields $\frac{\|f(x_n) - x_n\|}{\|x_n\|}$, which is exactly the right side of the inequality minus the constant, so we're done. ■

13.4. Implicit Function Theorem

Theorem (implicit function theorem): Let E be an open subset of $\mathbb{R}^n$, let $f : E \to \mathbb{R}$ be continuously differentiable, and let $y = (y_1, ..., y_n)$ be a point in E such that $f(y) = 0$ and $\frac{\partial f}{\partial x_n}(y) \neq 0$. Then there exists an open subset U of $\mathbb{R}^{n-1}$ containing $(y_1, ..., y_{n-1})$, an open subset V of E containing y, and a function $g : U \to \mathbb{R}$ such that $g(y_1, ..., y_{n-1}) = y_n$, and

$$\begin{aligned} &\{(x_1, ..., x_n) \in V : f(x_1, ..., x_n) = 0\} \\ &= \{(x_1, ..., x_{n-1}, g(x_1, ..., x_{n-1})) : (x_1, ..., x_{n-1}) \in U\}. \end{aligned}$$

Moreover, g is differentiable at $(y_1, ..., y_{n-1})$, and we have

$$\frac{\partial g}{\partial x_j}(y_1, ..., y_{n-1}) = -\frac{\partial f}{\partial x_j}(y) / \frac{\partial f}{\partial x_n}(y)$$

for all $1 \leq j \leq n - 1$.

Proof: Let $F : E \to \mathbb{R}^n$ be the function

$$F(x_1, ..., x_n) = (x_1, ..., x_{n-1}, f(x_1, ..., x_n)).$$

Since f is continuously differentiable, this one is also continuously differentiable. We have $F(y) = (y_1, ..., y_{n-1}, 0)$ and

$$DF(y) = \begin{pmatrix} 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \\ \frac{\partial f}{\partial x_1}(y) & \frac{\partial f}{\partial x_2}(y) & \cdots & \frac{\partial f}{\partial x_{n-1}}(y) & \frac{\partial f}{\partial x_n}(y) \end{pmatrix}.$$

Since the matrix is triangular, the determinant is the product of the diagonal entries, which is equal to $\frac{\partial f}{\partial x_n}(y) \neq 0$. Thus $F'(y)$ is invertible, so the inverse function theorem applies. Thus there exists an open set $V \subseteq E$ that contains y and open set $W \subseteq \mathbb{R}^n$ containing $F(y)$ such that $F : V \to W$ is a bijection and such that F^{-1} is differentiable at $F(y) = (y_1, ..., y_{n-1}, 0)$.

Writing F^{-1} in coordinates as $F^{-1}(x) = (h_1(x), ..., h_n(x))$, where $x \in W$. Since $F(F^{-1}(x)) = x$, we have $h_j(x) = x_j$ for all $1 \leq j \leq n-1$ and $x \in W$, and then we have

$$f(x_1, ..., x_{n-1}, h_n(x_1, ..., x_n)) = x_n.$$

Since F^{-1} is differentiable at $(y_1, ..., y_{n-1}, 0)$, we see that h_n is also differentiable there.

Set $U = \{(x_1, ..., x_{n-1}) \in \mathbb{R}^{n-1} : (x_1, ..., x_{n-1}, 0) \in W\}$. Note that U is open and contains $(y_1, ..., y_{n-1})$. Now define $g : U \to \mathbb{R}$ as $g(x_1, ..., x_{n-1}) = h_n(x_1, ..., x_{n-1}, 0)$. Then g is differentiable at $(y_1, ..., y_{n-1})$ since h is differentiable at $(y_1, ..., y_{n-1}, 0)$.

Now we prove the equality of the two sets. Suppose x is in the first then. Then $x = (x_1, ..., x_n) \in V$ and $f(x_1, ..., x_n) = 0$. Then $F(x) = (x_1, ..., x_{n-1}, 0)$. Since the output of F is in W, we have $(x_1, ..., x_{n-1}, 0) \in U$. Applying F^{-1} to both sides yields $(x_1, ..., x_n) = F^{-1}(x_1, ..., x_{n-1}, 0)$. This implies that $x_n = h_n(x_1, ..., x_{n-1}, 0)$, and thus by definition $x_n = g(x_1, ..., x_{n-1})$. Thus x lies in the second set, so the first set is a subset of the second set

Now suppose x is in the second set. Thus we can write it as $(x_1, ..., x_{n-1}, g(x_1, ..., x_{n-1}))$ for $(x_1, ..., x_{n-1}) \in U$. Letting $x_n = g(x_1, ..., x_{n-1})$, we have by definition that $x_n = h_n(x_1, ..., x_{n-1}, 0)$. Thus we have $F^{-1}(x_1, ..., x_{n-1}, 0) = (x_1, ..., x_n)$. Since the output of F^{-1} is in V, we have $(x_1, ..., x_n) \in V$. Applying F to both sides yields $(x_1, ..., x_{n-1}, 0) = F(x_1, ..., x_n)$. Thus from the definition of F, we have that $f(x_1, ..., x_n) = 0$. Thus x lies in the second set, so the second set is a subset of the first set. Since we have inclusions in both directions, the sets must be the same.

Thus, we have

$$f(x_1, ..., x_{n-1}, g(x_1, ..., x_{n-1})) = 0$$

for all $(x_1, ..., x_{n-1}) \in U$. Since g is differentiable at $(y_1, ..., y_{n-1})$ and f is differentiable at $(y_1, ..., y_{n-1}, g(y_1, ..., y_{n-1})) = y$, we can differentiate with respect to x_j, and the chain rule yields

$$\frac{\partial f}{\partial x_j}(y) + \frac{\partial f}{\partial x_n}(y) \frac{\partial g}{\partial x_j}(y_1, ..., y_{n-1}) = 0,$$

and rearranging yields the desired conclusion. ■

13.5. Problems

Problem: Let $f : \mathbb{R}^n \to \mathbb{R}^n$ be a continuously differentiable function such that $f'(x)$ is invertible for all $x \in \mathbb{R}^n$. Show that if $W \subseteq \mathbb{R}^n$ is open, then $f(W)$ is also open.

Solution: Consider $f(x_0) \in f(W)$ for some $x_0 \in \mathbb{R}^n$. We need to show that some neighborhood of $f(x_0)$ is contained in $f(W)$. By the inverse function theorem, there exist open sets U and V which contain x_0 and $f(x_0)$ respectively such that $f : U \to V$ is a bijection. Note that $U \cap W$ is open, since both sets are open. Since f is a continuous bijection on U, the set $f(U \cap W)$ must be open in as well (since f^{-1} is taking the role of f in the result about inverse images of open sets). Thus some neighborhood of $f(x_0)$ is contained in $f(U \cap W)$, but we also have that $f(U \cap W) \subseteq f(W)$. Thus some neighborhood of $f(x_0)$ is contained in $f(W)$, so $f(W)$ is open, as desired.

14. Multivariable Integration

Most of the stuff here until change of variables content is a weaker version of the Lebesgue integral with similar properties, so a good beginning chunk of this section will have results that aren't fully general, simply because the Lebesgue integral has even more general results than the most general with Riemann integrals. The concept of rectifiable sets (more commonly known as Jordan content) is also weaker than being measurable.

14.1. Integrals Over Rectangles

The theory here is almost identical to single variable integration.

Definition (volume of rectange): The *volume of a rectangle* $Q = [a_1, b_1] \times \cdots \times [a_n, b_n]$ is $(b_1 - a_1) \cdots (b_n - a_n)$ and is denoted $v(Q)$.

Definition (partition, subrectangle, mesh): Suppose $Q = [a_1, b_1] \times \cdots \times [a_n, b_n]$ is a closed rectange in $\mathbb{R}^n$. A *partition* P of Q is the tuple $(P_1, P_2, ..., P_n)$, where P_i is a partition of $[a_i, b_i]$. Taking a subinterval I_i from P_i for each I, the set $I_1, \times \cdots \times I_n$ is a *subrectangle* determined by P of Q. The maximum side length of one of the subrectangles is the *mesh* of P.

Definition (lower/upper sum): Let Q be a rectangle in $\mathbb{R}^n$. Let $f : Q \to \mathbb{R}$ and suppose f is bounded. Let P be a partition of Q. For each subrectangle R of P, let

$$m_R(f) = \inf\{f(x) : x \in \mathbb{R}\},$$
$$M_R(F) = \sup\{f(x) : x \in \mathbb{R}\}.$$

Then the *upper sum* and *lower sum* of f with respect to P are

$$L(f, P) = \sum_R m_R(f) v(R),$$
$$U(f, P) = \sum_R M_R(F) v(R),$$

where R runs over all subrectangles induced by P.

Definition (refinement): Let Q be a rectangle in $\mathbb{R}^n$. Suppose P, P' are partitions of Q such that $P_i \subseteq P_i'$ for all $1 \leq i \leq n$. Then P' is a *refinement* of P.

Proposition (refinment increase/decrease lower/upper sum): Let P be a partition of Q, and let $f : Q \to \mathbb{R}$ be a bounded function. If P' is a refinement of P, then

$$L(f, P) \leq L(f, P') \quad \text{and} \quad U(f, P') \leq U(f, P).$$

Proof: We prove the upper sum inequality, as the lower sum case follows similarly. It also suffices to show the claim for a refinement that only has one extra point in the partition, as then we can just induct on the number of new points.

Suppose $P = (P_1, P_2, ..., P_n)$ and $P' = (P_1 \cup \{q\}, P_2, ...P_n)$, where $a_1 \leq q \leq b_1$. Also suppose that q lies in the subinterval of P_1 given by $[t_{i-1}, t_i]$. All the subrectangles stay the same except for the ones that contain $[t_{i-1}, t_i]$ as one of their sides. Let S be a subrectangle of $[a_2, b_2] \times \cdots \times [a_n, b_n]$ given by $(P_2, ..., P_n)$. Then the new subrectangles that replace $[t_{i-1}, t_i] \times S$ are $[t_{i-1}, q] \times S$ and $[q, t_i] \times S$. Let R be the originaal subrectangle, and let R_1, R_2 denote the ones it is replaced by. Then

$$M_{R_1}(f), M_{R_2}(f) \leq M_R(f) \Rightarrow M_{R_1}(f)v(R_1) + M_{R_2}(f)v(R_2) \leq M_R(f)(v(R_1) + v(R_2)) = M_R(f)v(R).$$

This holds for any S, so overall $U(f, P') \leq U(f, P)$, as desired. ■

Proposition: Suppose Q is a rectangle and $f : Q \to \mathbb{R}$ is bounded. For any partitions P_1, P_2 of Q, we have

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

Proof: Consider P', where $P_i' = P_{1i} \cup P_{2i}$. Thus P' is refinement of P_1 and P_2, so we have

$$L(f, P_1) \leq L(f, P') \leq U(f, P') \leq U(f, P_2),$$

where the middle inequality easily follows from $m_R(f) \leq M_R(f)$. ■

Definition ((upper/lower) integral): Let Q be a rectangle and let $f : Q \to \mathbb{R}$ be a bounded function. As P ranges over all partitions of Q, define

$$\overline{\int_Q} f = \sup_P L(f, P) \quad \text{and} \quad \underline{\int_Q} f = \inf_P U(f, P)$$

as the upper and lower sum respectively (which may also be denoted as $U(f)$ and $L(f)$). If the two are equal, then f is *integrable* over Q, and we define

$$\int_Q f = U(f) = L(f).$$

The integral can also be denoted as

$$\int_{x \in Q} f(x).$$

Proposition (condition for integrability): Let Q be a rectangle, and suppose $f : Q \to \mathbb{R}$ is bounded. Then

$$\underline{\int_Q} f \leq \overline{\int_Q} f,$$

and equality holds if and only if for every $\varepsilon > 0$, there exists a partition P_ε of Q such that

$$U(f, P_\varepsilon) - L(f, P_\varepsilon) < \varepsilon.$$

Proof: The first inequality follows from previous proposition and taking sup/inf of either side. Now suppose we have equality. We know there exists P_1, P_2 such that $L(f) - L(f, P_1) < \frac{\varepsilon}{2}$ and $U(f, P_2) - U(f) < \frac{\varepsilon}{2}$. Let P_ε be their common refinement. Thus we can replace P_1, P_2 with P_ε. Then adding the two inequalities yields $U(f, P_\varepsilon) - L(f, P_\varepsilon) < \varepsilon$, as desired.

Now suppose that $L(f) < U(f)$. Then clearly there can't exist a partition P such that $U(f, P) - L(f, P) < U(f) - L(f)$, since we have $L(f, P) \leq L(f) < U(f) \leq U(f, P)$. ■

Proposition (continuous implies integrable): Suppose $f : Q \to \mathbb{R}$ is continuous on a closed rectangle $Q \subseteq \mathbb{R}^n$. Then f is integrable over Q.

Proof: Since Q is compact, f is uniformly continuous on Q. Pick $\varepsilon > 0$. By uniform continuity, there exists $\delta > 0$ such that $|x - y| < \delta \Rightarrow |f(x) - f(y)| < \frac{\varepsilon}{v(Q)}$ for $x, y \in Q$. Now partition each side of Q into intervals of length $\frac{\delta}{\sqrt{2n}}$, and let this partition be P_ε. Then clearly in any subrectangle of Q induced by P, the maximal distance between any two points is $\frac{\delta}{\sqrt{2}}$, so from uniform continuity, the function changes at most ε on the subrectangle. Thus on every subrectangle R, we have

$$M_R(f) - m_R(f) = \max_R f(x) - \min_R f(x) < \frac{\varepsilon}{v(Q)},$$

where continuity yielded the equality. Thus we have

$$U(f, P_\varepsilon) - L(f, P_\varepsilon) = \sum_R (M_R(f) - m_R(f))v(R) < \sum_R \frac{\varepsilon}{v(Q)} \cdot v(R) = \varepsilon,$$

so by the previous proposition, $\int_Q f$ exists. ■

Proposition (bound on integral): Let Q be a rectangle and let $f : Q \to \mathbb{R}$ be bounded. Then

$$m_Q(f)v(Q) \leq \underline{\int_Q} f \leq \overline{\int_Q} \leq M_Q(f)v(Q).$$

Proof: Let P be partition of Q. Then

$$U(f, P) = \sum_R M_R(f)v(R) \leq \sum_R M_Q(f)v(R) = M_Q(f)v(Q),$$

where the fact that $M_R(f) \leq M_Q(f)$ follows from the fact that the set of values attained in R is a subset of the set of values attained in Q. ∎

Proposition (integrability criterion from mesh): Let $f : Q \to \mathbb{R}$ be bounded, where $Q \subseteq \mathbb{R}^n$. Then f is integrable over Q if and only if given $\varepsilon > 0$, there is a $\delta > 0$ such that $U(f, P) - L(f, P) < \varepsilon$ for every partition P of mesh less than δ.

Proof: If the condition holds, then integrability follows easily from $U(f, P_\varepsilon) - L(f, P_\varepsilon) < \varepsilon$.

Now suppose the condition holds. Let $|f(x)| \leq M$ on Q, and suppose P is a partition of Q. Suppose we add a point into one of the component intervals of Q, creating a new partition P''. Then we can show (somewhat tediously) that

$$0 \leq L(f, P'') - L(f, P) \leq 2M(\text{mesh } P)(\text{wdith } Q)^{n-1}$$

and

$$0 \leq U(f, P) - U(f, P'') \leq 2M(\text{mesh } P)(\text{wdith } Q)^{n-1}.$$

From integrability, there exists P_ε such that $U(f, P_\varepsilon) - L(f, P_\varepsilon) < \frac{\varepsilon}{2}$. Suppose P_ε has N points in the partition. Let

$$\delta = \frac{\varepsilon}{8MN(\text{width } Q)^{n-1}},$$

and suppose some partition P of Q has mesh less than δ. Using the above inequalities N times and adding in the N points from P_ε into P (where we denote the ending common refinement P_N) yields

$$0 \leq L(f, P_N) - L(f, P) \leq 2NM(\text{mesh } P)(\text{width } Q)^{n-1}$$

and

$$0 \leq U(f, P) - U(f, P_N) \leq 2NM(\text{mesh } P)(\text{width } Q)^{n-1}.$$

Summing the two inequalities yields

$$0 \leq U(f, P) - L(f, P) - (U(f, P_N) - L(f, P_N)) \leq 4NM(\text{mesh } P)(\text{width } Q)^{n-1}.$$

Since mesh $P < \delta$, we obtain

$$U(f, P) - L(f, P) - (U(f, P_N) - L(f, P_N)) < \frac{\varepsilon}{2}.$$

Since P_N is a refinment of P_ε, we have $U(f, P_N) - L(f, P_N) < \frac{\varepsilon}{2}$, so bringing that over the other side yields the desired conclusion. ∎

14.1.1. Lebesgue's Integrability Criterion

Oscillations all the way. This is just a rehash of the one dimensional verson, so I'll just state the result here and move on.

Theorem (Lebesgue's integrability criterion): Let Q be a rectangle in $\mathbb{R}^n$, and let $f : Q \to \mathbb{R}$ be bounded. Let D be the set of points of Q at which f is not continuous. Then $\int_Q f$ exists if and only if D has measure zero in $\mathbb{R}^n$.

14.1.2. Fubini's Theorem

Again!

Theorem (Fubini's theorem): Let $Q = A \times B$, where A is a rectangle in $\mathbb{R}^k$ and B is a rectangle in $\mathbb{R}^n$. Let $f : Q \to \mathbb{R}$ be a bounded function and write in the form $f(x, y)$ for $x \in A$ and $y \in B$. For each $x \in A$, consider the lower and upper integrals

$$\underline{\int_B} f(x, y) \quad \text{and} \quad \overline{\int_B} f(x, y).$$

If f is integrable over Q, then these two functions of x are integrable over A, and

$$\int_Q f = \int_A \underline{\int_B} f(x, y) = \int_A \overline{\int_B} f(x, y).$$

Proof: Define

$$\underline{I}(x) = \underline{\int_B} f(x, y) \quad \text{and} \quad \overline{I}(x) = \overline{\int_B} f(x, y)$$

for $x \in A$. Assuming $\int_Q f$ exists, we show that $\overline{I}$ and $\underline{I}$ are integrable over A.

Let P be partition of Q. Then P consists of a partition P_A of A, and a partition P_B of B. If R_A is a subrectangle of A induced by P_A, and similarly for R_B, then $R_A \times R_B$ is a subrectangle of Q induced by P.

We show that $L(f, P) \leq L(\underline{I}, P_A)$. Pick some rectangle R in Q induced by P. Then from above, we can write $R = R_A \times R_B$. Fix some $x_0 \in R_A$. Then

$$m_R(f) \leq f(x_0, y)$$

for all $y \in R_B$, so taking the infinum over all $y \in R_B$ yields

$$m_R(f) \leq m_{R_B}(f(x_0, y)).$$

Now multiply both sides by $v(R_B)$ and sum over all R_B in B. We then have

$$\sum_{R_B} m_R(f) v(R_B) \leq \sum_{R_B} m_{R_B}(f(x_0, y)) v(R_B) = L(f(x_0, y), R_B) \leq \underline{I}(x_0).$$

This holds for all $x_0 \in R_A$, so $\sum_{R_B} m_R(f) v(R_B) \leq m_{R_A}(\underline{I})$. Then multiplying by $v(R_A)$ and summing over all of them yields

$$L(f,P) = \sum_R m_R(f)v(R) = \sum_{R_A}\sum_{R_B} m_R(f)v(R_A)v(R_B) \le \sum_{R_A} m_{R_A}(\underline{I})v(R_A) = L(\underline{I}, P_A).$$

The same method shows that $U(f,P) \ge U\left(\overline{I}, P_A\right)$.

Combinding everything we have

$$L(f,P) \le L(\underline{I}, P_A) \le L\left(\overline{I}, P_A\right), U(\underline{I}, P_A) \le U\left(\overline{I}, P_A\right) \le U(f,P).$$

Since f is integrable, there exists a partition P_ε for which the extreme bounds are ε apart. That implies that $L(\underline{I}, P_A)$ and $U(\underline{I}, P_A)$ are within ε of each other, and similarly for $\overline{I}$. Thus they are integrable over A. Note that since P is arbitrary, we must have that the two extreme ends must be equal, so combining everything, we obtain

$$\int_A \underline{I} = \int_A \overline{I} = \int_Q f.$$

■

Corollary: Let $Q = A \times B$, where A is a rectangle in $\mathbb{R}^k$ and B is a rectangle in $\mathbb{R}^n$. Let $f : Q \to \mathbb{R}$ be a bounded function. If $\int_Q f$ exists, and if $\int_{y \in B} f(x,y)$ exists for each $x \in A$, then

$$\int_Q f = \int_{x \in A} \int_{y \in B} f(x,y).$$

Corollary: Let $Q = I_1 \times \dots \times I_n$, where I_j is a closed interval in $\mathbb{R}$. If $f : Q \to \mathbb{R}$ is continuous, then

$$\int_Q f = \int_{x_1 \in I_1} \cdots \int_{x_n \in I_n} f(x_1, \dots, x_n).$$

14.2. Integrals Over Bounded Sets

Definition (integral over bounded set): Let S be a bounded set in $\mathbb{R}^n$ and suppose $f : S \to \mathbb{R}$ is bounded. Define $f_S : \mathbb{R}^n \to \mathbb{R}$ as $f\chi_S$. Choose a rectangle Q that contains S. Then the integral of f over S is

$$\int_S f = \int_Q f_S.$$

Proposition: Let Q and Q' be two rectangles in $\mathbb{R}^n$. If $f : \mathbb{R}^n \to \mathbb{R}$ is a bounded function that vanishes outside $Q \cap Q'$, then

$$\int_Q f = \int_{Q'} f,$$

where one integral exists if and only if the other one does.

Proof: Consider the case where $Q \subset Q'$, and let $E \subseteq \operatorname{int}(Q)$ be the set of points at which f fails to be continuous. Note that f can also be discontinuous on ∂Q, but this has measure zero in $\mathbb{R}^n$, so we can ignore it. Thus the first integral exists if and only if E has measure zero. Note that the set of discontinuities for f on Q is the same as for f on Q' plus some potential extras on the boundary of Q, but again this has measure zero so we can ignore it. Thus the second integral exists if and only if the first one does.

Now suppose both integrals exist. Let P be a partition of Q', and let P' be the refinement of P obtained by adding the component endpoints of Q into P. Then Q is made up of disjoint subrectangles, and on any subrectangle not inside Q, f is zero by definition. Thus

$$L(f, P') \leq \sum_{R \subseteq Q} m_R(f) v(R) \leq \int_Q f.$$

A similar argument shows that $\int_Q f \leq U(f, P')$. Since P was arbitrary, we can make the upper and lower bound arbitrarily close, and thus $\int_{Q'} f = \int_Q f$.

For the case where Q and Q' don't contain one another, enclose them in a larger rectange Q'' and apply the above. ■

Now here's a list of a bunch of properties that the integral has. The proofs are basically the same as the one dimensional case.

Proposition (properties of integral): Let S be a bounded set in $\mathbb{R}^n$, and let $f, g : S \to \mathbb{R}$ be bounded.

a) If f and g are integrable over S, so is $af + bg$, and

$$\int_S af + bg = a \int_S f + b \int_S g.$$

b) Suppose f and g are integrable over S. If $f(x) \leq g(x)$ for all $x \in S$, then

$$\int_S f \leq \int_S g.$$

c) If f is integrable, then $|f|$ is integrable, and

$$\left| \int_S f \right| \leq \int_S |f|.$$

d) Let $T \subset S$. If f is nonnegative on S and integrable over T and S, then

$$\int_T f \leq \int_S f.$$

e) If $S = S_1 \cup S_2$, and f is integrable over S_1 and S_2, then f is integrable over S and $S_1 \cap S_2$, and we have

$$\int_S f = \int_{S_1} f + \int_{S_2} f - \int_{S_1 \cap S_2} f.$$

Corollary: Let $S_1, S_2, ..., S_k$ be bounded sets in $\mathbb{R}^n$, and suppose $S_i \cap S_j$ has measure zero whenever $i \neq j$. Let $S = S_1 \cup \dots \cup S_k$. If $f : S \to \mathbb{R}$ is integrable over each S_i, then f is integrable over S around

$$\int_S f = \int_{S_1} f + \dots + \int_{S_k} f.$$

Proof: This follows from e) of the previous result, since integrals over sets of measure zero are zero. ■

Proposition: Let S be a bounded set in $\mathbb{R}^n$, and let $f : S \to \mathbb{R}$ be bounded and continuous. Let E be the set of points of ∂S at which the condition

$$\lim_{x \to x_0} f(x) = 0$$

fails to hold. If E has measure zero, then f is integrable over S.

Proof: Slap a rectangle Q around S, note that $f\chi_S$ on Q is only discontinuous at points in E, and since it has measure zero, $\int_Q f$ exists, and thus $\int_S f$ exists. ■

Proposition: Let S be a bounded set in $\mathbb{R}^n$, and suppose $f : S \to \mathbb{R}$ is bounded and continuous. Let $A = \operatorname{int}(S)$. If f is integrable over S, then f is integrable over A, and $\int_S f = \int_A f$.

Proof: f on S is continuous at basically every point that f on A is, so if f on S is integrable, then f on A is as well. ■

14.3. Rectifiable Sets

Definition (rectifiable set): Let S be a bounded set in $\mathbb{R}^n$. If the constant function 1 is integrable over S, we say that S is *rectifiable*, and we define the n-dimensional volume of S as

$$v(S) = \int_S 1.$$

Proposition: A subset $S \subseteq \mathbb{R}^n$ is rectifiable if and only if S is bounded and ∂S has measure zero.

Proof: χ_S is clearly continuous on $\operatorname{int}(S)$ and $\operatorname{ext}(S)$ and discontinuous on ∂S. Thus χ_S is integrable over S as long as we can put a rectangle around S and as long as ∂S has measure zero. ■

Proposition (properties of rectifiable sets):

a) If S is rectifiable, $v(S) \geq 0$.
b) If $S_1 \subseteq S_2$ are both rectifiable, then $v(S_1) \leq v(S_2)$.
c) If S_1 and S_2 are rectifiable, so are $S_1 \cup S_2$ and $S_1 \cap S_2$, and

$$v(S_1 \cup S_2) = v(S_1) + v(S_2) - v(S_1 \cap S_2).$$

d) Suppose S is rectifiable. Then $v(S) = 0$ if and only if S has measure zero.
e) If S is rectifiable, so is $\operatorname{int}\ (S)$, and they have equal volume.
f) If S is rectifiable, and if $f : S \to \mathbb{R}$ is a bounded continuous function, then f is integrable over S.

Proof: Follows from previous integral properties. ■

Definition (simple region): Let C be a compact rectifiable set in $\mathbb{R}^{n-1}$, and let $\phi, \psi : C \to \mathbb{R}$ be continuous functions such that $\phi(x) \leq \psi(x)$ for $x \in C$. The subset S of $\mathbb{R}^n$ defined by the equation

$$S = \{(x,t) : X \in C \ \text{ and } \ \phi(x) \leq t \leq \psi(x)\}$$

is called a *simple region* in $\mathbb{R}^n$.

Proposition: If S is a simple region in $\mathbb{R}^n$, then S is compact and rectifiable.

Proof: First we show S is compact. For a compact rectifiable C and continuous function f on C, define the graph G_f to be

$$G_f := \{(x, f(x)) : x \in C\}.$$

Further define

$$D := \{(x,t) : x \in \partial C \ \text{ and } \ \phi(x) \leq t \leq \psi(x)\}.$$

We claim $G_\phi \cup D \cup G_\psi$ combine to make up ∂S, and since clearly these sets are contained in S, it must be closed. S is also clearly bounded, since C is compact and both $\phi . \psi$ are continuous, which implies S is compact.

Suppose (x_0, t_0) is in none of the three sets above. Then either $x_0 \notin C$, $X_0 \in C$ with $t_0 < \phi(x_0)$ or $\psi(x_0) < t_0$, or $x_0 \in \operatorname{int}(C)$ with $\phi(x_0) < t_0 < \psi(x_0)$. In each case, we can easily construct a neighborhood around (x_0, t_0) such that the neighborhood stays within $\operatorname{int}(S)$ or $\operatorname{ext}(S)$. Thus, every other point in $\mathbb{R}^n$, which is contained in the above sets, it a boundary point, as desired.

Now we show each of the above sets has measure zero, which implies that ∂S has measure zero, and thus S is rectifiable.

First consider D. Note that since C is rectifiable in $\mathbb{R}^{n-1}$, there exists some rectangular cover of ∂C that has $n-1$ dimensional volume less than ε. Suppose $\min_{x \in \partial C} \phi(x) = m$ and $\max_{x \in \partial C} \psi(x) = M$. Then for each rectangle in the aforementioned cover, tack on $[m, M]$. It's clear that this now covers D, and has n dimensional volume $\varepsilon(M - m)$, which clearly can get arbitrarily small. Thus D has measure zero.

Now we show that G_ϕ and G_ψ have measure zero. We only show the second has measure zero, since the first follows similarly. Choose some rectangle Q in $\mathbb{R}^{n-1}$ that contains C, and pick $\varepsilon > 0$. By uniform continuity (since C is compact), there exists δ such that $|x - y| < \delta \Rightarrow |\psi(x) - \psi(y)|$. Now pick a partition P of Q with mess less than δ. For every subrectangle R of P that intersects C, we have that $|\psi(x) - \psi(y)| < \varepsilon$ for $x, y \in R \cap C$. For each such R, choose a point $x_R \in R \cap C$, and define $I_R := [\psi(x_R) - \varepsilon, \psi(x_R) + \varepsilon]$. Then $R \times I_R$ covers $\psi(R \cap C)$. Thus, unioning over all R, we obtain a set of n dimensional rectangles that cover G_ψ. Since

$$\sum_R v(R \times I_R) = \sum_R v(R) \cdot 2\varepsilon \leq 2\varepsilon v(Q),$$

which can be made arbitrarily small, G_ψ has measure zero, as desired. ■

Corollary (Fubini's theorem for simple regions): Let S be a simple region defined as above. Let $f : S \to \mathbb{R}$ be a continuous function. Then f is integrable over S, and

$$\int_S f = \int_{x \in C} \int_{t=\phi(x)}^{t=\psi(x)} f(x,t).$$

Proof: Let $Q \times [-M, M]$ be some rectangle containing S. Since f is continuous and bounded on S and S is rectifiable, f is integrable over S. Then Fubini's theorem applies, and we have

$$\int_S f = \int_Q f\chi_S = \int_{x \in Q} \int_{t=-M}^{t=M} f(x,t)\chi_S.$$

Since $f\chi_S$ is zero outside of S, and since $f(x,t)\chi_S$ is zero unless $\phi(x) \leq t \leq \psi(x)$, we have

$$\int_S = \int_{x \in C} \int_{t=\phi(x)}^{t=\psi(x)} f(x,t).$$

■

14.4. Improper Integrals

This section extends the integral to potentially unbounded sets and functions.

Definition (extended integral): Let A be an open set in $\mathbb{R}^n$, and let $f : A \to \mathbb{R}$ be a continuous function. If f is nonnegative on A, we define the *(extended) integral* of f over A to be the supremum of $\int_D f$ over all compact rectifiable subsets of A, provided the supremum is finite. Then f is *inetgrable* over A. More generally, if f is an arbitrary continuous function on A and both f^+ and f^- are integrable over A (where these are defined in the same way as for Lebesgue integration), then we set

$$\int_A f = \int_A f^+ - \int_A f^-.$$

Proposition: Let A be an open set in $\mathbb{R}^n$. Then there exists a sequence $C_1, C_2, \ldots$ of compact rectifiable subsets of A whose union is A, such that $C_N \subseteq \operatorname{int}(C_{N+1})$ for each N.

Proof: Let $d(x, B)$ denote the smalled distance between a point x and set B in $\mathbb{R}^n$, which exists when B is closed. Let $B = \mathbb{R}^n \setminus A$, which is closed and define

$$D_N := \left\{ x : d(x, B) \geq \frac{1}{N} \quad \text{and} \quad \|x\| \leq N \right\}.$$

Since $d(x, B)$ is continuous in x, D_N is closed, and obviously it's bounded, so D_N is compact. It's also easy to see that D_N is contained in A, since the first condition implies that no point of it is in B. It's also easy to check that $D_N \subseteq \text{int}(D_{N+1})$.

The D_N may not be rectifiable, so we rectify this (pun fully intended). For each $x \in D_N$, choose a closed cube that's centered at x and is contained in $\text{int}(D_{N+1})$. The interiors cover D_N, so by compactness, we have a finite subcover. Then this subcover be C_N. Since it's a finite union of rectangles, we have a compact rectifiable set. Further,

$$D_N \subseteq \text{int}(C_{N+1}) \subseteq C_N \subseteq \text{int}(D_{N+1}),$$

so we're done. ∎

Proposition (alternate formulation of extended integral): Let A be open in $\mathbb{R}^n$, and let $f : A \to \mathbb{R}$ be continuous. Choose a sequence C_N of compact recitifiable subsets of A whose union is A such that $C_N \subseteq \text{int}(C_{N+1})$ for each N. Then f is integrable over A if and only if the sequence $\int_{C_N} |f|$ is bounded. In this case,

$$\int_A f = \lim_{N \to \infty} \int_{C_N} f.$$

Proof: We first show the theorem for nonnegative f. Since $\int_{C_N} f$ is increasing by monotonicity, it converges if and only if it's bounded.

First suppose f is integrable over A. Then we have

$$\int_{C_N} f \leq \sup_D \int_D f = \int_A f \Rightarrow \lim_{N \to \infty} \int_{C_N} f \leq \int_A f,$$

where the supremum ranges over all compact rectifiable subsets of A, so the sequence is bounded. Now suppose the sequence is bounded. Then every compact rectifiable D will be contained in some C_N, since the union of the C_N's is A and they are nested. Thus $\left\{\int_D f\right\}$ over all such D is bounded, so the supremum of the set exists, and thus f is integrable over A. In particular, we have

$$\int_D f \leq \int_{C_N} f \leq \lim_{N \to \infty} \int_{C_N} f \Rightarrow \int_A f \leq \lim_{N \to \infty} \int_{C_N} f.$$

Thus we have the desired equality.

Now suppose $f : A \to \mathbb{R}$ is an arbitrary continuous function. By definition, f is integrable over A if and only if f^+ and f^- are integrable over A. This occurs when the sequences $\int_{C_N} f^+$ and $\int_{C_N} f^-$ are bounded by the above. Further note that $|f| = f^+ + f^-$ and $0 \leq f_\pm \leq |f|$, so the two sequences are bounded if and only if $\int_{C_N} |f|$ is bounded. In this case, the two sequences converge to $\int_A f^+$ and $\int_A f^-$. Thus we have

$$\int_{C_N} f = \int_{C_N} f^+ - \int_{C_N} f^-,$$

which converges to $\int_A f^+ - \int_A f^- = \int_A f$. ∎

Proposition (properties of extended integral): Let A be an open set in $\mathbb{R}^n$, and let $f, g : A \to \mathbb{R}$ be continuous functions.

a) If f and g are integrable over A, so is $af + bg$, and
$$\int_A af + bg = a \int_A f + b \int_A g.$$

b) Let f and g be integrable over A. If $f \leq g$, then
$$\int_A f \leq \int_A g.$$

c)
$$\left| \int_A f \right| \leq \int_A |f|.$$

d) Suppose $B \subseteq A$ is open. If f is nonnegative on A and integrable over A, then f is integrable over B and
$$\int_B f \leq \int_A f.$$

e) Suppose A and B are open in $\mathbb{R}^n$ and f is continuous on $A \cap B$. If f is integrable on A and B, then f is integrable on $A \cup B$ and $A \cap B$, and
$$\int_{A \cup B} f = \int_A f + \int_B f - \int_{A \cap B} f.$$

Proof: Same old same old. ∎

Proposition: Let A be a bounded open set in $\mathbb{R}^n$ and suppose $f : A \to \mathbb{R}$ is a bounded continuous function. Then the extended integral $\int_A f$ exists. If the ordinary integral also exists, then these two integrals are equal.

Proof: Let Q be a rectangle containing A. For a compact rectifiable subset D of A, we have
$$\int_D |f| \leq \int_D M \leq M \cdot v(Q),$$
where $M = \max_A |f|$. Thus f is integrable in the extended sense.

Now suppose the ordinary integral of f over A exists. By definition, it's equal to the integral of $f\chi_A$ over Q. Then if D is a compact rectifiable subset of A, we have
$$\int_D = \int_D f\chi_A \leq \int_Q f\chi_A = (\text{ordinary}) \int_A f.$$

Since D is arbitrary, we have
$$(\text{extended}) \int_A f \leq (\text{ordinary}) \int_A f.$$

On the other hand, let P be a partition of Q. Then union all subrectangles R that intersect A and let this be D, which is clearly compact rectifiable. Then we have

$$L(f\chi_A, P) \leq \int_D f \leq (\text{extended}) \int_A f.$$

Since P is arbitrary, we have

$$(\text{ordinary}) \int_A f \leq (\text{extended}) \int_A f.$$

Now suppose f is an arbitrary continuous function. Then we have

$$\begin{aligned}(\text{ordinary}) \int_A f &= (\text{ordinary}) \int_A f^+ - (\text{ordinary}) \int_A f^- \\ &= (\text{extended}) \int_A f^+ - (\text{extended}) \int_A f^- \\ &= (\text{extended}) \int_A f.\end{aligned}$$

■

Here's a third formulation of the extended integral that is easier to use to acutally compute it.

Proposition: Let A be open in $\mathbb{R}^n$, and suppose $f : A \to \mathbb{R}$ is continuous. Let $U_1 \subseteq U_2 \subseteq \cdots$ be a sequence of open sets whose union is A. Then $\int_A f$ exists if and only if the sequence $\int_{U_N} |f|$ exists and is bounded. In this case,

$$\int_A f = \lim_{N\to\infty} \int_{U_N} f.$$

Proof: First we show this for nonnegative f. If $\int_A f$ exists, then we have

$$\int_{U_N} f \leq \int_A f,$$

and thus the sequence on the left converges with

$$\lim_{N\to\infty} \int_{U_N} f \leq \int_A f.$$

Now suppose the sequence is bounded and thus converges. Given any compact rectifiable D, the U_N's form a cover for it, and thus by compactness they form a finite subcover, and by being nested, there exists some U_N that covers D. Thus we have

$$\int_D f \leq \int_{U_N} f \leq \lim_{N\to\infty} \int_{U_N} f.$$

Taking the supremum over all compact rectifiable D on the left yields the desired result. For arbitrary f, jsut split into positive and negative parts. ■

Proposition: Let A be open in $\mathbb{R}^n$, and let $f : A \to \mathbb{R}$ be continuous. If f vanishes outside the compact subset C of A, then the integrals $\int_A f$ and $\int_C f$ exist and are equal.

Proof: The integral $\int_C f$ exists because C is bounded and $f\chi_C$ is continuous on a compact set and thus bounded on C.

Now let C_i be a sequence of compact rectifiable sets whose union is A, such that $C_i \subseteq \operatorname{int}(C_{i+1})$ for each i. Then C is covered by finitely many sets $\operatorname{int}(C_i)$, and thus by one of them, say $\operatorname{int}(C_M)$. Since f vanishes outside C,

$$\int_C f = \int_{C_N} f$$

for all $N \geq M$. Applying this to $|f|$ shows that $\lim_{N\to\infty} \int_{C_N} |f|$ exists, so f is integrable over A. Then applying to f shows that $\int_C f = \lim_{N\to\infty} \int_{C_N} f = \int_A f$. ■

14.5. Partitions of Unity

The next two sections lead up to the general change of variables formula. Previously to evaluate integrals, we broke up sets into smaller pieces and took a limit. In these sections, we break up f into compactly supported functions and take limits. To do this, we need partitions of unity, which convert local constructions to global ones (don't quote me on that, I got this from MSE).

Proposition: Let Q be a rectangle in $\mathbb{R}^n$. There is a C^∞ function $\phi : \mathbb{R}^n \to \mathbb{R}$ such that $\phi(x) > 0$ for $x \in \operatorname{int}(Q)$ and $\phi(x) = 0$ otherwise.

Proof: Let $f : \mathbb{R} \to \mathbb{R}$ be defined by the equation

$$f(x) = \begin{cases} e^{-1/x} & \text{if } x > 0 \\ 0 & \text{otherwise.} \end{cases}$$

Then $f(x) > 0$ for $x > 0$ and C^∞. Define $g(x) = f(x)f(1-x)$. It is also C^∞ and positive for $0 < x < 1$ and vanishes everywhere else. Then, if $Q = \prod [a_i, b_i]$, we have

$$\phi(x) = \prod g\left(\frac{x_i - a_i}{b_i - a_i}\right).$$

■

Proposition: Let $\mathcal{A}$ be a collection of open sets in $\mathbb{R}^n$, and let A be their union. There there exists a countable collection $Q_1, Q_2, \ldots$ of rectangles contained in A such that

a) The sets $\operatorname{int}(Q_i)$ cover A.
b) Each Q_i is contained in an element of $\mathcal{A}$.
c) Each point of A has a neighborhood that intersects only finitely many of the sets Q_i.

Proof: Let $D_1, D_2, \ldots$ be a sequence of compact subsets of A whose union is A such that $D_i \subseteq \operatorname{int}(D_{i+1})$. Define D_i to be empty when $i \leq 0$. Now define

$$B_i := D_i \backslash \operatorname{int}(D_{i-1}).$$

Clearly these are bounded since they're in compact sets, and they're also closed, so they're compact. Note that B_i is also disjoint from the closed set D_{i-2}, since $D_{i-2} \subseteq \operatorname{int}(D_{i-1})$. For $x \in B_i$, we choose a closed cube C_x centered at x that is contained in A and is disjoint from D_{i-2}. Also choose C_x small enough that it is contained in an element of the collection of open sets $\mathcal{A}$. The interiors of the cubes C_x cover B_i, so finitely many of them cover B_i. Let $\mathcal{C}_i$ denote this finite collection.

Let $\mathcal{C}$ be the union of the $\mathcal{C}_i$'s. We show this collection satisfies the proposition. By construction, each element of $\mathcal{C}$ is a rectangle contained in an element of the collection $\mathcal{A}$. Given $x \in A$, let i be the smallest integer such that $x \in \operatorname{int}(D_i)$. Then x is an element of the set $B_i = D_i \backslash \operatorname{int}(D_{i-1})$. Since the interiors of the cubes belonging to the collection $\mathcal{C}_i$ cover B_i, the point x lies interior to one of these cubes. Thus the interiors of the rectangles cover A.

Now we check the last condition. Given $x \in A$, we have $x \in \operatorname{int}(D_i)$ for some i. Each cube belonging to one of the collections $\mathcal{C}_{i+2}, \mathcal{C}_{i+3}, \ldots$ is disjoint from D_i by construction. Thus $\operatorname{int}(D_i)$ can only intersect cubes belonging to $\mathcal{C}_1, \ldots, \mathcal{C}_{i+1}$, which is a finite collection of cubes.
■

Definition (support): If $\phi : \mathbb{R}^n \to \mathbb{R}$, then the *support* of ϕ is defined to be the closure of the set $\{x : \phi(x) \neq 0\}$, and denoted with $\operatorname{supp}(f)$.

Proposition (existence of a partition of unity): Let $\mathcal{A}$ be a collection of open sets in $\mathbb{R}^n$ and let A be their union. There exists a sequence $\phi_1, \phi_2, \ldots$ of continuous functions $\phi_i : \mathbb{R}^n \to \mathbb{R}$ such that:

a) $\phi_i(x) \geq 0$ for all x.
b) The set $S_i = \operatorname{supp}(\phi_i)$ is contained in A.
c) Each points of A has a neighborhood that intersects only finitely many of the sets S_i.
d) $\sum_{i=1}^{\infty} \phi_i(x) = 1$ for each $x \in A$.
e) The functions ϕ_i are of class C^∞.
f) The sets S_i are compact.
g) For each i, the set S_i is conatined in an element of $\mathcal{A}$.

Remark: A collection of function $\{\phi_i\}$ satisfying the first four conditions is called a *partition of unity* on A. If it satisfies e), then it's *of class* C^∞. If it satisfies f), it's said to have *compact supports*. If it satisfies g), it is said to be *dominated by the collection* $\mathcal{A}$.

Proof: Given $\mathcal{A}$ and A, let $Q_1, Q_2, \ldots$ be a sequence of rectangles in A satisfying the conditions stated in the previous proposition. For each i, let $\psi_i : \mathbb{R}^n \to \mathbb{R}$ be a C^∞ function that is positive on $\operatorname{int}(Q_i)$ and zero elsewhere. Then $\psi_i(x) \geq 0$ for all x. Furthermore, $\operatorname{supp}(\psi_i) = Q_i$, which is a compact subset of A that is contained in an element of $\mathcal{A}$ by the previous proposition. Also by the previous proposition, each point of A has a neighborhood that intersects only finitely many of the sets Q_i. Thus the collection $\{\psi_i\}$ satisfies all the conditions except d).

By condition c), we know that for $x \in A$, only finitely many of $\psi_1(x), \psi_2(x), \ldots$ are nonzero, so

$$\lambda(x) = \sum_{i=1}^{\infty} \psi_i(x)$$

converges. Because each $x \in A$ has a neighborhood on which $\lambda(x)$ equals a finite sum of C^∞ functions, $\lambda(x)$ is of class C^∞. Finally, $\lambda(x) > 0$ for each $x \in A$, since by construction some rectangle Q_i contains x, and thus $\psi_i(x) > 0$. Then define

$$\phi_x(x)\psi_i(x)/\lambda(x).$$

These functions satisfy all the conditions. ■

Example: Let $f : \mathbb{R} \to \mathbb{R}$ be defined by

$$f(x) = \begin{cases} (1 + \cos x)/2 & \text{for } -\pi \leq x \leq \pi \\ 0 & \text{otherwise} \end{cases}$$

Setting $\phi_m(x) = f(x + (-1)^m m\pi)$ for $m \geq 0$ yields a partition of unity.

Proposition: Let A be open in $\mathbb{R}^n$, and let $f : A \to \mathbb{R}$ be continuous. Let $\{\phi_i\}$ be a partition of unity on A having compact supports. The integral $\int_A f$ exists if and only if the series

$$\sum_{i=1}^{\infty} \int_A \phi_i |f|$$

converges. In this case,

$$\int_A f = \sum_{i=1}^{\infty} \int_A \phi_i f.$$

Proof: We consider when f is nonnegative on A. Suppose the series converges, and let D be an arbitrary compact rectifiable subset of A. Then there exists an M such that for $i > M$m the function ϕ_i vanishes on D (take some finite open subcover of D, only finitely many of the functions will be nonzero). Then

$$f(x) = \sum_{i=1}^{M} \phi_i(x) f(x)$$

for $x \in D$. Then

$$\begin{aligned}\int_D f &= \sum_{i=1}^{M} \int_D \phi_i f \\ &\leq \sum_{i=1}^{M} \int_{D \cup S_i} \phi_i f \\ &= \sum_{i=1}^{M} \int_A \phi_i f \\ &\leq \sum_{i=1}^{\infty} \int_A \phi_i f,\end{aligned}$$

where the second equality comes from the proposition in the previous section. Since D was arbitrary, we have $\int_A f \leq \sum_{i=1}^{\infty} \int_A \phi_i f$, as desired.

Now suppose f is nonnegative on A, and suppose f is integrable over A. Note that for any N, the set $D = S_1 \cup \cdots \cup S_N$ is compact. We also have that ϕ_i vanishes outside D, so $\int_A \phi_i f = \int_D \phi f$, again by the same proposition. Thus we have

$$\begin{aligned}\sum_{i=1}^{N} \int_A \phi_i f &= \sum_{i=1}^{N} \int_D \phi_i f \\ &= \int_D \sum_{i=1}^{N} \phi_i f \\ &\leq \int_D f \\ &\leq \int_A f.\end{aligned}$$

Thus the series converges since the partial sums are increasing and bounded, which sum less than $\int_A f$. Since we have the inequality in both directions, we have equality, as desired.

Now suppose f is arbitrary. Then $\int_A f$ exists if and only if $\int_A |f|$ exists, and by the above, this occurs if and only if

$$\sum_{i=1}^{\infty} \int_A \phi_i |f|$$

converges. Then we have

$$\begin{aligned}\int_A f &= \int_A f^+ - \int_A f^- \\ &= \sum_{i=1}^{\infty} \int_A \phi_i f^+ - \sum_{i=1}^{\infty} \int_A \phi_i f^- \\ &= \sum_{i=1}^{\infty} \int_A \phi_i f.\end{aligned}$$

■

14.6. Diffeomorphisms in $\mathbb{R}^n$ (INCOMPLETE)

Definition (diffeomorphism): Let A be open in $\mathbb{R}^n$, and let $g : A \to \mathbb{R}^n$ be an injective function of class C^r such that $\det(Dg(x)) \neq 0$ for $x \in A$. Then g is a *diffeomorphism* in $\mathbb{R}^n$.

Remark: By the inverse function theorem, we know that g is invertible on the image of A.

Proposition: Let A be open in $\mathbb{R}^n$, and suppose $g : A \to \mathbb{R}^n$ be a function of class C^1. If the subset E of A has measure zero in $\mathbb{R}^n$, then the set $g(E)$ also has measure zero in $\mathbb{R}^n$.

Proof: Let $\varepsilon, \delta > 0$. First we show that if a set S has measure zero in $\mathbb{R}^n$, then S can be covered by countably many closed cubes, each of width less than δ with total volume less than ε. ∎

Proposition: Let $g : A \to B$ be a diffeomorphism of class C^r, where A and B are open sets in $\mathbb{R}^n$. Let D be a compact subset of A, and let $E = g(D)$.
a) We have

$$g(\mathrm{int}(D)) = \mathrm{int}(E) \quad \text{and} \quad g(\partial D) = \partial E.$$

b) If D is rectifiable, so is E.

Proof:

∎

Definition (primitive diffeomorphism): Let $h : A \to B$ be a diffeomorphism of open sets in $\mathbb{R}^n$ ($n \geq 2$), given by

$$h(x) = (h_1(x), ..., h_n(x)).$$

Given i, we say that h preserves the ith coordinate if $h_i(x) = x_i$ for all $x \in A$. If h preserves the ith coordinate for some i, then h is called a *primitive diffeomorphism*.

Proposition: let $g : A \to B$ be a diffeomorphism of open sets in $\mathbb{R}^n$ ($n \geq 2$). Given $a \in A$, there exists a neighborhood U_0 of a contained in A, and a sequence of diffeomorphisms of open sets in $\mathbb{R}^n$

$$U_0 \xrightarrow{h_1} U_1 \xrightarrow{h_2} U_2 \longrightarrow \cdots \xrightarrow{h_k} U_k$$

such that $h_k \circ \cdots \circ h_2 \circ h_1 = g|_{U_0}$, and such that each h_i is a primitive diffeomorphism.

Proof:

■

14.7. Change of Variables and Applications (INCOMPLETE)

Theorem (change of variables): Let $g : A \to B$ be a diffeomorphism in $\mathbb{R}^n$. Let $f : B \to \mathbb{R}$ be continuous. Then f is integrable over B if and only if the function $(f \circ g)|\det(Dg)|$ is integrable over A, in which case

$$\int_B f = \int_A (f \circ g)|\det(Dg)|.$$

Proof:

■

Example: Let B be the open set in $\mathbb{R}^2$ defined by the equation

$$B = \{(x, y) : x, y > 0 \quad \text{and} \quad x^2 + y^2 < a^2\}.$$

Suppose we want to integrate x^2y^2 over B. We use the transformation

$$g(r, \theta) = (r\cos\theta, r\sin\theta).$$

It's easy to check that $\det(Dg) = r$ and that g carries the open rectangles

$$A = \left\{(r, \theta) : 0 < r < a \quad \text{and} \quad 0 < \theta < \frac{\pi}{2}\right\}$$

bijectively. Since $\det(Dg) = r > 0$ on A, the map $g : A \to B$ is a diffeomorphism. Then by change of variables, we have

$$\int_B x^2y^2 = \int_A (r\cos\theta)^2(r\sin\theta)^2 r,$$

the latter of which can be integrated easily using Fubini's theorem.

Proposition: Let A be an $n \times n$ matrix. Let $h : \mathbb{R}^n \to \mathbb{R}^n$ be the linear transformation $h(x) = Ax$. Let S be a rectifiable set in $\mathbb{R}^n$, and let $T = h(S)$. Then

$$v(T) = |\det A|v(s).$$

Proof: First suppose A is invertible. Then h is diffeomorphism from $\mathbb{R}^n$ to itself, so h carries $\text{int}(S)$ onto $\text{int}(T)$, and T will also be rectifiable. Then by change of variables, we have

$$v(T) = v(\text{int}(T)) = \int_{\text{int}(T)} 1 = \int_{\text{int}(S)} |\det(Dh)| = \int_{\text{int}(S)} |\det A| = |\det A|v(S).$$

Now suppose A is not invertible. Then $\det A = 0$, so we show that $v(T) = 0$. Since S is bounded, so is T. Note that since h carries $\mathbb{R}^n$ to a subspace with dimension less than n, is has measure zero. Since T is a subset of this, it also has measure zero, so $v(T) = \int_T 1 = 0$. ■

14.8. Problems