Elijah Thompson

Real Analysis: A Complete Guide

– In Pursuit of Abstract Nonsense –

Wednesday 20th April, 2022

Springer Nature

Preface

This text consists of a collection of analysis notes taken at the University of Calgary.

University of Calgary, *Elijah Thompson*

Contents

Part I Single Variable Analysis **1**

1 Topology and Construction of the Real Line **3**
- 1.1 Peano Arithmetic . . . 3
- 1.2 Construction of The Rational Field . . . 7
- 1.3 Divisibility . . . 9
- 1.4 Reals in terms of Cauchy Sequences . . . 10
- 1.5 Metric Properties of the Reals . . . 15
- Appendix: Cardinality . . . 19

2 Differentiation **25**
- 2.1 Introduction to Derivatives . . . 25
- 2.2 Differentiation Results . . . 26
- 2.3 Applications of Derivatives . . . 30
  - 2.3.1 Convexity . . . 35
- 2.4 Inverse Functions . . . 37
- Appendix A: Alternative Differentiation Formulation . . . 40
- Appendix B: Functional Sequences and Series . . . 44
  - 2.4.1 Power Series . . . 56
  - 2.4.2 Products of Series . . . 64

3 Integration **67**
- 3.1 Introduction to Definite Integrals . . . 67
- 3.2 Reimann Sums . . . 75
- 3.3 The Fundamental Theorem of Calculus . . . 77
- 3.4 Content of Sets . . . 81
- Appendix A: Trigonometric Functions . . . 84
- Appendix B: The Logarithm and Exponential Functions . . . 87

Part II Higher Dimensional Analysis **93**

4 Metric Spaces **95**
- 4.1 Euclidean Spaces . . . 95
  - 4.1.1 Sequences and Convergence . . . 96
  - 4.1.2 Topological Properties of Euclidean Space . . . 97
- 4.2 Metric Space Properties . . . 98
  - 4.2.1 Normed Linear Spaces . . . 99
  - 4.2.2 Sequences and Limits . . . 105

4.3 Topology of Metric Spaces . . . 108
4.3.1 Boundary and Limit Points . . . 113
4.4 Completion of a Metric Space . . . 116
4.5 Contractive Maps and Fixed Points . . . 120
4.6 Compactness for Metric Spaces . . . 122
4.7 Product Spaces . . . 129
4.8 Baire Category Theorem . . . 131
4.9 Continuous Functions on Metric Spaces . . . 135
4.9.1 Operators and NLS . . . 138
4.9.2 Maps on Compact Spaces . . . 147
4.9.3 Uniform Continuity . . . 148
4.9.4 Sequences and Series of Functions . . . 150
4.10 Connected Sets . . . 152

5 Approximations and Continuous Functions 155
5.1 Algebras . . . 155
5.2 Stone-Weierstrass . . . 157
5.3 Equicontinuity and Arzéla-Ascoli . . . 160

6 Higher-Dimensional Differentiation 165
6.1 Basic Notions of Higher-Dimensional Derivatives . . . 165
6.2 The Inverse and Implicit Function Theorems . . . 176

7 Higher-Dimensional Integration 183

8 Manifolds 185

9 Differential Forms 187

10 Integration On Chains 189

11 Integration on Manifolds 191

Part III Function Spaces 193

12 Normed Spaces 195

13 Hilbert Spaces 197

14 Banach Spaces 199

15 Differentiation and Integration on Functions 201

16 Banach Algebras 203

Part IV Measure Theory 205

17 Riemann Integration 207

18 Lebesgue Measure on a Line **213**

19 Basic Measure Spaces **217**

20 L^p Spaces **219**

21 Radon Measures **221**

Part I

Single Variable Analysis

Chapter 1

Topology and Construction of the Real Line

1.1 Peano Arithmetic

We begin by forming our number system from the ground up, starting wtih Giuseppe Peano's (1858-1932) axiomatized system for the natural numbers.

Axiom 1.1 (Peano's Axioms) *Peano's system for the naturals consists of two central axioms:*

1. *Assume that there exists a set* $\mathbb{N}$ *and an element* $0 \notin \mathbb{N}$. *Define, notationally,* $\tilde{\mathbb{N}} = \mathbb{N} \cup \{0\}$. *Then, assume there exists an injective map* $s : \tilde{\mathbb{N}} \to \mathbb{N}$ *called the* **successor function**
2. **Mathematical Induction**: *Whenever a subset* $S \subseteq \tilde{\mathbb{N}}$ *satisfies* $0 \in S$, *and if* $k \in S$ *then* $s(k) \in S$, *then this implies* $S = \tilde{\mathbb{N}}$. *Notationally we have*

$$(0 \in S \wedge (k \in S \implies s(k) \in S)) \implies S = \tilde{\mathbb{N}}$$

Given these axioms, we can define an addition operation on $\tilde{\mathbb{N}}$:

Definition 1.1 We define the binary operation $+ : \tilde{\mathbb{N}} \times \tilde{\mathbb{N}} \to \tilde{\mathbb{N}}$ inductively for $x, y \in \tilde{\mathbb{N}}$ by stating

1. $x + 0 = x$
2. $x + s(y) = s(x + y)$

The definition is performed inductively, or recursively. Fix $x \in \tilde{\mathbb{N}}$ and let $S = \{y \in \tilde{\mathbb{N}} : x + y \text{ is defined}\}$. By definition $0 \in S$. Further, if $y \in S$, $x + y$ is defined in $\tilde{\mathbb{N}}$, so $x + s(y) := s(x + y)$. Since $x + y \in \tilde{\mathbb{N}}$, by axiom 1 of Peano's arithmetic $s(x + y) \in \mathbb{N} \subset \tilde{\mathbb{N}}$. Thus, $s(x + y)$ is defined so $s(y) \in S$. Hence by mathematical induction $S = \tilde{\mathbb{N}}$ so $+$ is a well defined binary operation for all $x, y \in \tilde{\mathbb{N}}$.

We can similarly define multiplication:

Definition 1.2 We define the binary operation $\cdot : \tilde{\mathbb{N}} \times \tilde{\mathbb{N}} \to \tilde{\mathbb{N}}$ inductively for $x, y \in \tilde{\mathbb{N}}$ by stating

1. $x \cdot 0 = 0$
2. $x \cdot s(y) = x \cdot y + x$

This defines $x \cdot y$ inductively as in the case of $+$. Now, we define our unit:

Definition 1.3 We define $1 \in \mathbb{N}$ by $1 := s(0)$.

Now we can derive many of the standard properties of the naturals.

Proposition 1.1 *For all* $x \in \tilde{\mathbb{N}}$, $x + 1 = s(x)$.

Proof Let $x \in \tilde{\mathbb{N}}$. Then $x+1 = x+s(0)$, but then by our inductive definition for $+$, $x+s(0) = s(x+0)$, and $x + 0 = x$ by definition, so

$$x + 1 = s(x)$$

Proposition 1.2 *For all* $x \in \tilde{\mathbb{N}}$, $0 + x = x$.

Proof We proceed by induction on $x \in \tilde{\mathbb{N}}$. If $x = 0$, then $0 + 0 = 0 = x$, by definition. Suppose inductively that we have $x \in \tilde{\mathbb{N}}$ such that $0 + x = x$. Then $0 + s(x) = s(0 + x)$. But, $0 + x = x$ by the induction hypothesis, so $0 + s(x) = s(x)$ and our result holds for $s(x)$. Thus, by mathematical induction we conclude that $0 + x = x$ for all $x \in \tilde{\mathbb{N}}$. □

Proposition 1.3 *For all* $x, y \in \tilde{\mathbb{N}}$, $s(y) + x = s(y + x)$.

Proof Let $y \in \tilde{\mathbb{N}}$, and we proceed by induction on $x \in \tilde{\mathbb{N}}$. If $x = 0$ then $s(y) + 0 = s(y)$, and $s(y + 0) = s(y)$, so the base case holds. Now, suppose the proposition holds for some $x \in \tilde{\mathbb{N}}$. Then $s(y)+s(x) = s(s(y)+x)$ by definition, and by the induction hypothesis $s(y)+x = s(y+x) = y+s(x)$ by definition. Thus, $s(y)+s(x) = s(y+s(x)$, as desired. Thus, by mathematical induction we conclude taht $s(y) + x = s(y + x)$ for all $x, y \in \tilde{\mathbb{N}}$. □

Proposition 1.4 *For all* $x, y \in \tilde{\mathbb{N}}$, $x + y = y + x$ *(Commutative Law).*

Proof Fix $x \in \tilde{\mathbb{N}}$, and we proceed by induction on $y \in \tilde{\mathbb{N}}$. If $y = 0$ then $x + 0 = x$ by definition, and $0 + x = x$ by Proposition 1.2, so the base case holds. Now, suppose the proposition holds for some $y \in \tilde{\mathbb{N}}$. Then $x + s(y) = s(x + y)$ by definition, and by Proposition 1.3 we have $s(y) + x = s(y + x)$. By the induction hypothesis $x + y = y + x$, so $s(x + y) = s(y + x)$, and it follows that $s(y) + x = x + s(y)$. Hence, as x was arbitrary, we have by mathematical induction that $x + y = y + x$ for all $x, y \in \tilde{\mathbb{N}}$. □

Proposition 1.5 *For all* $x, y, z \in \tilde{\mathbb{N}}$, $x + (y + z) = (x + y) + z$ *(Associative Law).*

Proof Fix $x, y \in \tilde{\mathbb{N}}$, and proceed by mathematical induction on $z \in \tilde{\mathbb{N}}$. If $z = 0$, $x + (y + 0) = x + y = (x + y) + 0$, so the base case holds. Suppose the proposition holds for some $z \in \tilde{\mathbb{N}}$. Then it follows that

$$\begin{aligned} x + (y + s(z)) &= x + s(y + z) = s(x + (y + z)) \\ &= s((x + y) + z) && \text{(by Induction Hypothesis)} \\ &= (x + y) + s(z) \end{aligned}$$

as desired. Thus by the axiom of mathematical induction we have our result. □

Proposition 1.6 *For all* $x \in \tilde{\mathbb{N}}$*,* $x \cdot 1 = x$*.*

Proof Fix$x \in \tilde{\mathbb{N}}$. Then $x \cdot 1 = x \cdot s(0) := x \cdot 0 + x = 0 + x = x$, by Proposition 1.4. □

Proposition 1.7 *For all* $x \in \tilde{\mathbb{N}}$*,* $0 \cdot x = 0$*.*

Proof We proceed by induction on $x \in \tilde{\mathbb{N}}$. If $x = 0$, $0 \cdot 0 = 0$ by definition. If $0 \cdot x = 0$ for some $x \in \tilde{\mathbb{N}}$, then $0 \cdot s(x) = 0 \cdot x + 0 = 0$, so by mathematical induction we have our result. □

Proposition 1.8 *For all* $x, y \in \tilde{\mathbb{N}}$*,* $s(x) \cdot y = x \cdot y + y$*.*

Proof Fix $x \in \tilde{\mathbb{N}}$, and proceed by induction on $y \in \tilde{\mathbb{N}}$. If $y = 0$, $s(x) \cdot 0 = 0 = x \cdot 0 + 0$. Suppose the result holds for some $y \in \tilde{\mathbb{N}}$. Then $s(x) \cdot s(y) = s(x) \cdot y + s(x)$ by definition, and $s(x) \cdot y = x \cdot y + y$ by the induction hypothesis. By Proposition 1.5, $(x \cdot y + y) + s(x) = x \cdot y + (y + s(x)) = x \cdot y + s(y + x)$. By Proposition 1.4, $x \cdot y + s(y + x) = x \cdot y + s(x + y) = x \cdot y + (x + s(y))$. Finally, by Proposition 1.5 again we have $x \cdot y + (x + s(y)) = (x \cdot y + x) + s(y) = x \cdot s(y) + s(y)$, as desired. The result follows by the principle of mathematical induction. □

Proposition 1.9 *For all* $x, y \in \tilde{\mathbb{N}}$*,* $x \cdot y = y \cdot x$*.*

Proof Fix $x \in \tilde{\mathbb{N}}$, and proceed by mathematical induction on $y \in \tilde{\mathbb{N}}$. If $y = 0$, $x \cdot 0 = 0 = 0 \cdot x$ by Proposition 1.7, so the base case holds. Suppose it holds for some $y \in \tilde{\mathbb{N}}$. Then $x \cdot s(y) = x \cdot y + x = y \cdot x + x$ by the induction hypothesis, and by Proposition 1.8, $y \cdot x + x = s(y) \cdot x$. Thus by mathematical induction we have commutivity of multiplication. □

Proposition 1.10 *For all* $x, y, z \in \tilde{\mathbb{N}}$*,* $(x + y) \cdot z = x \cdot z + y \cdot z$ *(Distributivity).*

Proof Fix $x, z \in \tilde{\mathbb{N}}$ and proceed by induction on $y \in \tilde{\mathbb{N}}$. If $y = 0$, $(x + 0) \cdot z = x \cdot z = x \cdot z + 0 \cdot z$ by Proposition 1.7. Suppose it holds for some $y \in \tilde{\mathbb{N}}$. Then $(x + s(y)) \cdot z = s(x + y) \cdot z = (x + y) \cdot z + z$ by Proposition 1.8. By the induction hypothesis, associativity, and the same proposition again we have

$$(x + y) \cdot z + z = (x \cdot z + y \cdot z) + z = x \cdot z + (y \cdot z + z) = x \cdot z + s(y) \cdot z$$

as desired. Thus we have distributivity of multiplication by mathematical induction. □

Proposition 1.11 *For all* $x, y, z \in \tilde{\mathbb{N}}$*, if* $x + z = y + z$*, then* $x = y$*.*

Proof Let $x, y \in \tilde{\mathbb{N}}$, and we proceed by induction on $z \in \tilde{\mathbb{N}}$. If $z = 0$, $x + 0 = y + 0$ implies $x = y$, so the base case holds. If it holds for some $z \in \tilde{\mathbb{N}}$, then $x + s(z) = y + s(z)$ implies $s(x + z) = s(y + z)$. But s is an inductive function, so $x + z = y + z$ which implies $x = y$ by the induction hypothesis, and by mathematical induction we have our result. □

Proposition 1.12 *If* $x \cdot z = y \cdot z$ *and* $z \neq 0$*, then* $x = y$*.*

Proof Suppose $x, y, z \in \tilde{\mathbb{N}}$ such that $x \cdot z = y \cdot z$. We argue by contrapositive and suppose $x \neq y$. Then by Trichotomy (to be shown) $x < y$ or $y < x$. Without loss of generality suppose $x < y$. Then $y = x + u$ for some $u \in \mathbb{N}$. Then $y \cdot z = x \cdot z + u \cdot z$, and so $u \cdot z = 0$ by the cancellation property. If $z \neq 0$, $z = s(w)$, $w \in \tilde{\mathbb{N}}$, so $0 = u \cdot z = u \cdot w + u$. As $u \in \mathbb{N}$ this implies $u \cdot w < 0$, but $0 \leq k$ for all $k \in \tilde{\mathbb{N}}$ which contradicts trichotomy. □

As noted in the previous proof we used properties of the standard order relation on the naturals which we shall now define and prove.

Definition 1.4 If $x, y \in \tilde{\mathbb{N}}$ we say

1. $x < y$ if $y = x + u$ for some $u \in \mathbb{N}$
2. $x \leq y$ if $y = x + v$ for some $v \in \tilde{\mathbb{N}}$

We could also say $x \leq y$ if and only if $y \in Rx = \{x + v : v \in \tilde{\mathbb{N}}\}$. We define $y > x \iff x < y$ and $y \geq x \iff x \leq y$ for all $x, y \in \tilde{\mathbb{N}}$.

Proposition 1.13 *For* $x, y \in \tilde{\mathbb{N}}$*, if* $x \leq y$ *and* $y \leq x$*, then* $x = y$*.*

Proof As $y = x + v$, $v \in \tilde{\mathbb{N}}$, and $x = y + u$ for $u \in \tilde{\mathbb{N}}$ by definition, $x = x + v + u$, so by the cancellation property $v + u = 0$. Towards a contradiction suppose v or u is not 0. Without loss of generality suppose $v \neq 0$, so $v \in \mathbb{N}$ and there exists $m \in \tilde{\mathbb{N}}$ such that $v = s(m)$. Then $0 = u + s(m) = s(u + m) \in \mathbb{N}$, which contradicts our axiom that $0 \notin \mathbb{N}$. Thus, $v = u = 0$, so $x = y$. □

Proposition 1.14 (Trichotomy) *If* $x, y \in \tilde{\mathbb{N}}$*, then one and only one of the following hold:*

$$x < y \text{ or } x = y \text{ or } x > y$$

Proof Let $x, y \in \tilde{\mathbb{N}}$. If $x < y$ then $y = x + u$ for some $u \in \mathbb{N}$. If $x = y$ then $u = 0$, but $u \in \mathbb{N}$ so $u \neq 0$, and $x \neq y$. If $x > y$ then $x = y + v$, $v \in \mathbb{N}$, but by the proof of Proposition 1.13 this implies $u = v = 0$ contradicting the fact $u, v \in \mathbb{N}$. By similar arguments $x = y \implies x \not< y, x \not> y$, and $x > y$ follows from the first case. Let $y \in \tilde{\mathbb{N}}$ and proceed by induction on $x \in \tilde{\mathbb{N}}$. If $x = 0$, $y = y + 0$ so $y \geq x$. If $y = 0$, $x = y$, and if $y \neq 0$, $y \in \mathbb{N}$ so $y > x$. Suppose the claim holds for some $x \in \tilde{\mathbb{N}}$. If $x < y$, then $s(x) \leq y$. Then either $s(x) = y$ or $s(x) + u = y$ for some $u \neq 0$, so $u \in \mathbb{N}$ and $s(x) < y$. If $x = y$, $s(x) = x + 1 = y + 1$, so $y < s(x)$. A similar argument holds if $y < x$, since $y < s(x)$, completing the induction. □

We now define the partial function of subtraction on $\tilde{\mathbb{N}}$:

Definition 1.5 If $x, y \in \tilde{\mathbb{N}}$ with $x \leq y$, then we define $z := y - x \iff y = x + z$, where $z \in \tilde{\mathbb{N}}$.

Notice $y - x$ is well defined by the cancellation property of addition.

Proposition 1.15 *If $x, y, u \in \tilde{\mathbb{N}}$, with $x \leq y$, then $(y - x)u = yu - xu$.*

Proof Let $x, y \in \tilde{\mathbb{N}}$ with $x \leq y$ and let $u \in \tilde{\mathbb{N}}$. Then there exists $w \in \tilde{\mathbb{N}}$ such that $y = x + w$, so $yu = xu + wu$ by distributivity, Then by definition $yu - xy = wu = (y - x)u$. □

Next we move on to a central property of the natural numbers which is equivalent to the axiom of mathematical induction:

Theorem 1.1 (Well-Ordering Property of $\tilde{\mathbb{N}}$) *If $T \subseteq \tilde{\mathbb{N}}$ is non-empty, then T has a smallest element.*

Proof We proceed by contrapositive. Suppose $T \subseteq \tilde{\mathbb{N}}$ and T has no smallest element. Then $0 \notin T$, since for all $x \in \tilde{\mathbb{N}}$, $0 \leq x$, as either $x = 0$ or $x \in \mathbb{N}$ so $x = x + 0$ and $x > 0$. Let $S = \{x \in \tilde{\mathbb{N}} : x < y, \forall y \in T\}$, so $0 \in S$. Inductively suppose $x \in S$. If $s(x) \in S$, we're done, so suppose $s(x) \geq y$ for some $y \in T$. But $x < y$, so $y = x + s(w)$, for some $w \in \tilde{\mathbb{N}}$, and $y = s(x) + w$. Thus $s(x) \leq y$, so by Proposition 1.13, $s(x) = y \in T$. But $s(x) \leq t$ for all $t \in T$, so $s(x)$ is a minimal element of T, contradicting the hypothesis. Thus $s(x) \in S$, so by mathematical induction $S = \tilde{\mathbb{N}}$. Then as $T \subseteq \tilde{\mathbb{N}} \backslash S$, $T = \emptyset$ as desired. □

In the next section we perform an arithmetic closure of the naturals to obtain the rational field, $\mathbb{Q}$.

1.2 Construction of The Rational Field

First we need the notion of an equivalence relation for our constructions:

Definition 1.6 (Equivalence Relation) An equivalence relation on a set S is a subset $E \subseteq S \times S$ such that

1. For all $x \in S$, xEx (reflexivity)
2. For all $x, y \in S$, if xEy then yEx (symmetry)
3. For all $x, y, z \in S$, if xEy and yEz, then xEz (transitivity)

An important property of equivalence relations is there relation to partitions of a set: in particular, we have a bijection between partitions of a set and equivalence relations.

Definition 1.7 For an equivalence relation $\sim$ on a set S, and $x \in S$, the ***equivalence class*** for x is defined by

$$[x]_\sim := \{y \in S : x \sim y\}$$

Note that $[y]_\sim = [x]_\sim$ if and only if $x \sim y$. Further, the equivalence classes for $\sim$ form a partition on S. That is $S = \cup_{x \in S}[x]_\sim$, and if $[y]_\sim \neq [x]_\sim$, $[y]_\sim \cap [x]_\sim = \emptyset$.

Definition 1.8 Define $\sim$ on $\tilde{\mathbb{N}} \times \tilde{\mathbb{N}}$ by $(a, b) \sim (x, y)$ if and only if $a + y = x + b$.

We consider (a, b) to be $a - b$. We note that this defines an equivalence relation, we the proof left to the reader.

Definition 1.9 We define the ***Integers***, $\mathbb{Z}$, to be the set

$$\mathbb{Z} := \{[(x, a)]_\sim \in \mathcal{P}(\tilde{\mathbb{N}} \times \tilde{\mathbb{N}}) : x, a \in \tilde{\mathbb{N}}\} = \tilde{\mathbb{N}} \times \tilde{\mathbb{N}}/\sim$$

and we have the natural injection

$$\begin{aligned} \iota : \tilde{\mathbb{N}} &\to \mathbb{Z} \\ x &\mapsto [(x, 0)] \end{aligned}$$

We can define operations of addition and multiplication on $\mathbb{Z}$ inherited from $\tilde{\mathbb{N}}$.

Definition 1.10 For $[(x, a)], [(y, b)] \in \mathbb{Z}$, we define

$$\begin{aligned} [(x, a)] + [(y, b)] &= [(x + y, a + b)] \\ [(x, a)] \cdot [(y, b)] &= [(xy + ab, xb + ay)] \end{aligned}$$

To show these definitions are well defined we must show that the operation is independent of the choice of representative of each equivalence class. This is a routine check left to the reader.

Definition 1.11 We define $0 = [(0, 0)]$ and $-1 = [(0, 1)]$ in $\mathbb{Z}$, and if $m = [(x, a)]$, we define $-m := [(a, x)]$.

Proposition 1.16 *For all $m \in \mathbb{Z}$, $m \cdot (-1) = -m$.*

The proof is a quick calculation:

$$m \cdot (-1) = [(x, a)][(0, 1)] = [(x \cdot 0 + a, x + a \cdot 0)] = [(a, x)] = -m$$

Similarly, we have all the properties we derived for $\tilde{\mathbb{N}}$ for the operations on $\mathbb{Z}$, such as $m \cdot 0 = 0$, $m(n + k) = mn + mk$, $m + n = m + k \implies n = k$, and $m \cdot n = k \cdot n \implies m = k$ if $n \neq 0$. Now we have closed the $\tilde{\mathbb{N}}$ under ring operations, obtaining the integral domain $\mathbb{Z}$.

Next we perform a similar construction to obtain our field - this process is known as constructing a fraction field for an integral domain.

Definition 1.12 Define an equivalence relation $\sim$ on $\mathbb{Z} \times \mathbb{Z} \backslash \{0\}$ by $(x/a) \sim (y/b)$ if and only if $xb = ya$, for $(x/a), (y/b) \in \mathbb{Z} \times \mathbb{Z} \backslash \{0\}$.

It is routine to show that this is indeed an equivalence relation on the set. Next, we can define addition and multiplication operations:

Definition 1.13 We define the rationals to be

$$\mathbb{Q} = \mathbb{Z} \times \mathbb{Z} \backslash \{0\} / \sim$$

For $[m/n], [a/b] \in \mathbb{Q}$, we define

$$[m/n] + [a/b] = [(mb + an)/(nb)]$$

and

$$[m/n] \cdot [a/b] = [(ma)/(nb)]$$

It is a routine varification that these operations are well-defined and independent of the representative. If $x = [(a/b)] \in \mathbb{Q}$, and $x \neq 0$ so $a \neq 0$, then we can define

$$x^{-1} = \frac{1}{x} := [(b/a)] \in \mathbb{Q}$$

We also define $0 := [0/1], 1 := [1/1]$, and $-1 := [-1/1]$. So far we have the chain

$$\tilde{\mathbb{N}} \hookrightarrow \mathbb{Z} := \tilde{\mathbb{N}} \times \tilde{\mathbb{N}} / \sim \hookrightarrow \mathbb{Z} \times (\mathbb{Z} \backslash \{0\}) / \sim$$

1.3 Divisibility

We now go over some fundamental theorems of number theory involving divisibility.

Definition 1.14 We say $x \in \mathbb{N}$ is ***composite*** if $a, b \in \mathbb{N}$ such that $x = ab$ and $a, b \neq 1$. If x is not composite, and $x > 1$, then x is said to be ***prime***. That is, x is prime if and only if $x = ab$ implies $a = 1$ or $b = 1$.

Definition 1.15 If $x = ab$, $x, a, b \in \mathbb{N}$, then we say a ***divides*** x, or a is a ***divisor*** of x, and we write $a \mid x$.

Definition 1.16 Given $x \in \mathbb{N}$, define the collection of non-trivial divisors as

$$D_x := \{a \in \mathbb{N} \backslash \{1\} : a \mid x\} \subseteq \mathbb{N}$$

Note that if $a \mid x$, then $a \leq x$. As $D_x \subseteq \mathbb{N}$, and $x \in D_x$ so it is non-empty, D_x has a smallest element $p_1 \in D_x$. Then $x = p_1 x_1$ for some $x_1 \in \mathbb{N}$. Then p_1 is prime since if note it can be written as $p_1 = ab$ for $1 < a < p_1$, and then $a \mid x$ with $a < p_1$, contradicting its minimality. If $x_1 > 1$, then we can obtain $x_1 = p_2 x_2$, for p_2 prime and $p_2 \geq p_1$. Repeating in this fashion, since x is finite there must exist $N \in \mathbb{N}$ such that $x_N = 1$ and $x_{N-1} = p_N \cdot 1$. Then $x = p_1 \cdot \ldots \cdot p_N$. This is the existence portion of the following result.

Theorem 1.2 (Fundamental Theorem of Arithmetic) *Every natural number $x > 1$ has a unique factorization, up to reordering, into a product of prime numbers.*

Proof If $x = p_1 \ldots p_N = q_1 \ldots q_M$, then for each $1 \leq i \leq N$, $p_i \mid q_j$ for some $1 \leq j \leq M$. But, p_i and q_j are prime, so $p_i = q_j$, and after reordering $p_1 \ldots p_N = p_1 \ldots p_N q_{N=1} \ldots q_M$. By cancellation $q_{N+1} \ldots q_M = 1$. Thus, $N = M$ and the terms are equal up to reordering. □

Definition 1.17 We say $x, y \in \mathbb{N}$ are ***coprime*** if x and y have no common prime factors.

Proposition 1.17 *If $x, y \in \mathbb{N}$ are coprime, then there exists $m, n \in \mathbb{Z}$ such that*

$$xm + ny = 1$$

1.4 Reals in terms of Cauchy Sequences

To construct the reals we use the standard notion of a completion of metric spaces using equivalence classes of Cauchy sequences. But first we must define what a sequence is, and what it means for one to be Cauchy.

Definition 1.18 We define the ***absolute value function*** by

$$|x| = \begin{cases} x & \text{if } x \geq 0 \\ -x & \text{if } x < 0 \end{cases}$$

It is a standard proof by cases, that the absolute value function defines a norm on $\mathbb{Q}$.

Definition 1.19 A sequence is a function $a : \mathbb{N} \to \mathbb{Q}$, denoted $a(j) = a_j$ and $(a_j)_{j=1}^{\infty}$.

Definition 1.20 We say a sequence (a_j) converges to a number $a \in \mathbb{Q}$, and write $a_j \to a$, if for every $n \in \mathbb{N}$, there exists an index $K(n) \in \mathbb{N}$ such that if $j \geq K(n)$, then $|a_j - a| < \frac{1}{n}$.

Definition 1.21 A sequence (a_j) is said to be ***Cauchy*** if for all $n \in \mathbb{N}$, there exists $K(n) \in \mathbb{N}$ such that if $j, k \geq K(n)$, then

$$|a_j - a_k| < \frac{1}{n}$$

Proposition 1.18 *If* $a_j \to a$*, then* (a_j) *is Cauchy.*

Proof Since $a_j \to a$, for $n \in \mathbb{N}$ there exists $K(2n) \in \mathbb{N}$ such that if $j \geq K(2n)$, $|a_j - a| < \frac{1}{2n}$. Thus, if $k, j \geq K(2n)$, then

$$|a_j - a_k| \leq |a_j - a| + |a - a_k| < \varepsilon$$

as desired. □

Proposition 1.19 *If* (a_j) *is Cauchy then* (a_j) *is bounded.*

Proof Suppose (a_j) is Cauchy. Then there exists $K(1) \in \mathbb{N}$ such that for $k, j \geq K(1)$, $|a_k - a_j| < 1$. Then for all $j \geq K(1)$, $|a_j| < 1 + |a_{K(1)}|$. Letting $M = \max\{|a_1|, ..., |a_{K(1)-1}|, 1 + |a_{K(1)}|\}$, we have that $a_n \leq M$ or all $n \in \mathbb{N}$, so the sequence is bounded. □

We now have some standard results about convergence of sequences:

Proposition 1.20 *If* $a_j \to a$ *and* $b_j \to b$*, then*

$$a_j + b_j \to a + b, \quad \textit{and} \quad a_j b_j \to ab$$

If $b \neq 0$*, and* $b_j \neq 0$ *for all* j*, then*

$$a_j / b_j \to a/b$$

More generally we have

Definition 1.22 If (a_j), (b_j) are Cauchy sequences, then $(a_j + b_j)$ is Cauchy, $(a_j b_j)$ is cauchy, and if there exists $n \in \mathbb{N}$ such that $|b_j| > \frac{1}{n}$ for all j, then (a_j / b_j) is Cauchy.

Although for general metric spaces we have the inclusion

$$\begin{Bmatrix} Convergent \\ Sequences \end{Bmatrix} \subseteq \begin{Bmatrix} Cauchy \\ Sequences \end{Bmatrix}$$

the other inclusion is not in general true.

Example 1.1 Let $a_j = \sum_{l=0}^{j} \frac{1}{l!}$, in $(\mathbb{Q}, d)$, $d(x, y) = |x - y|$. a_j is a Cauchy sequence, as $|a_j - a_k| = \left|\sum_{l=k+1}^{j} \frac{1}{l!}\right| = \sum_{l=k+1}^{j} \frac{1}{l!} \to 0$. For $l \geq 2$ we have $\frac{1/(l+1)!}{1/l!} = \frac{1}{l} \leq \frac{1}{2}$. Then $\frac{1}{(2+j)!} \leq \frac{1}{2^j}\frac{1}{2}$. Then for $j > k \geq 2$,

$$\sum_{l=k+1}^{j} \frac{1}{l!} = \sum_{l=k-2}^{j-2} \frac{1}{(l+2)!} \leq \sum_{l=k-2}^{l-2} \frac{1}{2^l}\frac{1}{2} < \frac{1}{2}\frac{1}{1 - 1/2} = 1$$

Thus, a_j is a bounded increasing sequence and hence Cauchy. Now observe

$$\begin{aligned} a_{n+j} - a_n &= \frac{1}{(n+1)!} + ... + \frac{1}{(n+j)!} \\ &\leq \frac{1}{n!}\left(\frac{1}{n+1} + \frac{1}{(n+1)^2} + ... + \frac{1}{(n+1)^j}\right) \end{aligned}$$

$$\begin{aligned} &< \frac{1}{n!}\sum_{k=0}^{j-1}\frac{1}{(n+1)^{k+1}} \\ &= \frac{1}{(n+1)!}\frac{1-\frac{1}{(n+1)^{j-1}}}{1-\frac{1}{n+1}} < \frac{1}{n!n} \end{aligned}$$

So if we fix $N \in \mathbb{N}$, $N + j > N + k \geq N$, then $a_{N+j} - a_{N+k} < \frac{1}{N!N} < \frac{1}{N}$. Hence a_j is Cauchy. Since it is Cauchy, in the complete metric space $\mathbb{R}$ it is convergent, so let $a = \lim a_j$, so we observe $a = \sum_{l=0}^{\infty} \frac{1}{l!} = e$. But $e \notin \mathbb{Q}$, so this limit cannot be in the rationals and hence the rationals is not complete.

The following is a very important series known as the ***geometric series***:

Proposition 1.21 *If* $a \in \mathbb{Q}$*, with* $|a| < 1$*, then* $\sum_{j=0}^{\infty} a^j = \frac{1}{1-a}$.

Proposition 1.22 *If* $|a| < 1$*, then* $|a|^n \to 0$.

Proof If $a = 0$, then $|a|^n = 0$ for all n, so the result holds. Hence, suppose $a \neq 0$. Then $|a|^{j+1} < |a|^j$, so $|a|^n$ is a bounded decreasing sequence, and hence converges in $\mathbb{R}$. Hence $\lim_{n\to\infty} |a|^n = k$ for some $k \in \mathbb{R}$. Then $k = \lim_{n\to\infty} |a|^n = \lim_{n\to\infty} |a|^{n+1} = |a|k$. But $|a| \neq 1$, so $k = 0$. Thus, $|a|^n \to 0$, as desired. □

Proposition 1.23 (Bolzano-Weierstrass (Cauchy)) *If* (a_j) *is a bounded sequence, then there exists a Cauchy subsequence.*

Proof Since (a_j) is bounded, there exists $M > 0$ such that $|a_j| \leq M$ for all j. In particular, $a_j \in I_0 = [-M, M]$ for all j. Then either $[-M, 0]$ or $[0, M]$ contains an infinite number of a_j. Let I_1 be the one with such. Inductively, suppose there exists $k \in \tilde{\mathbb{N}}$ such that an infinite number of a_j are in I_k, for all $0 \leq l \leq k-1$ $I_{l+1} \subseteq I_l$, and $\ell(I_l) = \frac{2M}{2^l}$. Then, we have a sequence I_j of closed intervals containing infinitely many terms of a_j. Let $b_1 = a_1$, and let $b_k = a_{j(k)}$, where $j(k) = \min\{m \in \mathbb{N} : a_m \in I_k, m > j(k-1)\}$, which exists and is well defined by the construction of I_k and the well-ordering of $\mathbb{N}$. Then $b_k = a_j(k)$ is a subsequence of j, as $j(k) < j(k+1)$ for all k. Now, fix $n \in \mathbb{N}$. As $2^{-j} \to 0$, there exists $K(2Mn) \in \mathbb{N}$ such that for $j \geq K(2Mn)$, $\frac{1}{2^j} < \frac{1}{2Mn}$. Then, for $k, l \geq K(2Mn)$, $b_k, b_l \in I_{K(2Mn)}$, so

$$|b_k - b_l| \leq \ell(I_{K(2Mn)}) = \frac{2M}{2^{K(2Mn)}} < \frac{2M}{2Mn} = \frac{1}{n}$$

Thus b_j is Cauchy. □

Corollary 1.1 *Each bounded Monotone sequence is Cauchy.*

Proof Let (a_j) be a bounded monotone sequence. Then we have a Cauchy subsequence (a_{j_n}). Fix $n \in \mathbb{N}$. Then there exists $K(n) \in \mathbb{N}$ such that if $k, l \geq K(n)$, $|a_{j_k} - a_{j_l}| < \frac{1}{n}$. Let $K'(n) = j_{K(n)}$. Then for $k, l \geq K'(n)$, let $m \in \mathbb{N}$ such that $j_m \geq k, l$ and $m \geq K(n)$. Then as a_j is an increasing (decreasing) sequence, so

$$a_{j_{K(n)}} \leq a_k \leq a_l \leq a_{j_m} \text{ (respectively } a_{j_{K(n)}} \geq a_k \geq a_l \geq a_{j_m})$$

Then $0 \leq a_l - a_k \leq a_{j_m} - a_{j_{K(n)}}$, so $|a_l - a_k| < \frac{1}{n}$, and similarly for a decreasing sequence. Hence a_j is Cauchy. □

We now begine defining the reals using equivalence relations on our Cauchy sequences:

Definition 1.23 Let $\mathcal{S} = \{(a_j) \subseteq \mathbb{Q} : (a_j) \text{ is Cauchy}\}$. Define an equivalence relation $\sim$ on $\mathcal{S}$ by

$$(a_j) \sim (b_j) \iff a_j - b_j \to 0$$

It is a routine check that $\sim$ is an equivalence relation on $\mathcal{S}$.

Definition 1.24 We define $\mathbb{R} := \mathcal{S}/\sim$, so $x \in \mathbb{R}$ if and only if $x = [(a_j)]$ for some Cauchy sequence (a_j) in $\mathbb{Q}$.

Definition 1.25 If $x = [(a_j)], y = [(b_j)] \in \mathbb{R}$, we define

$$x + y := [(a_j + b_j)] \quad xy := [(a_j b_j)]$$

and $-x := [(-a_j)]$.

As with the notion of an equivalence relation, it is a routine check using the boundedness of Cauchy sequences to prove that these operations are well defined. We can then define a natural injection $\mathbb{Q} \hookrightarrow \mathbb{R}$ by $a \mapsto [(a, a, a, ...)]$. In particular, $0 := [(0, 0, 0, ...)]$ in $\mathbb{R}$.

Now, note that if $x = [(a_j)], y = [(b_j)] \in \mathbb{R}$, then $x \neq y$ if and only if $a_j - b_j$ does not converge to 0, so there exists $n \in \mathbb{N}$ such that for all $j \in \mathbb{N}$, there exists $k \geq j$ such that

$$|a_k - b_k| \geq \frac{1}{n}$$

Specializing to the case of $y = 0 = [(0, 0, 0, ...)]$, as (a_j) is Cauchy, there exists $K(2n) \in \mathbb{N}$ such that $k, l \geq K(2n)$, $|a_k - a_l| < \frac{1}{2n}$, so in particular $|a_j| \geq |a_k| - |a_j - a_k| > \frac{1}{2n}$ for all $j \geq K(2n)$. It follows that either $a_j > \frac{1}{2n} > 0$ for all $j \geq K(2n)$, or $a_j < \frac{-1}{2n} < 0$ for all $j \geq K(2n)$.

Thus, if $x \neq 0$, then $x = [(a_j)] = [(\alpha_j)]$ such that there exists $n \in \mathbb{N}$ such that either $\alpha_j \geq \frac{1}{2n}$ for all j, or $\alpha_j \leq \frac{-1}{2n}$ for all j. Then we can define $x^{-1} = \frac{1}{x} := [(\alpha_j^{-1})]$.

Definition 1.26 We define the following subsets of $\mathbb{R}$:

$$\mathbb{R}^+ = \{x = [(a_j)] : \exists n, K \in \mathbb{N}; a_j \geq \frac{1}{2n}, \forall j \geq K\}$$
$$\mathbb{R}^- = \{x = [(a_j)] : \exists n, K \in \mathbb{N}; a_j \leq \frac{-1}{2n}, \forall j \geq K\}$$

We have shown that if $x \neq 0$ then either $x \in \mathbb{R}^+$ or $x \in \mathbb{R}^-$. Thus

$$\mathbb{R} = \mathbb{R}^+ \sqcup \{0\} \sqcup \mathbb{R}^-$$

Proposition 1.24 *For $x \in \mathbb{R}$, $x \in \mathbb{R}^+$ if and only if $-x \in \mathbb{R}^-$, and $x \in \mathbb{R}^-$ if and only if $-x \in \mathbb{R}^+$.*

Definition 1.27 We define a total order $<$ on $\mathbb{R}$ by

$$x < y \iff y - x \in \mathbb{R}^+ \iff x = [(a_j)], y = [b_j)], \exists n, K \in \mathbb{N}; b_j - a_j \geq \frac{1}{2n} \forall j \geq K$$

We have a few standard results about the order relation on $\mathbb{R}$:

Proposition 1.25 *Let $x_1, x_2, y_1, y_2 \in \mathbb{R}$. Then*

- $x_1 < y_1, x_2 < y_2 \implies x_1 + x_2 < y_1 + y_2$
- $x_1 < y_1 \implies -y_1 < -x_1$
- $0 < x_1 < y_1$, $c > 0$, *then* $0 < cx < cy$
- $0 < x < y \iff 0 < \frac{1}{y} < \frac{1}{x}$

Note that in $\mathbb{N}$ we have well-ordering, but under the standard orders on $\mathbb{Q}$ and $\mathbb{R}$ this property does not hold.

Definition 1.28 For $S \subseteq \mathbb{R}$, we say x is an ***upper bound*** of S if $s \in S$ implies $s \leq x$. Dually, we say y is a ***lower bound*** for S is $s \in S$ implies $s \geq y$.

Definition 1.29 For $S \subseteq \mathbb{R}$, the ***least upper bound***, denoted $\sup S$, is an upper bound for S such that if y is any other upper bound for S then $\sup S \leq y$. Dually, the ***greatest lower bound***, denoted $\inf S$, is a lower bound for S such that if y is any other lower bound for S, then $y \leq \inf S$.

Theorem 1.3 (Completeness of $\mathbb{R}$) *If (x_j) is a Cauchy sequence of real numbers, then there exists $x = [(a_j)] \in \mathbb{R}$ such that $x_j \longrightarrow x$, of $x_j \sim a_j$, extending the equivalence relation to $\mathbb{R}$.*

Proposition 1.26 *If S is a non-empty subset of $\mathbb{R}$ that has an upper bound, then there exists $x \in \mathbb{R}$ such that $x = \sup S$.*

Proof By hypothesis, there exists $x_0 \in \mathbb{R}$ such that for all $s \in S$, $s \leq x_0$. As S is non-empty, there exists $s_0 \in S$. Define an interval $I_0 = [s_0, x_0]$, and divide it into 2 subintervals, I_0^l, I_0^r. If $I_0^r \cap S \neq \emptyset$ let $I_1^* = I_0^r$, and otherwise let $I_1^* = I_0^l$. In either case $I_1 = [s_1, x_1]$ is such that x_1 is an upper bound of S, and where we choose $s_1 \in I_1^*$ such that $s_1 \in S$. Further, $s_0 \leq s_1 \leq x_1 \leq x_0$, and letting $x_0 - s_0 = L$, $\ell(I_1) \leq \frac{L}{2}$. Proceeding inductively we find sequences $s_0 \leq s_1 \leq s_2 \leq \ldots$ in S and $x_0 \geq x_1 \geq x_2 \geq \ldots$, with $I_j = [s_j, x_j]$, $\ell(I_j) \leq \frac{L}{2^j}$, with $x_j \geq s$ for all $s \in S$, and all $j \in \tilde{\mathbb{N}}$. Note x_j is a decreasing bounded sequence, so it converges to some $x \in \mathbb{R}$, as $\mathbb{R}$ is complete. As $x_j \geq s$ for all $s \in S$, $x \geq s$ so x is an upper bound. Further, if $\varepsilon > 0$, there exists $K \in \mathbb{N}$ such that for all $j \geq K$, $\frac{1}{2^j} < \frac{\varepsilon}{L}$, so $0 \leq x_j - s_j < \frac{L}{2^j} < \varepsilon$ for all $j \geq K$. Then $x - \varepsilon < s_j$ for all $j \geq K$. Thus, $x - \varepsilon$ is not an upper bound of S, and hence x must be the least upper bound. $\square$

1.5 Metric Properties of the Reals

First we extend our definition of sequences to the reals:

Definition 1.30 A sequence (p_j) in $\mathbb{R}$ converges to a point $p \in \mathbb{R}$ if and only if for every $\varepsilon > 0$ there exists $N \in \mathbb{N}$ such that if $j \geq N$, then $|p_j - p| < \varepsilon$.

Using sequences we can define notions of our topology, such as closed and open sets, and limit points:

Definition 1.31 $S \subseteq \mathbb{R}$ is said to be ***closed*** if and only if whenever $(p_j) \subseteq S$ is a sequence in S which converges to a point $p \in \mathbb{R}$, then $p \in S$.

Definition 1.32 A point $p \in \mathbb{R}$ is said to be a ***limit point*** of S if there exists $(p_j) \subseteq S$ such that p_j converges to p, and $p_j \neq p$ for all $j \in \mathbb{N}$.

Note that S is closed if and only if S contains all of its limit points.

Definition 1.33 $U \subseteq \mathbb{R}$ is said to be ***open*** if and only if $U^c = \mathbb{R} \backslash U$ is closed.

Definition 1.34 For $S \subseteq \mathbb{R}$, the ***closure*** of S, $\overline{S}$, is defined as

$$\overline{S} := S \cup S'$$

where

$$S' = \{p \in \mathbb{R} : p \text{ is a limit point of } S\}$$

Proposition 1.27 *For all* $S \subseteq \mathbb{R}$, $\overline{\overline{S}} = \overline{S}$

Proof Let $(p_j) \subseteq \overline{S}$ which converges to some point $p \in \mathbb{R}$. Then for each j we have (b_{jk}) in S which converges to p_j. For each $j \in \mathbb{N}$, there exists $K(j) \in \mathbb{N}$ such that for $k \geq K(j)$, $|b_{jk} - p| < \frac{1}{j}$. Define (c_j) by $c_j = b_{jK(j)}$. Fix $\varepsilon > 0$. Then there exists $N \in \mathbb{N}$ such that $j \geq N$ implies $|p_j - p| < \frac{\varepsilon}{2}$. By the Archimedian property there exists $n \in \mathbb{N}$ such that $\frac{1}{n} < \frac{\varepsilon}{2}$. Then for a $j \geq \max\{n, N\}$,

$$|c_j - p| \leq |c_j - p_j| + |p_j - p| < \frac{1}{n} + \frac{\varepsilon}{2} < \varepsilon$$

Thus $(c_j) \subseteq S$ and $c_j \to p$, so $p \in \overline{S}$. Hence, $\overline{S} \supseteq \overline{\overline{S}}$, and by definition $\overline{S} \subseteq \overline{\overline{S}}$, so $\overline{S} = \overline{\overline{S}}$. □

Theorem 1.4 *Every Cauchy sequence in* $\mathbb{R}$ *has a limit point in* $\mathbb{R}$.

Note if $(x_j) = ([(a_{jk})])$ is Cauchy in $\mathbb{R}$, then (a_{jj}) is Cauchy in $\mathbb{Q}$ with $a_{jj} - x_j \to 0$, so then x_j converges to $[(a_{jj})]$.

Proposition 1.28 (Density of the Rationals) *For all $x \in \mathbb{R}$ and $\varepsilon > 0$, there exists $y \in \mathbb{Q}$ such that $|y - x| < \varepsilon$.*

In particular, for all $a < b$ in $\mathbb{R}$, there exists $c \in \mathbb{Q}$ such that $a < c < b$. Indeed, for $x = [(a_j)]$, $a_j \to x$ so there exists $N \in \mathbb{N}$ such that $|a_N - x| < \varepsilon$ for any $\varepsilon > 0$.

Definition 1.35 A subset $K \subseteq \mathbb{R}$ is ***sequentially compact*** if and only if for every ***infinite sequence*** $(p_j) \subseteq K$, there exists a subsequence which converges to a point in K.

Theorem 1.5 (Bolzano-Weierstrass Property) *Every bounded sequence of real numbers has a convergent subsequence.*

Theorem 1.6 *If $K \neq \emptyset$, $K \subseteq \mathbb{R}$, and K is closed and bounded, then K is sequentially compact.*

Proof If $K \subseteq \mathbb{R}$, $K \neq \emptyset$, is bounded and $(p_j) \subseteq K$, (p_j) has a convergent subsequence (p_{j_k}) by Bolzano-Weierstrass, so $p_{j_k} \to p$ for some $p \in \mathbb{R}$. But K is closed so $(p_{j_k}) \subseteq K$, so $p \in K$. $\square$

Note that if $K \subseteq \mathbb{R}$ is compact, then K is closed since all subsequences of a convergent sequence converge to the same point. Additionally, K is bounded as otherwise we can construct $p_1 \in K$, $p_2 \in K$ such that $|p_2| > |p_1| + 1$, and for $p_k \in K$, choose $p_{k+1} \in K$ such that $|p_{k+1}| > |p_k| + 1$. Thus, for all $j, k \in \mathbb{N}$, $|p_j - p_k| > 1$, so (p_j) has no convergent subsequence.

Theorem 1.7 (Heine-Borel) *If $K \neq \emptyset$, $K \subseteq \mathbb{R}$, the following are equivalent:*

- *K is sequentially compact*
- *K is closed and bounded*

If K is compact, $K \neq \emptyset$, in $\mathbb{R}$, then there exists $a, b \in K$ such that

$$a = \min K := \inf K \text{ and } b = \max K := \sup K$$

which is to say K contains its infimum and supremum.

Definition 1.36 A function $f : S \to \mathbb{R}$, $\emptyset \neq S \subseteq \mathbb{R}$, is said to be ***continuous*** at a point $p \in S$ if whenever $(p_j) \subseteq S$ such that $p_j \to p$, then $f(p_j) \to f(p)$

Definition 1.37 A point $p \in S$ is said to be an ***isolated point*** of S if there exists some $\varepsilon > 0$ such that $(p - \varepsilon, p + \varepsilon) \cap S = \{p\}$

Every function is continuous at isolated points of its domain.

Definition 1.38 If $f : S \to \mathbb{R}$ is continuous at every point $p \in S$, we say f is ***continuous*** on S.

Proposition 1.29 *If $K \subseteq \mathbb{R}$, $K \neq \emptyset$, is a compact subset of $\mathbb{R}$, and $f : K \longrightarrow \mathbb{R}$ is continuous, then $f(K)$ is compact.*

Proof Let $(q_k) \subseteq f(K)$. Then we have $(p_k) \subseteq K$ such that $f(p_k) = q_k$ for all k. Then as K is sequentially compact there exists $p \in K$ and a subsequence $(p_{k_j}) \subseteq K$ such that $p_{k_j} \longrightarrow p$. As f is continuous we have $q_{k_j} = f(p_{k_j}) \longrightarrow f(p)$, where $f(p) \in f(K)$. Thus $f(K)$ is sequentially compact as claimed. □

Proposition 1.30 *If $\emptyset \neq K \subseteq \mathbb{R}$ is sequentially compact and $f : K \longrightarrow \mathbb{R}$ is continuous on K, then there exist $q, p \in K$ such that*

$$f(p) = \max_K f, \quad f(q) = \min_K f$$

Theorem 1.8 (Intermediate Value Theorem) *If $f : [a, b] \longrightarrow \mathbb{R}$ is continuous on $[a, b]$ and $f(a) < c < f(b)$ (or $f(a) > c > f(b)$), then there exists $x \in (a, b)$ such that $f(x) = c$.*

Proof Let $S = \{y \in [a, b] : f(y) \leq c\}$. Without loss of generality suppose $f(a) < c < f(b)$ (if the other inequality holds, replace f with $-f$). THen $a \in S$ and $b \notin S$. Further, if $(y_j) \in S$, $y_j \longrightarrow y$, then by continuity $f(y_j) \longrightarrow f(y)$ and as $f(y_j) \leq c$ for all j, $f(y) \leq c$. Thus $y \in S$, so S is closed. As $S \subseteq [a, b]$, S is bounded, so by Heine-Borel S is compact. Let $x = \max S \in S$. Then $f(x) \leq c$. If $f(x) < c$, then for some $\varepsilon > 0$, $f(y) < c$ for all $y \in (x - \varepsilon, x + \varepsilon)$. Let $\varepsilon = c - f(x) > 0$. As f is continuous, there exists $\delta > 0$ such that if $|x - y| < \delta$, $|f(x) - f(y)| < \frac{\varepsilon}{2}$. Then $c - f(y) = c - f(x) - (f(y) - f(x)) \geq |c - f(x)| - |f(y) - f(x)|$, which is greater than or equal to $\varepsilon/2$, so $f(y) < c$. Taking $y = x + \frac{\delta}{2}$, $f(y) \leq c$, so $y \in S$, but $y > x = \max S$ which is a contradiction. Thus, $f(x) \not< c$, so $f(x) = c$. □

Proposition 1.31 *Suppose K is sequentially compact and $X_1 \supseteq X_2 \supseteq X_3 \supseteq \ldots$ is a sequence of closed subsets of K so all X_i are compact by Heine-Borel. If $X_m \neq \emptyset$, for all m, then $\bigcap_{m \geq 1} X_m \neq \emptyset$.*

Proof Let $x_m \in X_m$. As K is compact and $x_m \in K$, there exists a convergent subsequence $x_{m_j} \longrightarrow x \in K$. Since $X_1 \supseteq X_2 \supseteq \ldots$,

$$\{x_{m_l} : l \geq j\} \subseteq X_{m_j}$$

But X_{m_j} is closed, so $x \in X_{m_j}$, for all j. Then for all $m \in \mathbb{N}$, $x \in X_m$ as for all $n \in \mathbb{N}$ there exists $m_j \geq n$ so $x \in X_{m_j} \subseteq X_n$. Thus $x \in \bigcap_{m \geq 1} X_m$,a s claimed. □

Corollary 1.2 *If K is sequentially compact and $U_1 \subseteq U_2 \subseteq \ldots$ is a sequence of open sets such that $K \subseteq \bigcup_{j \geq 1} U_j$, then there exists $M \in \mathbb{N}$ such that $K \subseteq U_M$.*

Use Proposition 1.31 with $X_j = K \backslash U_j$.

We now discuss some general results on open and closed sets:

Proposition 1.32 *If $\{A_\alpha\}_{\alpha \in J}$ is a family of closed sets in $\mathbb{R}$, then $\bigcap_{\alpha \in J} A_\alpha$ is closed. If A and B are closed, then $A \cup B$ is closed.*

Proof Suppose $A_\alpha, \alpha \in J$ are as in the hypothesis. If $\emptyset = \bigcap_{\alpha \in J} A_\alpha$ then the claim vauously holds. Otherwise, let $(p_j) \subseteq \bigcap_{\alpha \in J} A_\alpha$ such that $p_j \longrightarrow p \in \mathbb{R}$. As A_α is closed for all $\alpha \in J$, it follows that $p \in A_\alpha$, so $p \in \bigcap_{\alpha \in J} A_\alpha$.

Next, let A and B be closed, and take $(p_j) \subseteq A \cup B$ such that $p_j \to p \in \mathbb{R}$. Either infinitely many points of the sequence are in A or infinitely many are in B. Without loss of generality suppose infinitely many are in A. Then there exists a subsequence $(p_{j_k}) \subseteq (p_j)$ contained in A, which will converge to p so as A is closed $p \in A$. Thus $p \in A \cup B$, so the union is closed. □

Corollary 1.3 *If* $\{U_\alpha\}_{\alpha \in J}$ *is a family of open sets in* $\mathbb{R}$*, then* $\bigcup_{\alpha \in J} U_\alpha$ *is open. If* A *and* B *are open, then* $A \cap B$ *is open.*

Conversely, we have $I_n = [-1 + \frac{1}{n}, 1 - \frac{1}{n}]$, a collection of closed intervals, who's union is $\bigcup_{n \geq 1} I_n = (-1, 1)$ is not closed. Further, if $U_n = (-\frac{1}{n}, 1 + \frac{1}{n})$, then $\bigcap_{n \geq 1} U_n = [0, 1]$ is not open.

Definition 1.39 The ball of radius $r > 0$ centered at $x \in \mathbb{R}$ is defined by

$$B_r(x) := \{y \in \mathbb{R} : |x - y| < r\}$$

Proposition 1.33 $U \subseteq \mathbb{R}$ *is open if and only if for all* $x \in U$*, there exists* $r_x > 0$ *such that* $B_{r_x}(x) \subseteq U$.

Proof First we prove the reverse implication:

- $\Longleftarrow$ If for all $x \in U$ there exists $r_x > 0$ such that $B_{r_x}(x) \subseteq U$, then $U = \bigcup_{x \in U} B_{r_x}(x)$. Each $B_{r_x}(x) = (x - r_x, x + r_x)$ is open as $(-\infty, x - r_x] \cup [x + r_x, \infty)$ is a finite union of closed sets. Thus U is open, being the union of open sets.
- $\Longrightarrow$ Towards a contradiction there exists $x \in U$ such that for all $r > 0$, $B_r(x) \cap U^c \neq \emptyset$. Then by the axiom of choice there exists $f : X \to \bigcup X$ with $X = \{B_{1/n}(x) \cap U^c : n \in \mathbb{N}\}$, and we can define a sequence $a(n) = f(B_{1/n}(x) \cap U^c)$ which converges to x, and $(a_n) \subseteq U^c$ which is closed. But $x \in U$ implies $x \notin U^c$ contradicting the closedness of U^c, so U must satisfy the hypothesis. □

We mention some relevant and important topological properties which $\mathbb{R}$ satisfies, but we leave the definition of a topology to later on.

Definition 1.40 A topological space (X, τ) is called ***separable*** if it has a ***countable dense subset***.

$\mathbb{R}$ is an example of a separable space, with countable dense subset $\mathbb{Q}$.

Definition 1.41 A topological space (X, τ) is called ***Lindelöf*** if and only if every open cover of X has a countable subcover.

Definition 1.42 A topological space (X, τ) is called ***second countable*** if τ has a countable base.

All metrizable spaces are second countable if and only if they are separable, as we can take the rational balls around points of the countably dense subset. THus, $\tau_{\mathbb{R}} = \bigcup \mathcal{B}$, where $\mathcal{B} = \{B_p(q) : p, q \in \mathbb{Q}, p > 0\}$ is a countable base.

Definition 1.43 A ***covering*** of a set F is any family of sets $\{X_\alpha\}_{\alpha\in\mathcal{A}}$, such that $F \subseteq \bigcup_{\alpha\in\mathcal{A}} X_\alpha$.

Definition 1.44 An ***open covering*** of a topological space is a covering by open sets.

Proposition 1.34 *If $K \subseteq \mathbb{R}$ is sequentially compact, then every open covering of K has a finite subcovering.*

Proof Suppose $\emptyset \neq K \subseteq \mathbb{R}$ is sequentially compact. Let $K \subseteq \bigcup_{\alpha\in J} U_\alpha$ be an open covering. Note $U_\alpha = \bigcup_{x\in U_\alpha} B_{p_{\alpha,x}}(q_{\alpha,x})$ such that $x \in B_{p_{\alpha,x}}(q_{\alpha,x})$, as U_α is open. Hence

$$K \subseteq \bigcup_{\alpha\in J} U_\alpha = \bigcup_{\alpha\in J} \bigcup_{x\in U_\alpha} B_{p_{\alpha,x}}(q_{\alpha,x})$$

but the right side consists of countably many distinct open sets. Now consider $K \subseteq \bigcup_{j\geq 1} V_j$ is countable. Let $J_n = \bigcup_{j=1}^{n} V_j$. Then $K \subseteq \bigcup_{j\geq 1} J_j$, and $J_1 \subseteq J_2 \subseteq \ldots$ so by Proposition **??** there exists $M \in \mathbb{N}$ such that $K \subseteq J_M = \bigcup_{j=1}^{M} V_j$. Thus, in particular there exist $B_{p_{\alpha,x_1}}(q_{\alpha,x_1}), \ldots, B_{p_{\alpha,x_M}}(q_{\alpha,x_M})$ such that $K \subseteq \bigcup_{j=1}^{M} B_{p_{\alpha,x_j}}(q_{\alpha,x_j})$. Thus $K \subseteq \bigcup_{j=1}^{M} U_{\alpha_j}$, so we have a finite subcovering of K as desired. □

Proposition 1.35 *If $K \subseteq \mathbb{R}$ is topologically compact, then K is sequentially compact.*

Proof We prove the equivalent notion of limit point compactness (which is equivalent for metric spaces). Then we argue by contrapostive, supposing S has no accumulation points. Then S and $S_x = S\backslash\{x\}$ is closed for all $x \in S$. Setting $U_x = \mathbb{R}\backslash S_x$, we have that $K = (\bigcup_{x\in S} U_x) \cup \mathbb{R}\backslash S$ is an open cover, and as K is topologically compact we have a finite subcover $U_{x_1}, \ldots, U_{x_N}, \mathbb{R}\backslash S$. Then $U_{x_1}, \ldots, U_{x_N}$ covers S, so $S = \bigcup_{i=1}^{N} U_{x_i} \cap S = \bigcup_{i=1}^{N}\{x_i\} = \{x_1, \ldots, x_N\}$, so S is finite. Thus, K is limit point compact as desired. □

Thus, we have the equivalence between sequentially compact, topologically compact, and limit point compact, for subsets of $\mathbb{R}$.

Appendix: Cardinality

In this appendix we investigate the notion of the size of a set. First, we define the set

$$I_n := \{j \in \mathbb{N} : j \leq n$$

the prototypical set of size n.

Lemma 1.1 $I_1 = \{1\}$, *and* $I_{n+1} = I_n \cup \{n+1\}$.

Proof If $n = 1$, $I_1 = \{j \in \mathbb{N} : j \leq 1\} = \{1\}$, so the base case holds. Further, $I_2 = \{j \in \mathbb{N} : j \leq 2\} = \{1, 2\} = I_1 \cup \{2\}$. Now, suppose for some $j \in \mathbb{N}$, $I_{j+1} = I_j \cup \{j+1\}$. Then

$$I_{j+2} = \{k \in \mathbb{N} : k \leq j+2\} = \{k \in \mathbb{N} : k \leq j+1 \vee k = j+2\} = I_{j+1} \cup \{j+2\}$$

Definition 1.45 A non-empty set S has n elements if and only if there exists a bijective map $\varphi : S \longrightarrow I_n$.

Proposition 1.36 *For $m, n \in \mathbb{N}$, if there exists an injection $\varphi : I_m \longrightarrow I_n$, then $m \leq n$.*

Proof If $n = 1$, then $I_n = I_1 = \{1\}$. Now, suppose $\varphi : I_m \longrightarrow I_1$ is an injection. If $x, y \in I_m$, then $\varphi(x) = 1 = \varphi(y)$, so by injectivity $x = y$. Thus, I_m has only one element, and since $1 \in I_m$, we must have $I_m = \{1\} = I_1$. Hence, $m = 1 \leq 1 = n$. Assume the result is true for some $1 \leq n < N$. Then let $\varphi : I_m \longrightarrow I_N$ be an injection. If $m = 1$ the result is immediately satisfied, so suppose $m \geq 2$.

(1)Suppose there exist $j \in I_m$ such that $\varphi(j) = N$. Then define $\psi : I_{m-1} \longrightarrow I_{N-1}$ by $\psi(l) = \varphi(l)$ for $l < j$, and $\psi(l) = \varphi(l+1)$ for $j \leq l \leq m-1$. Then ψ is injective because φ is injective, so the the induction hypothesis $m - 1 \leq N - 1$, so $m \leq N$.

(2)If there does not exist $j \in I_m$ such that $\varphi(j) = N$, then we can restrict the codomain of φ to obtain $\varphi : I_m \longrightarrow I_{N-1}$. By the induction hypothesis $m \leq N - 1 < N$, as desired. □

Corollary 1.4 *If there exists $\varphi : I_m \longrightarrow I_n$ bijective, then $m = n$.*

Corollary 1.5 *Suppose S is a set, $n, m \in \mathbb{N}$, and there exist bijections $\varphi : S \longrightarrow I_n$ and $\psi : S \longrightarrow I_m$, then $n = m$.*

The result follows from considering $\varphi \circ \psi^{-1} : I_m \longrightarrow I_n$.

Definition 1.46 If either $S = \emptyset$ or S has n elements for some $n \in \mathbb{N}$, then we say S is ***finite***. Otherwise, S is said to be ***infinite***.

Proposition 1.37 *If $n \in \mathbb{N}$ and $S \subseteq I_n$, then there exists $m \leq n$ and $\varphi : S \longrightarrow I_m$ bijective.*

Proof If $n = 1$, then $I_1 = \{1\}$ and the only non-empty subset is $S = \{1\} = I_1$ itself. Then $\varphi : S \longrightarrow I_1$ given by $\varphi(1) = 1$ is a bijection so the base case holds. Suppose for $k \geq 1$, if $S \subseteq I_k$, then there exists $m \leq k$ such that $\varphi_k : S \longrightarrow I_m$, bijective. Suppose $S \subseteq I_{k+1}$. If $S = I_{k+1}$, then $\mathrm{id} : I_{k+1} \longrightarrow I_{k+1}$ is the desired bijection. Otherwise, there exists $j \in I_{k+1}$ such that $j \notin S$. If $j = k + 1$, define $\varphi : S \longrightarrow I_k$ by $\varphi(s) = s$ for all $s \in S$. Then φ is an injection and $\varphi(S) \subseteq I_k$ so by the induction hypothesis there exists $m \leq k$ and a bijection $\psi : \varphi(S) \longrightarrow I_m$. Then $\psi \circ \varphi : S \longrightarrow I_m$ is bijective. If $j \neq k + 1$, then define $\varphi(S) \longrightarrow I_k$ by $\varphi(l) = l$ for $l \leq k$, and $\varphi(k + 1) = j$. Then φ is injective, and by the induction hypothesis there exists $m \leq k$ and a bijection $\psi : \varphi(S) \longrightarrow I_m$, so $\psi \circ \varphi : S \longrightarrow I_m$ is the desired bijection. □

Proposition 1.38 *$\mathbb{N}$ is not finite.*

Proof If $\mathbb{N}$ was finite then there would exist $n \in \mathbb{N}$ and a bijection $\varphi : I_n \longrightarrow \mathbb{N}$. As $I_{n+1} \subseteq \mathbb{N}$, if we restrict φ to $\varphi^{-1}(I_{n+1}) = S$, we have $\psi : S \longrightarrow I_{n+1}$, which is bijective. But $S \subseteq I_n$, so by Proposition 1.37 S has m elements with $m \leq n$, and as $n + 1 = m \leq n$, a contradiction. Thus, $\mathbb{N}$ must be infinite.□

Proposition 1.39 *If S is not finite, then there exists an injection $\varphi : \mathbb{N} \longrightarrow S$.*

Proof As S is non-empty, there exists $s_1 \in S$, so define $\varphi_1(1) = s_1$. Suppose there exists $k \in \mathbb{N}$ such that for $\varphi_k(l) = s_l \in S \backslash \{s_1, ..., s_{l-1}$ for all $1 \leq l \leq k$. Then as S is infinite, there exists $s_{k+1} \in S$ with $s_{k+1} \notin \{s_1, ..., s_k\}$, so we define $\varphi_{k+1}(k+1) = s_{k+1}$ and $\varphi_{k+1}(l) = \varphi_k(l)$ for all $1 \leq l \leq k$. Thus, for all $n \in \mathbb{N}$ we have $\varphi_n : I_n \longrightarrow S$ injective, and for all $m \leq n$, $\varphi_n|_{I_m} = \varphi_m : I_m \longrightarrow S$. Define $\Phi(m) = \varphi_n(m)$ for all $n \geq m$. Then $\Phi : \mathbb{N} \longrightarrow S$ is well defined and injective. □

Definition 1.47 We say that a set S is ***countably infinite*** if there exists a bijection $\varphi : S \longrightarrow \mathbb{N}$.

Definition 1.48 Two sets S and T have the same cardinality if there exists a bijection between them, and we write $\mathrm{Card}(S) = \mathrm{Card}(T)$.

Definition 1.49 If S is finite:

- $\mathrm{Card}(S) = 0$ if $S = \emptyset$
- $\mathrm{Card}(T) = n$ if S has n elements.

We now move on to a fundamental theorem on cardinalities of sets:

Theorem 1.9 (Schröder-Bernstein Theorem) *If S and T are two non-empty sets such that there exist injective maps*

$$\varphi : S \hookrightarrow T \quad \textit{and} \quad \psi : T \hookrightarrow S$$

then there exists a bijection $\Phi : S \longrightarrow T$.

Proof Let $\varphi : S \longrightarrow T$ and $\psi : T \longrightarrow S$ be the injections in the hypothesis. If $s \in S$ such that $\varphi(s) = t$, we say s is a parent of t. Similarly, if $t' \in T$ such that $\psi(t') = s' \in S$, then we say t' is a parent of s'. For elements in S and T there are three disjoint cases:

a) The set of elements who have an infinite number of ancestors

b) The set of elements whose last ancestor is an element of S

c) The set of elements whose last ancestor is an element of S

Define $S_a, T_a, S_b, T_b, S_c, T_c$ be the corresponding sets, so $S = S_a \sqcup S_b \sqcup S_c, T = T_a \sqcup T_b \sqcup T_c$. I claim that $\varphi|_{S_a} : S_a \longrightarrow T_a$ is bijective. Indeed, φ is injective. If $s \in S_a$, $\varphi(s) \in T_a$ as all ancestors of s are ancestors of $\varphi(s)$. If $t \in T_a$, then t has infinitely many ancestors, so in particular there exists $s \in S$ such that $\varphi(s) = t$. But, all ancestors of t, except s, are ancestors of s so $s \in S_a$. Hence $\varphi|_{S_a}$ is indeed well defined and bijective.

Next we claim that $\varphi|_{S_b} : S_b \longrightarrow T_b$ is bijective. If $s \in S_b$, then s's ancestors terminate in S, so $\varphi(s) \in T_b$. If $t \in T_b$, then t has at least one ancestor $s \in S$, and as $\varphi(s) \in T_b$, $s \in S_b$. Thus $\varphi|_{S_b}$ is well defined and bijective. Dually, we have that $\psi|_{T_c} : T_c \longrightarrow S_c$ is bijective. Then define $\Phi : S \longrightarrow T$ by $\Phi(s) = \varphi(s)$ for $s \in S_a \cup S_b$, and $\Phi(s) = \psi|_{T_c}^{-1}(s)$ for $s \in S_c$, which is by construction bijective.□

A classical application of this result is in the following example:

Example 1.2 Card($\mathbb{N}$) = Card($\mathbb{N} \times \mathbb{N}$). Define $\varphi : \mathbb{N} \to \mathbb{N} \times \mathbb{N}$ by $\varphi = \Delta$ is the diagonal, so φ is injective. Conversely, define $\psi : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$ by $\psi(n, m) = 2^n 3^m$, so by the fundamental theorem of arithmetic ψ is injective. Then by Schröder-Bernstein there exists a bijection $\Phi : \mathbb{N} \to \mathbb{N} \times \mathbb{N}$, as desired.

The following are the axioms founding the Zermelo-Frankel axioms of set theory with Choice:

Axiom 1.2 (ZFC Axioms) *The following constitute the axioms of ZFC:*

1. *(Emptyset) There is a set, denoted by $\emptyset$, which has no members:*

$$\exists x \forall t \ \neg t \in x$$

2. *(Pairset) For any two sets x and y there is a set p with the property that $t \in p$ if and only if $t = x$ or $t = y$. This set p is usually denoted $\{x, y\}$:*

$$\forall x \forall y \exists p (t \in p \iff (t = x \vee t = y))$$

3. *(Extensionality) For any two sets x and y, $x = y$ if and only if x and y have exactly the same members:*

$$\forall x \forall y (x = y \iff \forall t (t \in x \iff t \in y))$$

4. *(Union Set) For any set x there exists a set denote by $\bigcup x$ whose members are exactly the members of the members of x:*

$$\forall x \exists y \forall t (t \in y \iff \exists y (y \in x \wedge t \in y))$$

5. *(Infinity) There exists a set I which contains $0 = \emptyset$ as well as the successor of each of its members; that is, if $x \in I$, then $S(x) := x \cup \{x\} \in I$:*

$$\exists x (\emptyset \in I \wedge \forall x (x \in I \implies \bigcup \{x, \{x\}\}))$$

6. *(Powerset) For each set A there exists a set $\mathcal{P}(A)$ whose members are the subsets of A:*

$$\forall x \exists y \forall t (t \in y \iff \forall z (z \in t \implies z \in x))$$

7. *(Separation) Suppose P is a definite condition. For each set A there exists a set B whose members are exactly the members of A that satisfy P:*

$$\forall x \exists y \forall t (t \in B \iff (t \in x \wedge P(t)))$$

8. *(Replacement) Suppose P is a definite binary condition such that for each set x there is a unique set y for which $P(x, y)$ holds. Given a set A there exists a set B with the property that $y \in B$ if and*

only if there exists $x \in A$ such that $P(x, y)$:

$$\forall x \exists y \forall t (t \in y \iff \exists z (z \in x \land P(z, t)))$$

9. *(Regularity) Every non-empty set A contains an element that is disjoint from A:*

$$\forall x (\neg x = \emptyset \implies \exists t (t \in x \land \forall y (y \in t \implies \neg y \in x)))$$

10. *(Choice) Every non-empty set X whose members are all non-empty sets, there exists a function $f : X \to \bigcup X$ such that $f(A) \in A$ for all $A \in X$.*

Lemma 1.2 *If $\varphi : S \to T$ is onto, then there exists $\psi : T \to S$ which is injective.*

Proof Let $\varphi : S \to T$ be onto. Then let $X = \{\varphi^{-1}(t) : t \in T\}$, so $S = \bigcup X$, and all members of X are non-empty sets since φ is onto. By the axiom of choice there exists a function $f : X \to \bigcup X$ such that $f(\varphi^{-1}(t)) \in \varphi^{-1}(t)$ for all $t \in T$. Since φ is a well defined function f is injective. Let $g : T \to X$ by $g(t) = \varphi^{-1}(t)$. Then g is also injective as φ is well-defined, so their composite $f \circ g : T \to \bigcup X = S$ is an injection. □

As an application of our results, I claim that $\text{Card}(\mathbb{R}) \neq \text{Card}(\mathbb{N})$:

Proof First, $\iota : (0, 1) \to \mathbb{R}$, being the natural inclusion, is an injection, and $\varphi : \mathbb{R} \to (0, 1)$ given by $\varphi(x) = \frac{e^x}{e^x+1}$ is an injection, so by Schröder Bernstein there exists a bijection $\Phi : (0, 1) \to \mathbb{R}$. Towards a contradiction suppose $\text{Card}(\mathbb{N}) = \text{Card}(\mathbb{R}) = \text{Card}((0, 1))$, so we have a bijection $f : \mathbb{N} \to (0, 1)$. Expand the terms in their decimal expansion, so $f(j) = \sum_{n=1}^{\infty} a_{jn} 10^{-n}$, $a_{jn} \in \{0, 1, 2, ..., 9\}$. Define x by $x = \sum_{n=1}^{\infty} b_n 10^{-1}$ where $b_n = 2$ if $a_{nn} \neq 2$ and $b_n = 3$ if $a_{nn} = 2$. Thus $x \neq f(j)$ for all $j \in \mathbb{N}$, but by assumption f is bijective, and hence onto, which is a contradiction since $x \in (0, 1)$. Thus, no such bijection can exist. □

Definition 1.50 For sets S and T we define $\leq$ on the cardinals by $\text{Card}(S) \leq \text{Card}(T)$ if and only if there exists an injection $\varphi : S \to T$, and $\text{Card}(S) < \text{Card}(T)$ if and only if $\text{Card}(S) \leq \text{Card}(T)$ and $\text{Card}(S) \neq \text{Card}(T)$.

Then it follows that $\text{Card}(\mathbb{N}) < \text{Card}(\mathbb{R})$.

Conjecture 1.1 (Continuum Hypothesis) There exists no set with cardinality strictly between $\aleph_0 = \text{Card}(\mathbb{N})$ and $\aleph_1 = \text{Card}(\mathbb{R})$.

This hypothesis cannot be proven and taking it to be true or false in your system for set theory leads to no contradictions.

Proposition 1.40 *For any set S, $\mathcal{P}(S) \cong 2^S$.*

Proof Let $\varphi : \mathcal{P}(S) \to 2^S$ by $\varphi(A)(s) = 1$ if and only if $s \in A$, and 0 otherwise, for all $A \in \mathcal{P}(S)$. This is an injection, and a surjection as we can take $X_f = \{s \in S : f(s) = 1\}$ for all $f \in 2^S$, so $\varphi(X_f) = f$. □

Proposition 1.41 (Cantor's Theorem) *For any set S, Card(S) < Card$(\mathcal{P}(S))$*

Proof The inclusion $\iota : S \to \mathcal{P}(S)$ by $\iota(s) = \{s\}$ for all $s \in S$, is an injection so $\mathrm{Card}(S) \leq \mathrm{Card}(\mathcal{P}(S))$. Towards a contradiction we have $f : S \twoheadrightarrow \mathcal{P}(S)$. Let $B = \{s \in S : s \notin f(s)\}$. Then $f(s) = B$ for some $s \in S$. But then

$$s \notin B \iff s \notin f(s) \iff s \notin B$$

which is a contradiction, so no such f can exist. $\square$

Due to Cantor's theorem we obtain an infinite chain of infinite cardinals

$$\mathrm{Card}(\mathbb{N}) < \mathrm{Card}(\mathcal{P}(\mathbb{N})) < \mathrm{Card}(\mathcal{P}(\mathcal{P}(\mathbb{N}))) < \ldots$$

An important, but non-trivial result, is $\mathrm{Card}(\mathbb{R} \times \mathbb{R}) = \mathrm{Card}(\mathbb{R})$.

Chapter 2

Differentiation

2.1 Introduction to Derivatives

Definition 2.1 (Differentiability) A function $f : \mathbb{R} \to \mathbb{R}$ is said to be ***differentiable at*** a if

$$\lim_{h \to 0} \frac{f(a+h) - f(a)}{h} \tag{2.1.1}$$

exists. In this case the limit is denoted by $f'(a)$ and is called the ***derivative of*** f ***at*** a. We also say that f is ***differentiable*** if f is differentiable at a for all a in its domain.

Definition 2.2 We define the ***tangent line*** to the graph of f at $(a, f(a))$ to be the line through $(a, f(a))$ with slope $f'(a)$. That is, the tangent line at $(a, f(a))$ is well defined if and only if f is differentiable at a.

Remark 2.1 Given a function f, we denote by f' the function whose domain is the set of all numbers $a \in \mathbb{R}$ such that f is differentiable at a, and whose value at such a number a is

$$\lim_{h \to 0} \frac{f(a+h) - f(a)}{h} \tag{2.1.2}$$

The function f' is called the ***derivative*** of f.

Note 2.1 For a given function $f : \mathbb{R} \to \mathbb{R}$, the derivative f' is often denoted by

$$\frac{df(x)}{dx} \tag{2.1.3}$$

and the number $f'(a)$ is denoted by

$$\left.\frac{df(x)}{dx}\right|_{x=a} \tag{2.1.4}$$

Theorem 2.1 *If f is differentiable at a, then f is continuous at a.*

Proof Suppose f is differentiable at a point a. Then we have that the limit

$$\lim_{h\to 0}\frac{f(a+h)-f(a)}{h}$$

exists. It follows by **??** that

$$\begin{aligned}\lim_{h\to 0} f(a+h)-f(a) &= \lim_{h\to 0}\frac{f(a+h)-f(a)}{h}\cdot h\\ &= \lim_{h\to 0}\frac{f(a+h)-f(a)}{h}\cdot\lim_{h\to 0} h\\ &= f'(a)\cdot 0\\ &= 0\end{aligned}$$

Thus, by **??** the result that $\lim_{h\to 0} f(a+h)-f(a)=0$ is equivalent to $\lim_{h\to 0} f(a+h)=\lim_{h\to 0} f(a)=f(a)$. Thus, f is continuous at a, replacing $a+h$ with x and $h\to 0$ with $x\to a$. □

Definition 2.3 (Higher Order Derivatives) Since the derivative of a function f is also a function, we can take its derivative to obtain the function $(f')'=f''$. In general, we denote the $k+1$-th derivative of f inductively by

$$\begin{aligned}f^{(1)}&=f'\\ f^{(k+1)}&=(f^{(k)})'\end{aligned}$$

These are called ***higher order derivatives of*** f. We also define $f^{(0)}=f$. In Leibnitzian notation we write

$$\frac{d^k f(x)}{dx}=f^{(k)} \tag{2.1.5}$$

2.2 Differentiation Results

Theorem 2.2 *If f is a constant function, $f(x)=c$, then $f'(a)=0$ for all $a\in\mathbb{R}$.*

Proof Observe that for $a\in\mathbb{R}$,

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

as desired. □

Theorem 2.3 *If f is the identity function, $f(x)=x$, then $f'(a)=1$ for all $a\in\mathbb{R}$.*

Proof Observe that for $a\in\mathbb{R}$,

$$f'(a)=\lim_{h\to 0}\frac{f(a+h)-f(a)}{h}=\lim_{h\to 0}\frac{a+h-a}{h}=\lim_{h\to 0}1=1$$

as desired. □

Theorem 2.4 (Linearity) *If f and g are differentiable at a, then $f + cg$ is differentiable for all $c \in \mathbb{R}$*

Proof Observe that

$$\begin{aligned}
(f+cg)'(a) &= \lim_{h\to 0}\frac{(f+cg)(a+h)-(f+cg)(a)}{h}\\
&= \lim_{h\to 0}\frac{f(a+h)+cg(a+h)-[f(a)+cg(a)]}{h}\\
&= \lim_{h\to 0}\frac{[f(a+h)-f(a)]+c[g(a+h)-g(a)]}{h}\\
&= \lim_{h\to 0}\left(\frac{f(a+h)-f(a)}{h}+c\frac{g(a+h)-g(a)}{h}\right)\\
&= \lim_{h\to 0}\frac{f(a+h)-f(a)}{h}+\lim_{h\to 0}c\frac{g(a+h)-g(a)}{h}\\
&= f'(a)+c\lim_{h\to 0}\frac{g(a+h)-g(a)}{h}\\
&= f'(a)+cg'(a)
\end{aligned}$$

as desired. □

Theorem 2.5 (Product Rule) *If f and g are differentiable at a, then $f \cdot g$ is also differentiable at a and*

$$(f\cdot g)'(a) = f'(a)\cdot g(a) + f(a)\cdot g'(a)$$

Proof Observe that

$$\begin{aligned}
(f\cdot g)'(a) &= \lim_{h\to 0}\frac{(f\cdot g)(a+h)-(f\cdot g)(a)}{h}\\
&= \lim_{h\to 0}\frac{f(a+h)g(a+h)-f(a+h)g(a)+f(a+h)g(a)-f(a)g(a)}{h}\\
&= \lim_{h\to 0}\frac{f(a+h)[g(a+h)-g(a)]}{h}+\lim_{h\to 0}\frac{g(a)[f(a+h)-f(a)]}{h}\\
&= \lim_{h\to 0}f(a+h)\cdot\lim_{h\to 0}\frac{g(a+h)-g(a)}{h}+\lim_{h\to 0}\frac{f(a+h)-f(a)}{h}\cdot\lim_{h\to 0}g(a)\\
&= f(a)\cdot g'(a)+f'(a)\cdot g(a)
\end{aligned}$$

as claimed, where $\lim_{h\to 0} f(a+h) = f(a)$ since f is differentiable at a, which implies it is also continuous at a. □

Theorem 2.6 (Power Rule) *IF $f(x) = x^n$ for some natural number n, then*

$$f'(a) = na^{n-1}$$

for all a.

Proof For the proof we will proceed by induction on n. For $n = 1$ we have shown that $f'(a) = 1 = 1 \cdot a^0$, satisfying the base case. Assume that there exists $k \in \mathbb{N}$ such that if $n = k$, $f'(a) = ka^{k-1}$. Then, for the case of $n = k + 1$ we may write $g(x) = x \cdot x^k = I(x) \cdot f(x)$. Hence, by the product rule we have that for all a

$$\begin{aligned} g'(a) &= (I \cdot f)'(a) \\ &= I'(a) \cdot f(a) + I(a) \cdot f'(a) \\ &= 1 \cdot a^k + a \cdot ka^{k-1} \\ &= (k+1)a^k \end{aligned}$$

as claimed. Hence, by mathematical induction we conclude that if $f(x) = x^n$ for $n \in \mathbb{N}$, then $f'(a) = na^{n-1}$ for all $a \in \mathbb{R}$. □

Theorem 2.7 (Derivative of a Quotient) *If g is differentiable at a, and $g(a) \neq 0$, then $1/g$ is differentiable at a and*

$$\left(\frac{1}{g}\right)'(a) = \frac{-g'(a)}{|g(a)|^2}$$

Proof Note that since g is differentiable at a it is continuous at a. Moreover, since $g(a) \neq 0$, there exists $\delta > 0$ such that $g(a+h) \neq 0$ for $|h| < \delta$. Therefore, $(1/g)(a+h)$ is well defined for small enough h, and we can write

$$\begin{aligned} \lim_{h\to 0} \frac{(1/g)(a+h) - (1/g)(a)}{h} &= \lim_{h\to 0} \frac{1/g(a+h) - 1/g(a)}{h} \\ &= \lim_{h\to 0} \frac{g(a) - g(a+h)}{h[g(a) \cdot g(a+h)]} \\ &= \lim_{h\to 0} \frac{-[g(a+h) - g(a)]}{h} \cdot \lim_{h\to 0} \frac{1}{g(a) \cdot g(a+h)} \\ &= -g'(a) \cdot \frac{1}{|g(a)|^2} \end{aligned}$$

where $\lim\limits_{h\to 0} 1/g(a+h) = 1/g(a)$ by continuity of g. □

Theorem 2.8 (Quotient Rule) *If f and g are differentiable at a and $g(a) \neq 0$, then f/g is differentiable at a and*

$$(f/g)'(a) = \frac{g(a) \cdot f'(a) - f(a) \cdot g'(a)}{|g(a)|^2}$$

Proof Note that $f/g = f \cdot (1/g)$, so we have

$$\begin{aligned} (f/g)'(a) &= (f \cdot 1/g)'(a) \\ &= f'(a) \cdot (1/g)(a) + f(a) \cdot (1/g)'(a) && \text{(Product Rule)} \\ &= \frac{f'(a)}{g(a)} - \frac{f(a)g'(a)}{|g(a)|^2} && \text{(Quotient Derivative)} \\ &= \frac{f'(a)g(a) - f(a)g'(a)}{|g(a)|^2} \end{aligned}$$

as claimed. □

Theorem 2.9 (General Product Rule) *If $f_1, f_2, ..., f_n$ are differentiable at a for some $n \in \mathbb{N}$, then $f_1 \cdot f_2 \cdot ... \cdot f_n$ is differentiable at a and*

$$(f_1 \cdot ... \cdot f_n)'(a) = \sum_{i=1}^{n} f_1(a) \cdot ... \cdot f_{i-1}(a) \cdot f_i'(a) \cdot f_{i+1}(a) \cdot ... \cdot f_n(a)$$

Proof We proceed by induction on n. If $n = 1$ then $f_1'(a) = f_1'(a)$, so the base case holds. Now, suppose the claim is true for some $k \in \mathbb{N}$. Then it follows that if $n = k + 1$

$$\begin{aligned}(f_1 \cdot ... \cdot f_k \cdot f_{k+1})'(a) &= (f_1 \cdot ... \cdot f_k)'(a) f_{k+1}(a) + (f_1 \cdot ... \cdot f_k)(a) f_{k+1}'(a) && \text{(Product Rule)}\\ &= \left[\sum_{i=1}^{k} f_1(a) \cdot ... \cdot f_{i-1}(a) \cdot f_i'(a) \cdot f_{i+1}(a) \cdot ... \cdot f_k(a)\right] f_{k+1}(a)\\ &+ f_1(a) \cdot ... \cdot f_k(a) \cdot f_{k+1}'(a) && \text{(by Induction Hypothesis)}\\ &= \sum_{i=1}^{k+1} f_1(a) \cdot ... \cdot f_{i-1}(a) \cdot f_i'(a) \cdot f_{i+1}(a) \cdot ... \cdot f_{k+1}(a)\end{aligned}$$

as desired. Thus by mathematical induction we conclude that the formula holds for all $n \in \mathbb{N}$. □

Theorem 2.10 (Chain Rule) *If g is differentiable at a and f is differentiable at $g(a)$, then $f \circ g$ is differentiable at a and*

$$(f \circ g)'(a) = f'(g(a)) \cdot g'(a)$$

Proof Define a function ϕ as follows:

$$\phi(h) = \begin{cases} \frac{f(g(a+h)) - f(g(a))}{g(a+h) - g(a)}, & \text{if } g(a+h) - g(a) \neq 0 \\ f'(g(a)), & \text{if } g(a+h) - g(a) = 0 \end{cases} \tag{2.2.1}$$

Note that by differentiability of g at a, g is continuous at a as well so as $h \to 0$, $g(a+h) - g(a) \to 0$, so if $g(a+h) - g(a)$ is not zero, then $\phi(h)$ will approach $f'(g(a))$ as h goes to zero. If it is zero then $\phi(h)$ is exactly $f'(g(a))$. Note that as f is differentiable at $g(a)$ we have

$$\lim_{k \to 0} \frac{f(g(a) + k) - f(g(a))}{k} = f'(g(a))$$

Thus, if $\epsilon > 0$ there is some number $\delta' > 0$ such that, for all k,

$$(1)\ \text{ if } 0 < |k| < \delta', \text{ then } \left| \frac{f(g(a) + k) - f(g(a))}{k} - f'(g(a)) \right| < \epsilon$$

Now, g is differentiable at a, hence continuous at a, so there is $\delta > 0$ such that for all h,

$$(2)\ \text{ if } |h| < \delta, \text{ then } |g(a+h) - g(a)| < \delta'$$

Consider now any h with $|h| < \delta$. If $k = g(a+h) - g(a) \neq 0$, then

$$\phi(h) = \frac{f(g(a+h)) - f(g(a))}{g(a+h) - g(a)} = \frac{f(g(a)+k) - f(g(a))}{k}$$

it follows from (2) that $|k| < \delta'$, and hence from (1) that

$$|\phi(h) - f'(g(a))| < \epsilon$$

On the other hand, if $g(a+h) - g(a) = 0$, then $\phi(h) = f'(g(a))$, so it is surely true that

$$|\phi(h) - f'(g9a))| < \epsilon$$

We therefore have proved that

$$\lim_{h \to 0} \phi(h) = f'(g(a))$$

so ϕ is continuous at 0. If $h \neq 0$, then we have

$$\frac{f(g(a+h)) - f(g(a))}{h} = \phi(h) \cdot \frac{g(a+h) - g(a)}{h}$$

even if $g(a+h) - g(a) = 0$. Therefore, we have that

$$\begin{aligned}(f \circ g)'(a) &= \lim_{h \to 0} \frac{f(g(a+h)) - f(g(a))}{h} \\ &= \lim_{h \to 0} \phi(h) \cdot \lim_{h \to 0} \frac{g(a+h) - g(a)}{h} \\ &= f'(g(a)) \cdot g'(a)\end{aligned}$$

by continuity of $\phi(h)$ at 0. □

2.3 Applications of Derivatives

Definition 2.4 (Extrema) Let f be a function and A a set of numbers contained in the domain of f. A point $x \in A$ is ***maximum point*** for f on A if

$$f(x) \geq f(y) \forall y \in A \tag{2.3.1}$$

The number $f(x)$ is itself called the ***maximum value*** of f on A.

A point $x \in A$ is a ***minimum point*** for f on A if

$$f(x) \leq f(y) \forall y \in A \tag{2.3.2}$$

The number $f(x)$ is itself called the ***minimum value*** of f on A.

Theorem 2.11 *Let f be any function defined on (a, b). If x is an extremum point for f on (a, b), and f is differentiable at x, then $f'(x) = 0$.*

Proof Consider the case where f has a maximum at x. If h is any number such that $x + h \in (a, b)$, then

$$f(x) \geq f(x + h)$$

since f has a maximum on (a, b) at x. This implies that

$$f(x + h) - f(x) \leq 0$$

Thus, if $h > 0$ we have that

$$\frac{f(x + h) - f(x)}{h} \leq 0$$

and consequently

$$\lim_{h \to 0^+} \frac{f(x + h) - f(x)}{h} \leq 0$$

as otherwise $\frac{f(x+h)-f(x)}{h} > 0$ for some h, contradicting our initial assumptions. Similarly, if $h < 0$ we have

$$\frac{f(x + h) - f(x)}{h} \geq 0$$

so

$$\lim_{h \to 0^-} \frac{f(x + h) - f(x)}{h} \geq 0$$

By hypothesis f is differentiable at x, so these two limits must be equal, so in fact $f'(x) \leq 0$ and $f'(x) \geq 0$. Thus, $f'(x) = 0$.

On the other hand, suppose f has a minimum at x. Then $-f$ has a maximum at x. Indeed, for all $y \in (a, b)$ we have $f(y) \geq f(x)$, so $-f(y) \leq -f(x)$. Then, from our above argument and the differentiability of f at x, we have $-f'(x) = 0$, which implies that $f'(x) = 0$. □

Definition 2.5 (Local Extrema) Let f be a function, and A a set of numbers contained in the domain of f. A point x in A is a ***local maximum [minimum] point*** for f on A if there is some $\delta > 0$ such that x is a maximum [minimum] point for f on $A \cap (x - \delta, x + \delta)$.

Definition 2.6 A ***critical point*** of a function f is a number x such that

$$f'(x) = 0 \tag{2.3.3}$$

The number $f(x)$ itself is called a ***critical value*** of f.

Remark 2.2 Give a function continuous f, if x is an extrumum of f on $[a, b]$, then one of the following must be satisfied:

1. x is a critical point of f in $[a, b]$
2. $x = a$ or $x = b$ so x is an endpoint of $[a, b]$
3. x is a point in $[a, b]$ such that f is not differentiable at x

Theorem 2.12 (Rolle's Theorem) *If f is continuous on $[a, b]$ and differentiable on (a, b), and $f(a) = f(b)$, then there is a number $x \in (a, b)$ such that $f'(x) = 0$.*

Proof It follows from continuity of f on $[a, b]$ that f has a maximum or minimum value on $[a, b]$ (by the Extreme Value Theorem).

Suppose first that the maximum value occurs at a point $x \in (a, b0$. Then $f'(x) = 0$ by Theorem 2.11. On the other hand suppose that the minimum value of f occurs at some point x in (a, b). Then, again, $f'(x) = 0$ by Theorem 2.11.

Finally, suppose the maximum and minimum values both occur at the end points. Since $f(a) = f(b)$, the maximum and minimum values of f are equal, so f is a constant function, and for a constant function we can choose any $x \in (a, b)$ and have $f'(x) = 0$, completing the proof. □

Theorem 2.13 (The Mean Value Theorem) *If f is continuous on $[a, b]$ and differentiable on (a, b), then there is a number $x \in (a, b)$ such that*

$$f'(x) = \frac{f(b) - f(a)}{b - a} \tag{2.3.4}$$

Proof Let

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

Evidently, h is continuous on $[a, b]$ and differentiable on (a, b) as it is the sum of correspondingly continuous and differentiable functions. Moreover,

$$\begin{aligned} h(a) &= f(a) \\ h(b) &= f(b) - \left[\frac{f(b) - f(a)}{b - a}\right](b - a) \\ &= f(a) \end{aligned}$$

Consequently, we may apply 2.12 to h and conclude that there exists $x \in (a, b)$ such that

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

so that

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

as desired. □

Corollary 2.1 *If f is defined on an interval and $f'(x) = 0$ for all x in the interval, then f is constant on the interval.*

Proof Let a and b be any two points in the interval with $a \neq b$. Then there is some $x \in (a, b)$ such that

$$0 = f'(x) = \frac{f(b) - f(a)}{b - a}$$

so $f(b) - f(a) = 0$ and consequently $f(a) = f(b)$. Thus the value of f at any two points in the interval is the same, so f is constant on the interval. □

Corollary 2.2 *If f and g are defined on the same interval, and $f'(x) = g'(x)$ for all x in the interval, then there is come number c such that $f = g + c$.*

Proof For all x in the interval we have $(f - g)'(x) = f'(x) - g'(x) = 0$, so by the previous corollary there is some number c such that $f - g = c$. □

Definition 2.7 A function is ***increasing*** on an interval I if $f(a) < f(b)$ whenever $a, b \in I$ with $a < b$. The function f is ***decreasing*** on an interval I if $f(a) > f(b)$ for all $a, b \in I$ with $a < b$.

Corollary 2.3 *If $f'(x) > 0$ for all x in an interval, then f is increasing on the interval; if $f'(x) < 0$ for all x in the interval, then f is decreasing on the interval.*

Proof Consider the case where $f'(x) > 0$. Let $a, b \in I$ with $a < b$. Then by 2.13 there exists $x \in (a, b)$ such that

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

But, $f'(x) > 0$ for all $x \in (a, b)$, so

$$\frac{f(b) - f(a)}{b - a} > 0$$

Since $b - a > 0$ we conclude that $f(b) > f(a)$ so f is increasing.

Next, consider the case for $f'(x) < 0$. Then $-f'(x) > 0$ for all $x \in I$, so by the first case we have that for all $a, b \in I$ with $a < b$, $-f(a) < -f(b)$. Multiplying both sides by -1 we have that $f(a) > f(b)$ for all $a, b \in I$ such that $a < b$, so f is decreasing, as desired. □

Theorem 2.14 (Second Derivative Test) *Suppose $f'(a) = 0$. If $f''(a) > 0$, then f has a local minimum at a; if $f''(a) < 0$ then f has a local maximum at a.*

Proof By definition

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

Since $f'(a) = 0$ by assumption, we can write

$$f''(a) = \lim_{h \to 0} \frac{f'(a + h)}{h}$$

Suppose now that $f''(a) > 0$. Then there exists $\delta > 0$ such that if $|h| < \delta$ $f'(a + h)/h > 0$. Thus, for $|h| < \delta$, if $h < 0$ we must have $f'(a + h) < 0$ and if $h > 0$ we must have $f'(a + h) > 0$. This means by our previous corollary that f is increasing in the interval $(a, a + \delta)$, and decreasing in $(a - \delta, a)$. Thus, as $f'(a) = 0$, $f(a)$ must be a local minimum.

If $f''(a) < 0$, then $-f''(a) > 0$ so $-f(a)$ is must be a local minimum. That is, there exists $\delta > 0$ such that if $x \in (a - \delta, a + \delta)$, then $-f(x) \geq -f(a)$. Hence, it follows that $f(x) \leq f(a)$ for all $x \in (a - \delta, a + \delta)$, so $f(a)$ is a local maximum of f. □

Theorem 2.15 *Suppose $f''(a)$ exists. If f has a local minimum at a, then $f''(a) \geq 0$; if f has a local maximum at a, then $f''(a) \leq 0$.*

Proof Suppose f has a local minimum at a. If $f''(a) < 0$ then by our previous result f would have a local maximum at a. But, this implies that f would be constant in some interval containing a, so that $f''(a) = 0$, which is a contradiction. Thus, we must have that $f''(a) \geq 0$.

The case for a local maximum is analogous. □

Theorem 2.16 *Suppose that f is continuous at a, and that $f'(x)$ exists for all x in some interval containing a, except perhaps for $x = a$. Suppose, moreover, that $\lim_{x \to a} f'(x)$ exists. Then $f'(a)$ also exists and*

$$f'(a) = \lim_{x \to a} f'(x) \tag{2.3.5}$$

Proof By definition

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

For sufficiently small $h > 0$ the function f will be continuous on $[a, a+h]$, and differentiable on $(a, a+h)$, by assumption (similarly for sufficiently small $h < 0$). By 2.13 there is a number $\alpha_h \in (a, a+h)$ such that

$$\frac{f(a+h) - f(a)}{h} = f'(\alpha_h)$$

Now, α_h approaches a as h approaches 0, because α_h is in $(a, a+h)$. Since $\lim_{x \to a} f'(x)$ exists, it follows that

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

For this last equality write $\lim_{x \to a} f'(x) = L \in \mathbb{R}$. Fix $\epsilon > 0$. Then there exists $\delta > 0$ such that for all $x \in (a - \delta, a + \delta)$, $|f'(x) - L| < \epsilon$. It follows that for $|h| < \delta$, if $h > 0$ and $\alpha_h \in (a, a+h) \subset (a - \delta, a + \delta)$ we have $|f'(\alpha_h) - L| < \epsilon$ and if $h < 0$ and $\alpha_h \in (a+h, a) \subset (a - \delta, a + \delta)$, then $|f'(\alpha_h) - L| < \epsilon$. Thus, by definition we have that $\lim_{h \to 0^+} f'(\alpha_h) = \lim_{h \to 0^-} f'(\alpha_h) = L$, so in particular $\lim_{h \to 0} f'(\alpha_h) = L = \lim_{x \to a} f'(x)$, completing the proof. □

Theorem 2.17 (The Cauchy Mean Value Theorem) *If f and g are continuous on $[a, b]$ and differentiable on (a, b), then there is a number $x \in (a, b)$ such that*

$$[f(b) - f(a)]g'(x) = [g(b) - g(a)]f'(x) \tag{2.3.6}$$

Proof Let

$$h(x) = f(x)[g(b) - g(a)] - g(x)[f(b) - f(a)]$$

Then h is continuous on $[a, b]$, differentiable on (a, b), and

$$h(a) = f(a)g(b) - g(a)f(b) = h(b)$$

It follows by 2.12 that $h'(x) = 0$ for some $x \in (a, b)$, which implies that

$$0 = h'(x) = f'(x)[g(b) - g(a)] - g'(x)[f(b) - f(a)]$$

completing the proof. □

Theorem 2.18 (L'Hôpital's Rule) *Suppose that*

$$\lim_{x\to a} f(x) = 0 \textit{ and } \lim_{x\to a} g(x) = 0 \tag{2.3.7}$$

and suppose also that $\lim\limits_{x\to a} f'(x)/g'(x)$ *exists. Then* $\lim\limits_{x\to a} f(x)/g(x)$ *exists, and*

$$\lim_{x\to a} \frac{f(x)}{g(x)} = \lim_{x\to a} \frac{f'(x)}{g'(x)} \tag{2.3.8}$$

Proof The hypothesis that $\lim\limits_{x\to a} f'(x)/g'(x)$ exists contains two implicit assumptions:

1. there is an interval $(a-\delta, a+\delta)$ such that $f'(x)$ and $g'(x)$ exist for all $x \in (a-\delta, a+\delta)$, except, perhaps, $x = a$,
2. in this interval $g'(x) \neq 0$, with the possible exception of $x = a$

If we define $f(a) = g(a) = 0$, then f and g are continuous at a. If $x \in (a, a+\delta)$, then 2.13 and 2.17 apply to f and g on $[a, x]$ (a similar statement holds for $x \in (a-\delta, a)$). First, applying the 2.13 to g, we see that $g(x) \neq 0$, for if $g(x) = 0$ there would exist $x_1 \in (a, x)$ with $g'(x_1) = 0$, contradicting 2.. Now, applying 2.17 to f and g, we see that there is a number $\alpha_x \in (a, x)$ such that

$$[f(x) - 0]g'(\alpha_x) = [g(x) - 0]f'(\alpha_x)$$

or

$$\frac{f(x)}{g(x)} = \frac{f'(\alpha_x)}{g'(\alpha_x)}$$

Now, let $\lim_{y\to a} f'(y)/g'(y) = L \in \mathbb{R}$. Fix $\epsilon > 0$. Then there exists $\delta' > 0$ such that if $y \in (a-\delta', a+\delta')$ then $|f'(y)/g'(y) - L| < \epsilon$. Then, for $x \in (a, a+\delta)$ (or $x \in (a-\delta, a)$) we have $(a, x) \subset (a-\delta, a+\delta)$ (or $(x, a) \subset (a-\delta, a+\delta)$. Thus, for $|x-a| < \delta$ we have $\alpha_x \in (a, x) \subset (a-\delta, a+\delta)$ (or $\alpha_x \in (x, a) \subset (a-\delta, a+\delta)$), so $|f'(\alpha_x)/g'(\alpha_x) - L| < \epsilon$. Therefore, we conclude that

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

so in particular

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

completing the proof. □

2.3.1 Convexity

Definition 2.8 A function f is ***convex*** on an interval I, if for all $a, b \in I$, the line segment joining $(a, f(a))$ and $(b, f(b))$ lies above the graph of f.

This is equivalent to stating that for all $x \in (a, b)$,

$$\frac{f(x)-f(a)}{x-a} < \frac{f(b)-f(a)}{b-a} \tag{2.3.9}$$

Definition 2.9 A function f is ***concave*** on an interval I, if for all $a, b \in I$, the line segment joining $(a, f(a))$ and $(b, f(b))$ lies below the graph of f.

This is equivalent to stating that for all $x \in (a, b)$,

$$\frac{f(x)-f(a)}{x-a} > \frac{f(b)-f(a)}{b-a} \tag{2.3.10}$$

Theorem 2.19 *Let f be convex. If f is differentiable at a, then the graph of f lies above the tangent line through $(a, f(a))$, except at $(a, f(a))$ itself. If $a < b$ and f is differentiable at a and b, then $f'(a) < f'(b)$.*

Proof If $0 < h_1 < h_2$, then $a < a + h_1 < a + h_2$, and applying f's convexity we have that

$$\frac{f(a+h_1)-f(a)}{h_1} < \frac{f(a+h_2)-f(a)}{h_2}$$

This implies that the values of $[f(a+h)-f(a)]/h$ decrease as $h \longrightarrow 0^+$. Consequently,

$$f'(a) < \frac{f(a+h)-f(a)}{h}, h > 0$$

In fact, $f'(a)$ is the infimum of these numbers. Similarly, for h negative, if $h_2 < h_1 < 0$, then

$$\frac{f(a+h_1)-f(a)}{h_1} > \frac{f(a+h_2)-f(a)}{h_2}$$

This shows that the slope of the tangent line is greater that $[f(a+h)-f(a)]/h$ for $h < 0$. In fact, $f'(a)$ is the supremum of all these numbers, so $f(a+h)$ lies above the tangent line if $h < 0$. This satisfies the first part of the theorem. Now, suppose $a < b$. Then we have that

$$f'(a) < \frac{f(a+(b-a))-f(a)}{b-a} = \frac{f(b)-f(a)}{b-a}$$

since $b - a > 0$ and

$$f'(b) > \frac{f(b+(a-b))-f(b)}{a-b} = \frac{f(a)-f(b)}{a-b} = \frac{f(b)-f(a)}{b-a}$$

since $a - b < 0$. Combining these inequalities we obtain $f'(a) < f'(b)$, as desired. □

Lemma 2.1 *Suppose f is differentiable and f' is increasing. If $a < b$ and $f(a) = f(b)$, then $f(x) < f(a) = f(b)$ for $a < x < b$.*

Proof Suppose towards a contradiction that $f(x) \geq f(a) = f(b)$ for some $x \in (a, b)$. Then the maximum of f on $[a, b]$ occurs at some point $x_0 \in (a, b)$ with $f(x_0) \geq f(a)$ and, of course, $f'(x_0) = 0$. On the other hand, applying 2.13 to the interval $[a, x_0]$, we find that there is x_1 with $a < x_1 < x_0$ and

$$f'(x_1) = \frac{f(x_0) - f(a)}{x_0 - a} \geq 0$$

contradicting the fact that f' is increasing (since $f'(x_0) = 0$ and $x_1 < x_0$). □

Theorem 2.20 *If f is differentiable and f' is increasing, then f is convex.*

Proof Let $a < b$. Define g by

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

It is easy to see that g' is also increasing; moreover, $g(a) = g(b) = f(a)$. Applying the lemma to g we conclude that

$$a < x < b \implies g(x) < f(a)$$

In other words, if $a < x < b$, then

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

or

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

Hence, f is convex. □

Theorem 2.21 *If f is differentiable and the graph of f lies above each tangent line except at the point of contact, then f is convex.*

Proof Let $a < b$. Since the tangent lien at $(a, f(a))$ is the graph of the function

$$g(x) = f'(a)(x - a) + f(a)$$

and since $(b, f(b))$ lies above the tangent line, we have

$$(1) \quad f(b) > f'(a)(b - a) + f(a)$$

Similarly, since the tangent line at $(b, f(b))$ is the graph of $h(x) = f'(b)(x - b) + f(b)$, and $(a, f(a))$ lies above the tangent line at $(b, f(b))$, we have

$$(2) \quad f(a) > f'(b)(a - b) + f(b)$$

It follows from (1) and (2) that $f'(a) < f'(b)$. Then, from our previous theorem we have that f is convex. □

2.4 Inverse Functions

Definition 2.10 For any function f. the ***inverse image*** of f, denoted by f^{-1}, is the set of all pairs (a, b) such that $(b, a) \in f$.

Remark 2.3 f^{-1} is a function if and only if f is one-to-one.

Theorem 2.22 *If f is increasing (decreasing) on an interval I, then f is injective on I so f^{-1} is a function and in fact f^{-1} is increasing (decreasing).*

Proof Consider the case that f is increasing. Then suppose $a, b \in I$ with $a \neq b$. Without loss of generality suppose $a < b$. Then since f is increasing $f(a) < f(b)$ so in particular $f(a) \neq f(b)$. Therefore, f is injective as claimed, so f^{-1} is a well-defined function on I. Now, consider $a' < b'$ in $f(I) = I'$. Then there exist $x, y \in I$ such that $f(x) = a'$ and $f(y) = b'$, so in particular $f^{-1}(a') = x$ and $f^{-1}(b') = y$. Since f is increasing and $f(x) = a' < b' = f(y)$ we must have that $x < y$. Thus, $f^{-1}(a') = x < y = f^{-1}(b')$, so f^{-1} is increasing as claimed.

Consider the case that f is decreasing. Then $-f$ is increasing so it is injective and $-f^{-1}$ is increasing by the first case. Hence, we have that f^{-1} is decreasing as desired. □

Theorem 2.23 *If f is continuous and one-to-one on an interval I, then f is either increasing or decreasing on I.*

Proof We proceed in three steps:

(1) If $a < b < c$ are three points in I, then I claim either $f(a) < f(b) < f(c)$ or $f(a) > f(b) > f(c)$. Indeed, suppose that $f(a) < f(c)$. If we have $f(b) < f(a)$, then the **??** applied to $[b, c]$ gives an $x \in (b, c)$ such that $f(x) = f(a)$, contradicting the fact that f is injective on $[a, c]$. Similarly, if $f(b) > f(c)$ we would find a contradiction, so $f(a) < f(b) < f(c)$. Similar argumentation leads to the result that $f(a) > f(b) > f(c)$ in the second case.

(2) If $a < b < c < d$ are four points in I, then I claim that either $f(a) < f(b) < f(c) < f(d)$ or $f(a) > f(b) > f(c) > f(d)$. Indeed we can apply (1) to $a < b < c$ and then to $b < c < d$.

(3) Take any $a < b$ in I, and suppose $f(a) < f(b)$. Then f is increasing, for if $c, d \in I$ are any two points, we can apply (2) to the collection $\{a, b, c, d\}$ after arranging them in increasing order. □

Theorem 2.24 *If f is continuous and one-to-one on an interval, then f^{-1} is also continuous.*

Proof Since f is continuous and injective on the interval, it is either increasing or decreasing. Consider the case that f is increasing. We must show that

$$\lim_{x \to b} f^{-1}(x) = f^{-1}(b)$$

for each b in the domain of f^{-1}. Such a number b is of the form $f(a)$ for some a in the domain of f. For any $\epsilon > 0$, we want to find a $\delta > 0$ such that for all x, if $x \in (f(a)-\delta, f(a)+\delta)$, then $|f^{-1}(x)-a| < \epsilon$, as $a = f^{-1}(b) = f^{-1}(f(a))$. Now, since $a - \epsilon < a < a + \epsilon$ we have that $f(a-\epsilon) < f(a) < f(a+\epsilon)$ since f is presumed increasing. Let $\delta = \min(f(a+\epsilon) - f(a), f(a) - f(a-\epsilon))$. Our choice of δ ensures that

$$f(a-\epsilon) \leq f(a) - \delta \ and \ f(a) + \delta \leq f(a+\epsilon)$$

Consequently, if

$$f(a) - \delta < x < f(a) + \delta$$

then

$$f(a-\epsilon) < x < f(a+\epsilon)$$

SInce f is increasing, f^{-1} is also increasing, and we obtain

$$f^{-1}(f(a-\epsilon)) < f^{-1}(x) < f^{-1}(f(a+\epsilon))$$

so $a-\epsilon < f^{-1}(x) < a+\epsilon$, which is precisely $|f^{-1}(x)-a| < \epsilon$, as desired. ☐

Theorem 2.25 *If f is a continuous one-to-one function defined on an interval I, and $f'(f^{-1}(a)) = 0$, then f^{-1} is not differentiable at a.*

Proof We have $f(f^{-1}(x)) = x$. If f^{-1} were differentiable at a, then the chain rule would imply that

$$f'(f^{-1}(a)) \cdot (f^{-1})'(a) = 1$$

hence

$$0 \cdot (f^{-1})'(a) = 1$$

which is impossible. ☐

Theorem 2.26 *Let f be a continuous one-to-one function defined on an interval I, and suppose that f is differentiable at $f^{-1}(b)$, with derivative $f'(f^{-1}(b)) \neq 0$. Then f^{-1} is differentiable at b, and*

$$(f^{-1})'(b) = \frac{1}{f'(f^{-1}(b))} \tag{2.4.1}$$

Proof Let $b = f(a)$. Then

$$\lim_{h\to 0} \frac{f^{-1}(b+h) - f^{-1}(b)}{h} = \lim_{h\to 0} \frac{f^{-1}(b+h) - a}{h}$$

Now, every number $b+h$ in the domain of f^{-1} can be written in the form $b+h = f(a+k)$ for a unique $k(h)$. Then

$$\begin{aligned}\lim_{h\to 0} \frac{f^{-1}(b+h) - a}{h} &= \lim_{h\to 0} \frac{f^{-1}(f(a+k(h))) - a}{f(a+k(h)) - b} \\ &= \lim_{h\to 0} \frac{k(h)}{f(a+k(h)) - f(a)}\end{aligned}$$

Since $b+h = f(a+k(h))$ we have $f^{-1}(b+h) = a+k(h)$, or $k(h) = f^{-1}(b+h) - f^{-1}(b)$. Now, since f is continuous on I, f^{-1} is also continuous on its domain, and in particular it is continuous at b. This means that $\lim_{h\to 0} k(h) = 0$, so $k(h)$ goes to zero as h goes to 0. Hence, as

$$\lim_{k\to 0} \frac{f(a+k) - f(a)}{k} = f'(a) = f'(f^{-1}(b)) \neq 0$$

this implies that f^{-1} is differentiable at b and

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

Appendix A: Alternative Differentiation Formulation

Definition 2.11 A function $f : (a, b) \subseteq \mathbb{R} \to \mathbb{R}$ (or $\mathbb{C}$) is said to be ***differentiable*** at $x \in (a, b)$ with derivative $f'(x)$ if the limit

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

exists.

If f is differentiable at $x \in (a, b)$, it is immediate that f is continuous at x. Indeed, let $\varepsilon > 0$. Then there exists $\delta > 0$ such that if $0 < |h| < \delta$, $\left|\frac{f(x+h)-f(x)}{h} - f'(x)\right| < \varepsilon$. Then let $\delta' = \min\{\delta, |f'(x)| + \varepsilon\}$. Then for $x + h \in B_{\delta'}(x)$, $h \neq 0$, we ahve

$$|f(x+h) - f(x)| = |h| \left|\frac{f(x+h) - f(x)}{h}\right| < \delta'(|f'(x)| + \varepsilon) \leq \varepsilon$$

Now a useful equivalent condition to differentiability at $x \in (a, b)$ is the existence of a function $r(x, h)$ for h close to 0 such that $f(x + h) = f(x) + Dh + r(x, h)$ such that $\frac{r(x,h)}{h} \to 0$ as $h \to 0$, and then $D = f'(x)$. Indeed if f is differentiable set $r(x, h) = f(x + h) - (f(x) + f'(x)h)$. On the other hand, if such an $r(x, h)$ exists, then

$$\frac{f(x+h) - f(x)}{h} = \frac{r(x,h)}{h} + D \to D$$

so the limit exists and is D.

Definition 2.12 We say that f is differentiable on (a, b) if it is differentiable at all $x \in (a, b)$.

Proposition 2.1 *Let f and g be differentiable at x. Then*

- *$f \pm g$ is differentiable at x with $(f \pm g)'(x) = f'(x) \pm g'(x)$ (additivity)*
- *fg is differentiable at x with $(fg)'(x) = f'(x)g(x) + f(x)g'(x)$ (Liebnitz's rule)*
- *If $g(y) \neq 0$ in a neighborhood of x, then $(1/g)$ is differentiable at x with $(1/g)'(x) = -\frac{g'(x)}{g(x)^2}$*
- *For all $c \in \mathbb{R}$ cf is differentiable at x with $(cf)'(x) = cf'(x)$*

Proof The first bullet is by linearity of limits. The second bullet follows from the computation

$$\frac{(fg)(x+h) - (fg)(x)}{h} = \frac{(f(x+h) - f(x))g(x+h)}{h} + \frac{f(x)(g(x+h) - g(x)}{h}$$

and the fact that g is continuous at x and the product of limits is the limit of the product, if the limits involved all exist. The quotient also follows by a similar computation

$$\frac{\frac{1}{g(x+h)} - \frac{1}{g(x)}}{h} = \frac{-(g(x+h) - g(x))}{hg(x+h)g(x)}$$

and the same properties of limits and g at x. Finally, bullet follows simply by the calculation

$$\frac{cf(x+h)-cf(x)}{h}=c\frac{f(x+h)-f(x)}{h}$$

Some basic derivatives from the definition are $f'(x)=0$ if $f(x)=c\in\mathbb{R}$ is a constant, and later we shall show the converse also holds. Additionally, $id'(x)=1$, for $id(x)=x$. Then by induction $\frac{d}{dx}x^n=nx^{n-1}$ for all $n=0,1,2,3,....$

Proposition 2.2 *If $f:(a,b)\to(\alpha,\beta)$ is differentiable at $x\in(a,b)$ and $g:(\alpha,\beta)\to\mathbb{R}$ is differentiable at $f(x)\in(\alpha,\beta)$, then $g\circ f$ is differentiable at x with*

$$(g\circ f)'(x)=g'(f(x))f'(x)$$

Proof We use the remainder definition of differentiability to prove the claim. As f is differentiable at x we have $r_f(x,h)$ such that $f(x+h)=f(x)+f'(x)h+r_f(x,h)$, and as g is differentiable at $f(x)$ we have $r_g(x,h)$ such that $g(f(x)+h)=g(f(x))+g'(f(x))h+r_g(f(x),h)$. Then

$$\begin{aligned} g\circ f(x+h) &= g(f(x)+f'(x)h+r_f(x,h)) \\ &= g(f(x))+g'(f(x))(f'(x)h+r_f(x,h))+r_g(f(x),\tilde{h}) \quad (\tilde{h}:=f'(x)h+r_f(x,h)) \end{aligned}$$

Let $r_{g\circ f}(x,h)=g'(f(x))r_f(x,h)+r_g(f(x),\tilde{h})$. Then we have $\lim\limits_{h\to0}g'(f(x))\frac{r_f(x,h)}{h}=0$ and $\lim\limits_{h\to0}\tilde{h}=0$, so

$$\lim_{h\to0}\frac{r_g(f(x),\tilde{h})}{h}=\lim_{h\to0}\frac{r_g(f(x),\tilde{h})}{\tilde{h}}\frac{\tilde{h}}{h}=0\cdot f'(x)=0$$

Thus, we have that $g\circ f$ is differentiable at x with $(g\circ f)'(x)=g'(f(x))f'(x)$, as desired. □

Proposition 2.3 *If $f:(a,b)\to\mathbb{R}$ and $x\in(a,b)$ such that*

$$f(x)\geq f(y)(or\ f(x)\leq f(y))\forall y\in(a,b)$$

then if f is differentiable at x it follows that

$$f'(x)=0$$

Proof First, note that

$$f'(x)=\lim_{h\to0}\frac{f(x+h)-f(x)}{h}=\lim_{h\to0^+}\frac{f(x+h)-f(x)}{h}=\lim_{h\to0^-}\frac{f(x+h)-f(x)}{h}$$

But for h such that $x+h\in(a,b)$, and $h>0$, we have

$$\frac{f(x+h)-f(x)}{h}\leq 0(\text{respectively }\geq 0)$$

while if $h<0$,

$$\frac{f(x+h)-f(x)}{h}\geq 0(\text{respectively }\leq 0)$$

Thus, by order properties of limits $f'(x) \leq 0$ and $f'(x) \geq 0$, so $f'(x) = 0$. □

Next we explore a funcdamental result for differentiable functions on an interval, which we will use to prove the fundamental theorem of calculus.

Theorem 2.27 (Mean Value Theorem) *Suppose f is continuous on $[a, b]$ and is differentiable on (a, b). Then there exists $\xi \in (a, b)$ such that*

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

Proof Consider $f(x) = f(x) - (x - a)\frac{f(b)-f(a)}{b-a}$. Then g is continuous on $[a, b]$, differentiable on (a, b), and $g(a) = f(a)$ while $g(b) = f(a)$ as well, so $\frac{g(b)-g(a)}{b-a} = 0$. Since g is continuous on the compact set $[a, b]$, g attains a maximum on $[a, b]$. If it occurs at $\xi \in (a, b)$, then $g'(\xi) = 0$. On the other hand, if the maximum occurs at a or b, then since $g(a) = g(b) = \max_{[a,b]} g$, g must attain its minimum in (a, b). Thus, there exists $\zeta \in (a, b)$ such that $g(\zeta)$ is a minimum and $g'(\zeta) = 0$. Thus in either case we have a $\xi \in (a, b)$ such that $f'(\xi) - \frac{f(b)-f(a)}{b-a} = 0$, so $f'(\xi) = \frac{f(b)-f(a)}{b-a}$ as desired.□

Theorem 2.28 (Inverse Function Theorem) *Suppose f is continuous on $[a, b]$ and differentiable on (a, b), and there exists $\gamma_0, \gamma_1 \in \mathbb{R}$ such that $0 < \gamma_0 \leq f'(x) \leq \gamma_1 < \infty$ (or $-\infty < \gamma_0 \leq f'(x) \leq \gamma_1 < 0$) for all $x \in (a, b)$. Let $\alpha = f(a)$ and $\beta = f(b)$. Then there exists an inverse function $g : [\alpha, \beta] \to [a, b]$ (or $g : [\beta, \alpha] \to [a, b]$) which is continuous on $[\alpha, \beta]$ and differentiable on (α, β), with derivative*

$$(f^{-1})'(y) = g'(y) = \frac{1}{f'(g(y))} = \frac{1}{f'(f^{-1}(y))}$$

for all $y \in (\alpha, \beta)$.

Proof Let $x_1, x_2 \in [a, b]$, with $a \leq x_1 < x_2 \leq b$. By the mean value theorem there exists $\xi \in (x_1, x_2)$ such that

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

This implies

$$0 < \gamma_0 \leq \frac{f(x_2) - f(x_1)}{x_2 - x_2} \leq \gamma_1 < \infty$$

so

$$0 < \gamma_0(x_2 - x_1) \leq f(x_2) - f(x_1) \leq \gamma_1(x_2 - x_1)$$

for all $a \leq x_1 < x_2 \leq b$. This implies that f is strictly increasing, and so injective. Further, $f(x) \in [\alpha, \beta]$ for all $x \in [a, b]$ since $\alpha = f(a)$ and $\beta = f(b)$. By the intermediate value theorem it follows that f is surjective onto $[\alpha, \beta]$. Since $f : [a, b] \to [\alpha, \beta]$ is a continuous bijection on a compact set, f is a homeomorphism and $f^{-1} : [\alpha, \beta] \to [a, b]$ is also a homeomorphism. As f is differentiable on (a, b), there exists $r_f(x, h)$ satisfying certain properties discussed previously. Then for $y \in (\alpha, \beta)$, there exists $x \in (a, b)$ such that $f^{-1}(y) = x$, so $f(x) = y$. Then

$$\begin{aligned} f^{-1}(y) + h = x + h = f^{-1}(f(x + h)) &= f^{-1}(f(x) + f'(x)h + r_f(x, h)) \\ &= f^{-1}(y + f'(x)h + r_f(x, h)) \end{aligned}$$

Let $\tilde{h} = f'(x)h + r_f(x, h)$, so $h = \frac{\tilde{h}}{f'(x)} - \frac{r_f(x,h)}{f'(x)}$. Then

$$f^{-1}(y + \tilde{h}) = f^{-1}(y) + \frac{1}{f'(x)}\tilde{h} - \frac{r_f(x, h)}{f'(x)}$$

Recall $r_f(x, h) = f(x + h) - f(x) - f'(x)h$. Then let $r_g(x, \tilde{h}) = -\frac{[f(x+h)-f(x)-f'(x)h]}{f'(x)}$. Note $\gamma_0 h + r_f(x, h) \leq \tilde{h} \leq \gamma_1 h + r_f(x, h)$, so

$$\gamma_0 + \frac{r_f(x, h)}{h} \leq \frac{\tilde{h}}{h} \leq \gamma_1 + \frac{r_f(x, h)}{h}$$

which implies $\frac{\tilde{h}}{h}$ and $\frac{h}{\tilde{h}}$ are bounded. Thus

$$\frac{r_g(x, \tilde{h})}{\tilde{h}} = -\frac{1}{f'(x)}\frac{h}{\tilde{h}}\frac{r_f(x, h)}{h}$$

which goes to 0 as $\tilde{h}$ goes to 0, since $\tilde{h} \to 0$ implies $h \to 0$. Thus, $g = f^{-1}$ is differentiable at $y = f(x)$, and

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

The result for a negative derivative follows by replacing f with $-f$.

Note that if f is differentiable in (a, b) then $f'(x)$ is a function on (a, b). If $f'(x)$ is differentiable at x_0 we may write

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

and in general $f^{(k+1)}(x) = \frac{d}{dx}f^{(k)}(x)$, for all $k \in \mathbb{N} \cup \{0\}$.

Proposition 2.4 *If f is differentiable on (a, b) and $x_0 \in (a, b)$, $f'(x_0) = 0$ but $f''(x_0) > 0$, then there exists $\delta > 0$:*

$$f(x_0) < f(x), \quad \forall x \in (x_0 - \delta, x_0 + \delta)\backslash\{x_0\}$$

We say that f has a local minimum at x_0.

Proof Note

$$f''(x_0) = \lim_{h\to 0} \frac{f'(x_0 + h) - f'(x_0)}{h} > 0$$

Then there exists $\delta > 0$ such that $\frac{f'(x_0+h)-f'(x_0)}{h} > 0$ for all $h \in [-\delta, \delta]$. In particular, as $f'(x_0) = 0$, $f'(x_0 + h) > 0$ for $h \in (0, \delta]$ and $f'(x_0 + h) < 0$ for $h \in [-\delta, 0)$. Consider $h \in (0, \delta]$. By the mean value theorem there exists $c \in (x_0, x_0 + h)$ such that

$$f(x_0 + h) - f(x_0) = hf'(c) > 0$$

so $f(x_0) < f(x_0 + h)$. If $h \in [-\delta, 0)$, by the mean value theorem there exists $c \in (x_0 + h, x_0)$ such that

$$f(x_0 + h) - f(x_0) = hf'(c) < 0$$

so $f(x_0) < f(x_0 + h)$ again. Thus, for all $x \in (x_0 - \delta, x_0 + \delta)\backslash\{x_0\}$, we have $f(x_0) < f(x)$. □

The result for a local maximum is then given by replacing f by $-f$ in the previous result.

Proposition 2.5 *Suppose f is twice differentiable on (a, b) and $f''(x) > 0$ on (a, b). Then for all $a < x_0 < x_1 < b$ and $\lambda \in (0, 1)$,*

$$f(\lambda x_0 + (1 - \lambda)x_1) < \lambda f(x_0) + (1 - \lambda)f(x_1)$$

Proof Let $g(s) = sf(x_0) + (1 - s)f(x_1) - f(sx_0 + (1 - s)x_1)$. Then $g(0) = f(x_1) - f(x_1) = 0$ and $g(1) = f(x_0) - f(x_0) = 0$. Note g is also twice differentiable. Towards a contradiction suppose there exists $c \in (0, 1)$ such that $g(c) < 0$. Since g is continuous it attains its minimum, so there exists $s_0 \in (0, 1)$ such that $g(s_0) \leq g(c) < 0$. Further, $g'(s_0) = 0$, where $g'(s) = f(x_0) - f(x_1) - f'(sx_0 + (1 - s)x_1)(x_0 - x_1)$. Then

$$f'(s_0x_0 + (1 - s_0)x_1) = \frac{f(x_1) - f(x_0)}{x_1 - x_0}$$

and $g''(s) = -f''(sx_0 + (1 - s)x_0)(x_0 - x_1)^2 < 0$. In particular, $g''(s_0) < 0$, contradicting the fact that $g(s_0)$ is a minimum and Proposition 2.4 □

Appendix B: Functional Sequences and Series

We are now interested in the study of series of functions, or in other words functions of the form

$$f(x) = f_1(x) + f_2(x) + f_3(x) + \ldots$$

In such a situation $\{f_n\}$ will be some sequence of functions; for each x we obtain a sequence of numbers $\{f_n(x)\}$, and $f(x)$ is the sum of this sequence. Recall that each sum $f_1(x)+f_2(x)+f_3(x)+\ldots$ is, by definition, the limit of the sequence $f_1(x), f_1(x) + f_2(x), f_1(x) + f_2(x) + f_3(x), \ldots$. If we define a new sequence of functions $\{s_n\}$ by

$$s_n = f_1 + \ldots + f_n$$

then we can express this fact more succinctly by writing

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

for some $x \in \mathbb{R}$.

First let us consider functions of the form

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

All this form may seem simple, it is very important to note that ***nothing one would hope to be true actually is***. Instead we have a flurry of lovely counter-examples.

Example 2.1 (Counter-Example 1) Even if each f_n is continuous, the function f may not be! Indeed, consider the sequence of functions

$$f_n(x) = \begin{cases} x^n, & 0 \le x \le 1 \\ 1, & x \ge 1 \end{cases}$$

These functions are all continuous, but the function $f(x) = \lim_{n\to\infty} f_n(x)$ is not continuous; in fact,

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

Another example of this phenomenon is illustrated by the family of functions

$$f_n(x) = \begin{cases} -1, & x \le -\frac{1}{n} \\ nx, & -\frac{1}{n} \le x \le \frac{1}{n} \\ 1, & \frac{1}{n} \le x \end{cases}$$

In this case, if $x < 0$ $f_n(x)$ is eventually -1, and if $x > 0$, then $f_n(x)$ is eventually 1, while $f_n(0) = 0$ for all n. Thus,

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

so once again $f(x) = \lim_{n\to\infty} f_n(x)$ is not continuous.

Example 2.2 (Counter-Example 2) It is even possible to produce a sequence of differentiable functions $\{f_n\}$ for which the function $f(x) = \lim_{n\to\infty} f_n(x)$ is not continuous. One such sequence is

$$f_n(x) = \begin{cases} -1, & x \le -\frac{1}{n} \\ \sin\left(\frac{n\pi x}{2}\right), & -\frac{1}{n} \le x \le \frac{1}{n} \\ 1, & \frac{1}{n} \le x \end{cases}$$

These functions are differentiable, but we still have

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

Definition 2.13 If f is a function defined on some set A, and a sequence of functions $\{f_n\}$, all defined on the same set A, are such that only

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

for all $x \in A$. Precisely, $\{f_n\}$ is said to ***converge pointwise to f on A*** if for all $\varepsilon > 0$, and for all $x \in A$, there is some N such that if $n \ge N$, then $|f(x) - f_n(x)| < \varepsilon$.

Definition 2.14 (Pointwise Convergence (alt.)) Suppose $S \subseteq \mathbb{R}$ and $f_n : S \to \mathbb{R}$ is a real-valued function for each $n \in \mathbb{N}$. We say that the sequence of functions $\{f_n\}$ ***converges pointwise on S***

to $f : S \to \mathbb{R}$ if for every $x \in S$ and every $\varepsilon > 0$, there exists $N \in \mathbb{N}$ such that if $n \geq N$, $|f_n(x) - f(x)| < \varepsilon$.

Example 2.3 Take $f_n(x) = x^n$ on $S = [0, 1]$. If $0 \leq x < 1$ notice that $\lim\limits_{n\to\infty} x^n = 0$ (geometric sequence). If $x = 1$, then $f_n(1) = 1^n = 1$, which converges to 1 as n goes to infinity. Thus, f_n converges pointwise to

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

Notice each f_n is continuous on $[0, 1]$, but f is not.

This example answers the following question in the negative:

? [

Question] Suppose f_n converges pointwise to f on $S \subseteq \mathbb{R}$. If $a \in S'$ (an accumulation/limit point for S), $\lim\limits_{x\to a} f(x)$ exists and $\lim\limits_{x\to a} f_n(x)$ exists for all n, is it true that

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

In particular, in our example we take $a = 1$, then $\lim\limits_{x\to a} f_n(x) = \lim\limits_{x\to 1} x^n = 1$, so $\lim\limits_{n\to\infty} 1 = 1$, but $\lim\limits_{x\to 1} f(x) = 0$.

Example 2.4 Consider the sequence $g_n(x) = \frac{1}{1+x^n}$ on $S = (-\infty, -1) \cup (-1, \infty)$. As n goes to infinity we have that the sequence converges pointwise to 1 for $|x| < 1$, 1/2 for $x = 1$, and 0 for $|x| > 1$.

Example 2.5 Let $S = [0, \infty)$ and define

$$f_n(x) = \begin{cases} n^2 x & 0 \leq x \leq \frac{1}{n} \\ -n^2\left(x - \frac{2}{n}\right) & \frac{1}{n} < x \leq \frac{2}{n} \\ 0 & x > \frac{2}{n} \end{cases}$$

We claim that $\lim\limits_{n\to\infty} f_n(x) = 0$ for all $x \geq 0$. When $x = 0$, $f_n(0) = 0$ which converges to 0. If $0 < x$: By the Archemedean property there exists $N \in \mathbb{N}$ such that $\frac{2}{N} < x$. Then $f_N(x) = 0$ and $f_n(x) = 0$ for all $n \geq N$. Thus, $\lim\limits_{n\to\infty} f_n(x) = 0$, as claimed. This argument can be intuitively realized by noting that for n large enough, the tent is always to the left of any $0 < x$.

We note that this gives an example of an unbounded sequence $\{f_n\}$ converging pointwise to a bounded function.

Example 2.6 Take $S = \mathbb{R}$ and define $f_n(x) = \frac{\sin(nx)}{\sqrt{n}}$, waves of declining amplitude but increasing frequency. Note that $0 \leq |f_n(x)| = \left|\frac{\sin(nx)}{\sqrt{n}}\right| \leq \frac{1}{\sqrt{n}}$ which goes to 0 as n goes to infinity, so the sequence converges pointwise to 0. Notice $f_n'(x) = \frac{n}{\sqrt{n}}\cos(nx) = \sqrt{n}\cos(nx)$, which has no limit for any $x \in \mathbb{R}$.

This is an example of pointwise convergence where the derivatives do not converge pointwise. Moreover, as we will see, the original sequence actually converges uniformly (the bound on the terms does not depend on x), which suggests we need a stronger requirement for convergence of derivatives.

Definition 2.15 Let $\{f_n\}$ be a sequence of functions defined on A, and let f be a function which is also defined on A. Then f is called the ***uniform limit of*** $\{f_n\}$ ***on*** A if for every $\varepsilon > 0$ there is some N such that for all $x \in A$,

$$\text{if } n > N, \text{ then } |f(x) - f_n(x)| < \varepsilon$$

We also say that $\{f_n\}$ ***converges uniformly to*** f ***on*** A, or that f_n ***approaches*** f ***uniformly on*** A.

Definition 2.16 (Uniform Convergence (alt.)) Suppose $S \subseteq \mathbb{R}$ and $f_n : S \to \mathbb{R}$ are real-valued functions for each $n \in \mathbb{N}$. We say that f_n ***converges uniformly on*** S to $f : S \to \mathbb{R}$ if for all $\varepsilon > 0$, there is an $N \in \mathbb{N}$ (depends only on ε) such that $n \geq N$ implies

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

for all $x \in S$.

We will use the notation $f_n \to_u f$ sometimes to denote uniform convergence. Intuitively uniform convergence can be understood by saying that given any band or tube containing f on S, there is a point past which the tail functions of the sequence reside entirely in this band.

Remark 2.4 Note that uniform convergence implies pointwise convergence, but the converse is not true.

Definition 2.17 (Uniform Norm) Suppose $S \subseteq \mathbb{R}$ and f is a function on S. The ***uniform norm*** of f on S is given by

$$||f||_S := \sup_{x \in S} |f(x)|$$

We note that this may not be finite. When the context is clear we will write $||f||_\infty$.

Remark 2.5 Suppose $f : S \to \mathbb{R}$ is a function:

(i) f is bounded on S if and only if $||f||_S < \infty$

(ii) If f is continous on S and S is compact, then $||f||_S = \max_{s \in S} |f(x)|$ (e.g. $||x^n||_{[0,1]} = 1$)

Proposition 2.6 *A sequence of functions f_n converges uniformly to f on S if and only if $||f_n - f||_S \to 0$.*

Proof Let f_n be a sequence of functions, each defined on S, and let f be another function on S.

First, let us suppose that the f_n converge uniformly to f on S. Fix $\varepsilon > 0$. Then there exists $N \in \mathbb{N}$ such that for $n \geq N$, $|f_n(x) - f(x)| < \varepsilon/2$ for all $x \in S$. This implies that $\varepsilon/2$ is an upper bound of all $|f_n(x) - f(x)|$, so by definition $||f_n - f||_S \leq \varepsilon/2 < \varepsilon$ for all $n \geq N$. Hence, as $||f_n - f||_S = |||f_n - f||_S - 0|$, we have that the sequence $||f_n - f||_S$ converges to 0 in $\mathbb{R}$.

Conversely, suppose that $||f_n - f||_S$ converges to 0, and let $\varepsilon > 0$. Then, there exists $N \in \mathbb{N}$ such that for all $n \geq N$, $||f_n - f||_S < \varepsilon$. It follows that for all $x \in S$, $|f_n(x) - f(x)| \leq ||f_n - f||_S < \varepsilon$ for $n \geq N$, so we find that f_n converges uniformly to f on S by definition. □

Example 2.7 Let $S = [0, 1]$ and $f_n(x) = x^n$. f_n does not converge uniformly to any function. Indeed if f_n converges uniformly to anything on $[0, 1]$, it must be the pointwise limit $f(x) = 0, 0 \leq x < 1$ and $f(x) = 1, x = 1$, since uniform implies pointwise convergence (and limits are unique in Hausdorff spaces). This implies $\sup_{x\in[0,1]} |f_n(x) - f(x)|$ goes to zero as n goes to infinity. But if $x = 1 - \frac{1}{n}$, then $|f_n(1 - 1/n) - f(1 - 1/n)| = (1 - 1/n)^n$, which converges to $e^{-1} = \frac{1}{e}$, and not zero. Further, this must be less than what the limit of the supremums converges to, so the supremums cannot converge to 0, as that would result in a contradiction. Hence, the sequence does not converge uniformly.

The tent function is another example of pointwise but not uniform convergence, since $||f_n - 0||_\infty = n$, which does not converge to 0.

Theorem 2.29 (Cauchy Criterion for Uniform Convergence) *Suppose $f_n : S \to \mathbb{R}$. Then f_n converges uniformly on S if and only if for all $\varepsilon > 0$ there is an $N \in \mathbb{N}$ such that $m, n \geq N$ implies*

$$||f_n - f_m||_\infty < \varepsilon$$

Proof Let $f_n : S \to \mathbb{R}$ be a sequence of functions.

First, suppose f_n converges uniformly on S to some $f : S \to \mathbb{R}$. Fix $\varepsilon > 0$. Then there exists $N \in \mathbb{N}$ such that for $n \geq N$, $|f_n(x) - f(x)| < \varepsilon/3$, for all $x \in S$. It follows by the triangle inequality that for $k, m \geq N$, $|f_k(x) - f_m(x)| < 2\varepsilon/3$ for all $x \in S$. Thus, $||f_k - f_m||_\infty \leq 2\varepsilon/3 < \varepsilon$, so f_n is uniformly Cauchy (i.e. it is Cauchy with respect to the uniform norm $|| \cdot ||_\infty$)

Conversely, suppose f_n is uniformly Cauchy on S. Let $\varepsilon > 0$. Then there exists $N \in \mathbb{N}$ such that for $k, m \geq N$, $||f_k - f_m||_\infty < \varepsilon/2$. Then for all $x \in S$, $|f_k(x) - f_m(x)| < \varepsilon/2$ for $k, m \geq N$. As $(\mathbb{R}, | \cdot |)$ is a complete metric space, for each $x \in S$ there exists $f(x) \in \mathbb{R}$ such that $f_k(x)$ converges to $f(x)$. Then, taking the limit as m goes to infinity in $|f_k(x) - f_m(x)| < \varepsilon/2$, and using the order properties of limits in $\mathbb{R}$, we have that $|f_k(x) - f(x)| \leq \varepsilon/2$ for all $x \in S$ and $k \geq N$. Thus, for $k \geq N$ we have that $||f_k - f||_\infty < \varepsilon$, so f_k converges to f uniformly on S, as desired. □

Example 2.8 Let $S = [0, 1]$ and $f_n(x) = nx(1 - x^2)^n$. We claim that f_n converges pointwise to 0, but not uniformly. If $x = 0, 1$, then $f_n(x) = 0 \to 0$. If $0 < x < 1$, then $0 < 1 - x^2 < 1$. Then

$1/(1-x^2) > 1$, so there exists $y > 0$ such that $1/(1-x^2) = 1 + y$. Then, for $n \geq 2$ we have by the binomial theorem that $(1+y)^n \leq 1 + ny + n(n+1)/2y^2 \leq 1 + ny + n^2y^2/2$. It follows that

$$\begin{aligned} 0 \leq \lim_{n\to\infty} nx(1-x^2)^n &\leq \lim_{n\to\infty} \frac{nx}{1+ny+n^2y^2/2} && \text{(for some } y > 0\text{)} \\ &= x \lim_{n\to\infty} \frac{1}{1/n + y + ny^2/2} \\ &= x \cdot 0 = 0 \end{aligned}$$

as claimed. But, the convergence is not uniform since if $||f_n||_\infty \to 0$, then $|f_n(x_n)| \to 0$ for any sequence x_n in $[0,1]$. Let $x_n = \frac{1}{\sqrt{n}}$. Then $f_n(x_n) = n\frac{1}{\sqrt{n}}(1-1/n)^n = \sqrt{n}(1-1/n)^n$, which converges to ∞ as $\sqrt{n} \to \infty$ and $(1-1/n)^n \to e^{-1}$.

Example 2.9 Let $S = [0,b]$ for some $b > 0$. Then $f_n(x) = n\sin(x/n)$ converges uniformly on S but not on $[0,\infty)$. Note if we fix x and take n large enough, then $\sin(x/n) \approx x/n$. Then

$$\lim_{n\to\infty} f_n(x) = \lim_{n\to\infty} \frac{\sin(x/n)}{x/n} x = x$$

so $f_n(x) \to x$ for all x. We first show $f_n \not\to_u x$ on $[0,\infty)$. Suppose it did. Then $|f_n(x_n) - x_n| \to 0$ for any sequence x_n in $[0,\infty)$. Try $x_n = n^2$. Then $|f_n(n^2) - n^2| = |n\sin(n) - n^2| = n|\sin(n) - n| \to \infty$, which is a contradiction, so the claim holds. To show uniform convergence on $[0,b]$, note that $\sin(x/n) \leq x/n$ for all $x \geq 0$, so $n\sin(x/n) \leq x$. Let $g(x) = x - n\sin(x/n) \geq 0$. Then $g'(x) = 1 - \cos(x/n) \geq 0$. Thus, g is a monotone increasing function on $[0,b]$. Thus, $||g||_\infty = g(b) = b - n\sin(b/n)$. Then $||f_n - x||_{[0,b]} = ||g||_{[0,b]} = b - n\sin(b/n)$, which converges to $b - b = 0$ as $n \to \infty$.

Note that $[0,\infty) = \bigcup_{b>0}[0,b]$, sp this gives an example of uniform convergence on many sets, but not on the union of those sets.

Theorem 2.30 *Suppose $S \subseteq \mathbb{R}$, and $f_n : S \to \mathbb{R}$ are continuous functions. If f_n converges uniformly to f on S, then f is also continuous on S.*

Proof First, fix $\varepsilon > 0$. Then there exists $N \in \mathbb{N}$ such that for $k \geq N$, $||f_k - f||_S < \varepsilon/3$. Let $x \in S$. As f_N is continuous at x, there exists $\delta > 0$ such that for $y \in B^*_\delta(x)$, $|f_N(y) - f_N(x)| < \varepsilon/3$. Then, for $|x - y| < \delta$ we have that

$$|f(x) - f(y)| \leq |f(x) - f_N(x)| + |f_N(x) - f_N(y)| + |f_N(y) - f(y)| < 3\varepsilon/3 = \varepsilon$$

so f is indeed continuous at x, and hence on S. □

Proposition 2.7 *If f_n converges uniformly to f on S and each f_n is uniformly continuous on S, then f is uniformly continuous on S.*

Proof First, fix $\varepsilon > 0$. As $f_n \to_u f$, there exists $N \in \mathbb{N}$ such that for $k \geq N$, $||f_k - f||_S < \varepsilon/3$. Then, as f_N is uniformly continuous there exists $\delta > 0$ such that if $|x - y| < \delta$, $|f_N(x) - f_N(y)| < \varepsilon/3$ for all $x, y \in S$. Thus, if $|x - y| < \delta$ and $x, y \in S$, it follows that

$$|f(x) - f(y)| \leq |f(x) - f_N(x)| + |f_N(x) - f_N(y)| + |f_N(y) - f(y)|$$

$$< 3\varepsilon/3 = \varepsilon$$

so f is uniformly continuous on S, as desired. □

Example 2.10 Let $S = [0,1]$ and $f_n(x) = \frac{1}{1+x^n}$. We have seen that $f_n(x) \to 1$ for $x \in [0,1)$, and $f_n(1) \to 1/2$, pointwise. Evidently, the convergence is not uniform since each f_n is continuous on $[0,1]$, but the limit function is not.

Lemma 2.2 *Suppose f_n converges uniformly to f on S. If each f_n are bounded on S, then f is bounded on S as well. Moreover, the sequence is* **uniformly bounded***:*

$$\sup_{n\in\mathbb{N}} ||f_n||_S < \infty$$

Proof First, there exists $N \in \mathbb{N}$ such that if $k \geq N$, $||f_k - f||_S < 1$, by uniform convergence. Then, for all $x \in S$, $|f(x)| \leq |f(x) - f_N(x)| + |f_N(x)| < ||f_N||_S + 1$. But f_N is assumed bounded on S, so $||f_N||_S < \infty$, and hence $||f||_S \leq ||f_N||_S + 1 < \infty$, so f is also bounded on S. Now, for all $k \geq N$ we have that $||f_k||_S \leq ||f||_S + 1$. Let $M = \max\{||f_1||_S, ..., ||f_{N-1}||_S, ||f_N||_S + 1\}$. Then it follows that $||f_n||_S \leq M$ for all $n \in \mathbb{N}$, and $M < \infty$ is finite, so $\sup_{n\in\mathbb{N}} ||f_n||_S \leq M < \infty$, as desired. □

Theorem 2.31 *Suppose that f_n are bounded and integrable functions on $[a,b]$, $f_n \in \mathcal{R}([a,b])$, converging uniformly to f on $[a,b]$. Then f is integrable on $[a,b]$ and*

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

Proof By the previous lemma we have that f is bounded on $[a,b]$. First, if f is integrable on $[a,b]$, then

$$\begin{aligned}\left|\int_a^b f_n(x)dx - \int_a^b f(x)dx\right| &\leq \int_a^b |f_n(x) - f(x)|dx \\ &\leq \int_a^b ||f_n - f||_\infty dx \\ &= (b-a)||f_n - f||_\infty\end{aligned}$$

which goes to 0 as $n \to \infty$ by uniform convergence. Thus, all we need to show is f is integrable. Fix $\varepsilon > 0$. Then there exists $N \in \mathbb{N}$ such that $||f_n - f|| < \frac{\varepsilon}{3(b-a)}$. Since each f_n is integrable, there exists a partition $P_n \in \prod([a,b])$ with

$$U(f_n, P_n) - L(f_n, P_n) < \frac{\varepsilon}{3}$$

Write $P_n = \{a = x_{0,n}, x_{1,n}, ..., x_{M_n,n} = b\}$. Let $J_{k,n} = [x_{k,n}, x_{k+1,n}]$ for $0 \leq k \leq N_n$. Then $||f - f_n||_{J_k} < \frac{\varepsilon}{3(b-a)}$ for $n \geq N$. Then

$$||f||_{J_k} \leq \frac{\varepsilon}{3(b-a)} + ||f_n||_{J_k}$$

and

$$||f_n||_{J_k} \leq \frac{\varepsilon}{3(b-a)}$$

, so

$$|\sup_{J_k}(f) - \sup_{J_k}(f_n)| \leq \frac{\varepsilon}{3(b-a)}$$

for all $n \geq N$. It follows that

$$\begin{aligned}
|U(f,P_n) - U(f_n,P_n)| &= \left|\sum_{k=0}^{M_n-1} (\sup_{J_k}(f) - \sup_{J_k}(f_n))\ell(J_k)\right| \\
&\leq \sum_{k=0}^{M_n-1} |\sup_{J_k}(f) - \sup_{J_k}(f_n)|\ell(J_k) \\
&< \sum_{k=0}^{M_n-1} \frac{\varepsilon}{3(b-a)}\ell(J_k) \\
&= \frac{\varepsilon}{3(b-a)}\ell([a,b]) = \frac{\varepsilon}{3}
\end{aligned}$$

Similarly, $|L(f,P_n) - L(f_n,P_n)| < \varepsilon/3$. Finally,

$$\begin{aligned}
U(f,P_n) - L(f,P_n) &= U(f,P_n) - U(f_n,P_n) + U(f_n,P_n) - L(f_n,P_n) + L(f_n,P_n) - L(f,P_n) \\
&\leq |U(f,P_n) - U(f_n,P_n)| + |U(f_n,P_n) - L(f_n,P_n)| + |L(f_n,P_n) - L(f,P_n)| \\
&< 3 \cdot \frac{\varepsilon}{3} = \varepsilon
\end{aligned}$$

Thus, f is Riemann integrable on $[a,b]$. □

Remark 2.6 Although these last two theorems are great successes, differentiability sadly fails. Even if each f_n is differentiable and $\{f_n\}$ converges uniformly to f, it need not be the case that f is differentiable. Moreover, even if f is itself differentiable, it need not be the case that

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

Example 2.11 (Counter Example 3) Consider the family of functions

$$f_n(x) = \frac{1}{n}\sin(n^2x)$$

then $\{f_n\}$ converges uniformly to the function $f(x) = 0$, but

$$f_n'(x) = n\cos(n^2x)$$

and $\lim_{n\to\infty} n\cos(n^2x)$ does not even always exist (for example if $x = 0$).

Theorem 2.32 *Suppose that $\{f_n\}$ is a sequence of functions which are differentiable on $[a,b]$, with integrable derivatives f_n', and that $\{f_n\}$ converges (pointwise) to f. Suppose, moreover, that $\{f_n'\}$ converges uniformly on $[a,b]$ to some continuous function g. Then f is differentiable and*

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

Proof Applying Theorem 1 to the interval $[a, x]$, we see that for each x we have

$$\begin{aligned}\int_a^x g &= \lim_{n\to\infty} \int_a^x f_n' \\ &= \lim_{n\to\infty} [f_n(x) - f_n(a)] && \text{(by ??)} \\ &= f(x) - f(a)\end{aligned}$$

Since g is continuous, it follows that $f'(x) = g(x) = \lim\limits_{n\to\infty} f_n'(x)$ for all x in the interval $[a, b]$, by **??**. □

Example 2.12 Consider $\int_0^\pi \frac{n+\sin x}{3n+\sin^2(nx)} dx$. Observe that

$$\begin{aligned}\left|\frac{n+\sin x}{3n+\sin^2(nx)} - \frac{1}{3}\right| &= \left|\frac{3n + 3\sin x - 3n - \sin^2(nx)}{3(3n+\sin^2(nx))}\right| \\ &\le \frac{|3\sin x - \sin^2(nx)|}{9n} \le \frac{3+1}{9n} = \frac{4}{9n}\end{aligned}$$

for all $x \in [0, \pi]$. Thus, $||f_n - \frac{1}{3}||_{[0,\pi]} \le \frac{4}{9n} \to 0$, so it converges uniformly. It follows that

$$\begin{aligned}\lim_{n\to\infty} \int_0^\pi \frac{n+\sin x}{3n+\sin^2(nx)} dx &= \int_0^\pi \lim_{n\to\infty} \frac{n+\sin x}{3n+\sin^2(nx)} dx \\ &= \int_0^\pi \frac{1}{3} dx \\ &= \frac{\pi}{3}\end{aligned}$$

Definition 2.18 The series $\sum\limits_{n=1}^{\infty} f_n$ ***converges uniformly*** (more formally, the sequence $\{f_n\}$ is ***uniformly summable***) ***to f on A***, if the sequence

$$f_1, f_1 + f_2, f_1 + f_2 + f_3, \ldots$$

converges uniformly to f on A.

Definition 2.19 (Series of Functions (alt.)) Suppose S is a subset of $\mathbb{R}$ and $f_n : S \to \mathbb{R}$. We say that the series $\sum_{n=1}^{\infty} f_n(x)$ ***converges pointwise*** if the sequence of partial sums $s_k(x) = \sum_{n=1}^{k} f_n(x)$ converges pointwise to a function $f : S \to \mathbb{R}$. In this case, write $f(x) = \sum_{n=1}^{\infty} f_n(x)$. We say that the series $\sum_{n=1}^{\infty} f_n(x)$ ***converges uniformly*** if s_k converges uniformly to f on S.

Corollary 2.4 *Let $\sum\limits_{n=1}^{\infty} f_n$ converge uniformly to f on $[a, b]$.*

1. If each f_n is continuous on $[a, b]$, then f is continuous on $[a, b]$.

2. If f and each f_n is integrable on $[a,b]$, then

$$\int_a^b f = \int_a^b \sum_{n=1}^{\infty} f_n = \sum_{n=1}^{\infty} \int_a^b f_n$$

Moreover, if $\sum\limits_{n=1}^{\infty} f_n$ converges (pointwise) to f on $[a,b]$, each f_n has an integrable derivative f_n' and $\sum\limits_{n=1}^{\infty} f_n'$ converges uniformly on $[a,b]$ to some continuous function, then

3. $f'(x) = \sum\limits_{n=1}^{\infty} f_n'(x)$ for all $x \in [a,b]$

Proof Let $\{s_n\}$ be the sequence of partial sums of the $\{f_n\}$. Then since each f_n is continuous, so is each s_n. Then as $\{s_n\}$ converges uniformly to f we have by a previous theorem that f is also continuous on $[a,b]$. Next, since each f_n is integrable on $[a,b]$, so is each s_n. Then as $\{s_n\}$ converges uniformly to f we have that

$$\begin{aligned}\int_a^b f &= \lim_{n\to\infty} \int_a^b s_n \\ &= \lim_{n\to\infty} t_n \\ &= \sum_{n=1}^{\infty} \int_a^b f_n\end{aligned}$$

where $\{t_n\}$ is the sequence such that

$$t_n = \sum_{i=1}^{n} \int_a^b f_n = \int_a^b s_n$$

Finally, suppose s_n converges (pointwise) to f on $[a,b]$, and each f_n has an integrable derivative f_n'. Then each s_n has an integrable derivative s_n' on $[a,b]$ by the linearity of the derivative and integral operators. Moreover, suppose s_n' converges uniformly on $[a,b]$ to some continuous function g. Then it follows that for all $x \in [a,b]$

$$f'(x) = \lim_{n\to\infty} s_n'(x) = \sum_{n=1}^{\infty} f_n'(x)$$

Example 2.13 Let $S = [-r,r]$ with $0 < r < 1$. Let $f(t) = \sum_{n=0}^{infty}(-r)^n$. Then as $t \in [-r,r] \subset (-1,1)$ we have $f(t) = \frac{1}{t}$. Convergence is uniform on $[-r,r]$ since $|t| \le r$, so $|(-t)^n| \le r^n$, and $\sum_{n=0}^{\infty} r^n$ converges. By the last corollary, we have that for $x \in [-r,r]$,

$$\log(1+x) = \int_0^x \frac{1}{1+t}dt \qquad \text{(by the FTC)}$$

$$
\begin{aligned}
&= \int_0^x \sum_{n=0}^{\infty} (-t)^n dt \\
&= \sum_{n=0}^{\infty} \int_0^x (-1)^n t^n dt \\
&= \sum_{n=0}^{\infty} (-1)^n \frac{x^{n+1}}{n+1} && \text{(by the FTC)}
\end{aligned}
$$

Thus, $\log(1+x) = \sum_{n=0}^{\infty} (-1)^n \frac{x^{n+1}}{n+1}$ for all $x \in (-1, 1)$

Theorem 2.33 (The Weierstrass M-Test) *Let $\{f_n\}$ be a sequence of functions defined on A, and suppose that $\{M_n\}$ is a sequence of numbers such that*

$$|f_n(x)| \leq M_n, \forall x \in A$$

Suppose moreover that $\sum\limits_{n=1}^{\infty} M_n$ converges. Then for each x in A the series $\sum\limits_{n=1}^{\infty} f_n(x)$ converges absolutely, and $\sum\limits_{n=1}^{\infty} f_n$ converges uniformly on A to the function

$$f(x) = \sum_{n=1}^{\infty} f_n(x)$$

Proof For each $x \in A$, the series $\sum\limits_{n=1}^{\infty} |f_n(x)|$ converges by **??**; consequently $\sum\limits_{n=1}^{\infty} f_n(x)$ converges absolutely. Moreover, for all $x \in A$ we have

$$
\begin{aligned}
\left| f(x) - \sum_{i=1}^{N} f(x) \right| &= \left| \sum_{n=N+1}^{\infty} f_n(x) \right| \\
&\leq \sum_{n=N+1}^{\infty} |f_n(x)| \\
&\leq \sum_{n=N+1}^{\infty} M_n
\end{aligned}
$$

Since $\sum\limits_{n=1}^{\infty} M_n$ converges, the number $\sum\limits_{n=N+1}^{\infty} M_n$ can be made as small as desired (by **??**), by choosing N sufficiently large. □

Example 2.14 The series $\sum_{n=0}^{\infty} \frac{x^n}{n!}$ converges uniformly on $[-R, R]$ for any $R > 0$, but not on $\mathbb{R}$. If $-R \leq x \leq R$, then $\left|\frac{x^n}{n!}\right| \leq \frac{R^n}{n!} =: M_n$. Note that $\sum_{n=0}^{\infty} M_n$ converges by the ratio test. This implies by the Weierstrass M-test that $\sum_{n=0}^{\infty} \frac{x^n}{n!}$ converges uniformly on $[-R, R]$. Note if $\sum_n f_n$ converges uniformly on S to some function s, then $f_n \to_u 0$. In particular, $||f_n||_\infty \to 0$. But, $||x^n/n!||_{\mathbb{R}} = \infty$, so the series cannot be uniformly convergent on all of $\mathbb{R}$.

The converse to the uniform limit of the terms needing to go to zero for the series to converge is false. For example, if $f_n(x) = 1/n, x \in \mathbb{R}$, then $||f_n||_{\mathbb{R}} = 1/n \to 0$, but $\sum_{n=1}^{\infty} \frac{1}{n}$ diverges (the harmonic series).

Example 2.15 Recall we saw that $\sum_{n=0}^{\infty} \frac{x^n}{n!}$ converges uniformly on $[-R, R]$ for any $R > 0$ by Weierstrass M-test. Since $\sum_{n=0}^{N} \frac{x^n}{n!}$ are continuous, $\sum_{n=0}^{\infty} \frac{x^n}{n!}$ is continuous on $[-R, R]$ for all R so it is continuous on $\mathbb{R}$.

Theorem 2.34 (Riemann 1861; Weierstrass 1872; du Bois-Reymond 1875) *There is a continuous function on $\mathbb{R}$ that is nowhere differentiable.*

Proof Let $g(x) = |x|$ on $[-1, 1]$, and continuously and periodically extend to $\mathbb{R}$ so $g(x+2) = g(x)$ for all x. Let $r, s \in \mathbb{R}$. If $|r-s| > 1$, then $|r-s| > 1 \geq |g(r)-g(s)|$, as $-1 = 0-1 \geq g(r)-g(s) \leq 1-0 = 1$. Otherwise, suppose $|r - s| \leq 1$. Without loss of generality suppose $r \geq s$. If there does not exist $s \leq n \leq r$ for some $n \in \mathbb{Z}$, then there exists $n \in \mathbb{Z}$ such that $n \leq s \leq r \leq n + 1$. It follows that either $g(r) - g(s) = r - n - (s - n) = r - s$, or $g(r) - g(s) = -(r - n - 1) + (s - n - 1) = s - r$. In either case $|g(r) - g(s)| = |r - s|$. Otherwise, there exists $n \in \mathbb{Z}$ such that $n - 1 \leq s \leq n \leq r \leq n + 1$. Then

$$g(r) - g(s) = (r - n) - |s - n| = r - n - n + s = r + s - 2n \leq r + s - 2s = r - s = |r - s|$$

or

$$g(r)-g(s) = |r-(n+1)|-(s-(n-1)) = n+1-r-s+n-1 = 2n-r-s \leq 2r-r-s = r-s = |r-s|$$

Thus, in any case we have $|g(r)-g(s)| \leq |r-s|$ for all $r, s \in \mathbb{R}$. This says that g is Lipshitz continuous, and in particular is uniformly continuous on $\mathbb{R}$. Now, for all $n \in \mathbb{N}$, let $g_n(x) = \frac{3^n}{4^n} g(4^n x)$. Notice that $|g_n(x)| \leq \left(\frac{3}{4}\right)^n$, so the sum $\sum_n g_n$ converges uniformly on $\mathbb{R}$ by the Weierstrass M-test. Let $f(x) = \sum_{n=0}^{\infty} g_n(x)$, where $g_0(x) = g(x)$. Then f is continuous on $\mathbb{R}$ as the g_n are continuous and they converge uniformly (in particular, f is uniformly continuous on $\mathbb{R}$). Fix $x \in \mathbb{R}$ and $m \in \mathbb{N}$. Define $\delta_m = \pm\frac{1}{2} \cdot \frac{1}{4^m}$, where $+$ or $-$ is selected so that there is no integer between $4^m x$ and $4^m(x + \delta)$, which has distance $\frac{1}{2}$. We have $|4^m x - 4^m(x + \delta_m)| = \frac{1}{2}$. We will show $\left|\frac{f(x+\delta_m)-f(x)}{\delta_m}\right| \to \infty$ as $m \to \infty$, so $\delta_m \to 0$. We compute

$$\begin{aligned} f(x + \delta_m) - f(x) &= \sum_{n=0}^{\infty} g_n(x + \delta_m) - g_n(x) \\ &= \sum_{n=0}^{\infty} \frac{3^n}{4^n} \left[g(4^n x + 4^n \delta_m) - g(4^n x)\right] \end{aligned}$$

We have three cases on each term:

(i) $n > m$: Then $4^n \delta_m = \frac{\pm 4^{n-m}}{2}$ is an even integer. Since $g(t + 2) = g(t)$ for all t, $g(4^n x + 4^n \delta_m) = g(4^n x)$. So

$$f(x + \delta_m) - f(x) = \sum_{n=0}^{m} \frac{3^n}{4^n} \left[g(4^n x + 4^n \delta_m) - g(4^n x)\right]$$

(ii)$n < m$: We have $|g(r) - g(s)| \leq |r - s|$ for all $r, s \in \mathbb{R}$. Thus

$$|g(4^n x + 4^n \delta_m) - g(4^n x)| \le 4^n |\delta_m|$$

(iii)$n = m$: There is no integer between $4^n x + 4^n \delta_n$ and $4^n x$ which means g maps these two points to the same line segment. Then $|g(4^n x + 4^n \delta_n) - g(4^n x)| = |4^n \delta_n| = \frac{1}{2}$

Now we compute

$$\begin{aligned}
\left|\frac{f(x+\delta_m) - f(x)}{\delta_m}\right| &= \left|\sum_{n=0}^{m} 3^n 4^n \frac{g(4^n x + 4^n \delta_m) - g(4^n x)}{\delta_m}\right| \\
&\ge \left|\frac{3^m}{4^m}\frac{1/2}{\delta_m}\right| - \sum_{n=0}^{m-1}\left|\frac{3^n(g(4^n x + 4^n \delta_m) - g(4^n x))}{4^n \delta_m}\right| \\
&\ge 3^m - \sum_{n=0}^{m-1} \frac{4^n |\delta_m|}{|\delta_m|}\frac{3^n}{4^n} \\
&= 3^m - \sum_{n=0}^{m-1} 3^n \\
&= 3^m - \frac{1-3^m}{1-3} \\
&= \frac{2 \cdot 3^m + 1 - 3^m}{2} \\
&= \frac{1}{2}(3^m + 1) \to \infty
\end{aligned}$$

as $m \to \infty$, as desired. Thus, as if f was differentiable this limit would be its derivative, so f is not differentiable at any $x \in \mathbb{R}$. □

Example 2.16 Let $S = (0,1)$ and $f(x) = \sum_{n=0}^{\infty} \frac{1}{n^2 x + 1}$. Determine where f is continuous and where the convergence of the series is uniform. Note that $f_n(1/n^2) = \frac{1}{1+1} = \frac{1}{2} \not\to 0$, so f_n does not converge uniformly to 0 on S, and hence the sum does not converge uniformly. Let $0 < a < 1$, and consider f on $[a,1)$. Then $|f_n(x)| = \frac{1}{n^2 x + 1} \le \frac{1}{n^2 a + 1}$ for $x \in [a,1)$. But $\sum_{n=0}^{\infty} \frac{1}{n^2 a + 1} < \infty$, so by the Weierstrass M-test, $\sum_{n=0}^{\infty} f_n$ converges uniformly on $[a,1)$. Since each f_n is continuous on $[a,1)$, this implies f is continuous on $[a,1)$. Finally, $0 < a < 1$ is arbitrary, so $f(x) = \sum_{n=0}^{\infty} f_n(x)$ is continuous on $S = (0,1)$.

2.4.1 Power Series

Definition 2.20 An infinite sum of functions of the form

$$f(x) = \sum_{n=0}^{\infty} a_n (x-a)^n$$

is called a ***power series centered at** a*. One especially important family of power series are those of the form

$$\sum_{n=0}^{\infty} \frac{f^{(n)}(a)}{n!}(x-a)^n$$

where f is some infinitely differentiable function at a; this series is called the ***Taylor series for** f **at** a*.

A power series is a formal notion. It may not converge for many values of x. It always converges at $x = 0$ (a, its center).

Remark 2.7 Given a function f infinitely differentiable at a, we have for $x \in \mathbb{R}$ that

$$f(x) = \sum_{n=0}^{\infty} \frac{f^{(n)}(a)}{n!}(x-a)^n$$

if and only if the remainder terms satisfy $\lim\limits_{n\to\infty} R_{n,a}(x) = 0$.

Theorem 2.35 *Suppose that the series*

$$f(x_0) = \sum_{n=0}^{\infty} a_n x_0^n$$

converges, and let a be any number with $0 < a < |x_0|$. *Then on* $[-a, a]$ *the series*

$$f(x) = \sum_{n=0}^{\infty} a_n x^n$$

converges uniformly (and absolutely). Moreover, the same is true for the series

$$g(x) = \sum_{n=1}^{\infty} n a_n x^{n-1}$$

Finally, f is differentiable and

$$f'(x) = \sum_{n=1}^{\infty} n a_n x^{n-1}$$

for all x with $|x| < |x_0|$.

Proof First, since $\sum\limits_{n=0}^{\infty} a_n x_0^n$ converges, $\lim\limits_{n\to\infty} a_n x_0^n = 0$. Hence, the sequence $\{a_n x_0^n\}$ is surely bounded: there is some number M such that

$$|a_n x_0|^n = |a_n| \cdot |x_0|^n \le M$$

for all n. Now if x is in $[-a, a]$, then $|x| \le |a|$, so

$$|a_n x^n| = |a_n| \cdot |x|^n$$

$$
\begin{aligned}
&\leq |a_n| \cdot |a|^n \\
&= |a_n| \cdot |x_0|^n \cdot \left|\frac{a}{x_0}\right|^n \\
&\leq M \left|\frac{a}{x_0}\right|^n
\end{aligned}
$$

But $|a/x_0| < 1$, so the (geometric) series

$$\sum_{n=0}^{\infty} M \left|\frac{a}{x_0}\right|^n = M \sum_{n=0}^{\infty} \left|\frac{a}{x_0}\right|^n$$

converges. Choosing $M \cdot |a/x_0|^n$ as the number M_n in 2.33, it follows that $\sum\limits_{n=0}^{\infty} a_n x^n$ converges uniformly on $[-a, a]$.

To prove the same assertion for $g(x) = \sum\limits_{n=1}^{\infty} n a_n x^{n-1}$ notice that

$$
\begin{aligned}
|n a_n x^{n-1}| &= n|a_n| \cdot |x^{n-1}| \\
&\leq n|a_n| \cdot |a^{n-1}| \\
&= \frac{|a_n|}{|a|} \cdot |x_0|^n n \left|\frac{a}{x_0}\right|^n \\
&\leq \frac{M}{|a|} n \left|\frac{a}{x_0}\right|^n
\end{aligned}
$$

Since $|a/x_0| < 1$, the series

$$\sum_{n=1}^{\infty} \frac{M}{|a|} n \left|\frac{a}{x_0}\right|^n = \frac{M}{|a|} \sum_{n=1}^{\infty} n \left|\frac{a}{x_0}\right|^n$$

converges (by an application of the Ratio Tes). Another appeal to 2.33 proves that $\sum\limits_{n=1}^{\infty} n a_n x^{n-1}$ converges uniformly on $[-a, a]$.

Finally, our corollary proves, first that g is continuous, and then that

$$f'(x) = g(x) = \sum_{n=1}^{\infty} n a_n x^{n-1}$$

for all $x \in [-a, a]$. Since we could have chosen any a with $0 < a < |x_0|$, this result holds for all x with $|x| < |x_0|$. □

Theorem 2.36 (Ratio and Root Test) *If* $\sum_{n=0}^{\infty} a_n$ *for* $a_n \in \mathbb{C}$ *non-zero, then we have*

- *If* $\limsup \frac{|a_{n+1}|}{|a_n|} < 1$, *then* $\sum_n a_n$ *converges absolutely*
- *If* $\limsup \frac{|a_{n+1}|}{|a_n|} > 1$, *then* $\sum_n a_n$ *diverges*

- *If* $\limsup |a_n|^{\frac{1}{n}} < 1$*, then* $\sum_n a_n$ *converges absolutely*
- *If* $\limsup |a_n|^{\frac{1}{n}} > 1$*, then* $\sum_n a_n$ *diverges*

Example 2.17 If $a_n = 1$ for all n, we get the geometric series $\sum_{n=0}^{\infty} x^n = \frac{1}{1-x}$ provided $|x| < 1$.

The natural domain of a power series $f(x) = \sum_{n=0}^{\infty} a_n x^n$ is all x for which the sum converges.

Proposition 2.8 (Radius of Convergence) *Let* $\sum_n a_n x^n$ *be a power series and define* $\alpha := \limsup |a_n|^{1/n}$*. Define*

$$R = \begin{cases} \frac{1}{\alpha} & 0 < \alpha < \infty \\ \infty & \alpha = 0 \\ 0 & \alpha = \infty \end{cases}$$

which we call the **radius of convergence** *of the power series. Then* $\sum_n a_n x^n$ *converges absolutely for* $|x| < R$*. If* $R = 0$*, it converges only when* $x = 0$*. If* $R = \infty$*, it converges absolutely on all of* $\mathbb{R}$*. The series may or may not converge when* $|x| = R$*.*

Proof Let $b_n = a_n x^n$, and

$$\beta = \limsup |b_n|^{1/n} = \limsup |a_n|^{1/n} |x^n|^{1/n} = |x|\alpha$$

Apply the root test to $\sum_{n=0}^{\infty} b_n$:

(i) If $\alpha = 0$ then $\beta = 0 < 1$, so the series converges absolutely for all $x \in \mathbb{R}$

(ii)If $\alpha = \infty$, $\beta = 0$ if $x = 0$ and ∞ otherwise, so the series converges absolutely only at $x = 0$

(iii)If $0 < \alpha < \infty$, $\beta = |x|\alpha < 1$ if and only if $|x| < \frac{1}{\alpha} = R$, so the series converges absolutely for $|x| < R$ and diverges for $|x| > R$. □

Corollary 2.5 *If* $\lim\limits_{n\to\infty} \frac{|a_n|}{|a_{n+1}|}$ *exists, it equals R above.*

Example 2.18 Consider $\sum_{n=0}^{\infty} \frac{n! x^n}{n^n}$, so $a_n = \frac{n!}{n^n}$. Observe that

$$\lim_{n\to\infty} \frac{a_n}{a_{n+1}} = \lim_{n\to\infty} \frac{(n+1)^n}{n^n} = \lim_{n\to\infty} \left(1 + \frac{1}{n}\right)^n = e = R$$

For $x = \pm e$ we must use stirling's approximation $n! \approx \sqrt{2\pi n}\frac{n^n}{e^n}$, which becomes exact in the limit.

Proposition 2.9 *Suppose* $\sum_n a_n x^n$ *has radius of convergence* $R > 0$*. Then* $\sum_n a_n x^n$ *converges uniformly on any set of the form* $[-K, K]$ *for* $0 \le K < R$*.*

Proof Suppose $x \in [-K, K]$. Then $|a_n x^n| \le |a_n| K^n$. But $\sum_{n=0}^{\infty} |a_n| K^n$ converges since $K \in (-R, R)$. Thus, by the Weierstrass M test $\sum_{n=0}^{\infty} a_n x^n$ converges uniformly on $[-K, K]$. □

In particular, this tells us that $\sum_{n=0}^{\infty} a_n x^n$ is continuous on $[-K, K]$ for any $0 < K < R$, so it is continuous on $(-R, R)$.

Proposition 2.10 *Suppose* $f(x) = \sum_n a_n x^n$ *has radius of convergence* $R > 0$*. Then* f *is differentiable on* $(-R, R)$ *and* $f'(x) = \sum_n n a_n x^{n-1}$*. Moreover,* f' *also has radius of convergence* R*.*

Proof We have shown previously that if $f_n \to f$ pointwise, f_n' are continuous, and $f_n' \to_u g$ for some g, then f is differentiable and $f' = g$. Let $s_m(x) = \sum_{n=0}^{m} a_n x^n$, which converges uniformly to $\sum_{n=0}^{\infty} a_n x^n = f(x)$ on $[-K, K]$ for any $0 \leq K < R$. Note s_m is differentiable with $s_m'(x) = \sum_{n=1}^{m} n a_n x^{n-1}$, which converges uniformly to $\sum_{n=1}^{\infty} n a_n x^{n-1} =: g(x)$ on $[-K, K]$ for all $0 \leq K < R$ as $\frac{1}{R'} = \limsup \frac{|(n+1)a_{n+1}|}{|na_n|} = \limsup \frac{|a_{n+1}|}{|a_n|} = \frac{1}{R}$, so g's radius of convergence is also R. This implies that f is differentiable on $(-K, K)$ and $f'(x) = g(x)$ for all $x \in (-K, K)$. As this holds for all $0 \leq K < R$, this implies $f'(x) = g(x)$ on $(-R, R)$ □

Corollary 2.6 *The power series* $f(x) = \sum_n a_n x^n$ *with radius of convergence* $R > 0$ *has derivatives of all orders on* $(-R, R)$ *and* $f^{(k)}(x) = \sum_{n \geq k} \frac{n!}{(n-k)!} a_n x^{n-k}$ *on* $(-R, R)$*.*

Proof From the previous proposition we have $f'(x) = \sum_{n=1}^{\infty} n a_n x^{n-1}$ with radius of convergence R. Applying this result repeatedly we have $f^{(k)}(x) = \sum_{n=k}^{\infty} \frac{n!}{(n-k)!} a_n x^{n-k}$ has radius of convergence R. □

Note that in the formula above, it follows that $f^{(k)}(0) = k! a_k$, so $a_k = \frac{f^{(k)}(0)}{k!}$.

Corollary 2.7 *If* $\sum_n a_n x^n = \sum_n b_n x^n$ *for all* $x \in (-R, R)$*, where* R *is the radius of convergence of both series, then* $a_n = b_n$ *for all* n*.*

Using the remark above, letting $f(x) = \sum_n a_n x^n = \sum_n b_n x^n$, we have $a_n = \frac{f^{(n)}(0)}{n!} = b_n$ for all n.

Definition 2.21 (Power Series Representation) A function $f : S \to \mathbb{R}$ has a ***power series representation on S*** if there exists a sequence (a_n) such that

$$f(x) = \sum_{n=0}^{\infty} a_n x^n$$

for all $x \in S$. In particular, the series on the right hand side above converges for all $x \in S$.

We recall Taylor's Theorem:

Theorem 2.37 (Taylor's Theorem) *Suppose* f *is* $(n+1)$ *times differentiable on* (a, b) *containing* 0*, then*

$$f(x) = \underbrace{\sum_{k=0}^{n} \frac{f^{(k)}(0)}{k!} x^k}_{\text{Taylor Polynomial } P_n(x)} + \underbrace{\frac{f^{(n+1)}(c)}{(n+1)!} x^{n+1}}_{\text{Taylor remainder } R_n(x)}$$

for somce $c \in (0, x)$ (or $(x, 0)$). We get $|f(x) - P_n(x)| = |R_n(x)|$. If $R_n(x) \to 0$ as $n \to \infty$, then $f(x) = \lim\limits_{n\to\infty} P_n(x) = \sum_{k=0}^{\infty} \frac{f^{(k)}(0)}{k!} x^k$.

This result gives a sufficient condition for a function admitting a power series representation. Note that if f admits a power series representation, it automatically has derivatives of all orders on the radius of convergence.

Example 2.19 Consider

$$f(x) = \begin{cases} e^{-1/x^2} & x \neq 0 \\ 0 & x = 0 \end{cases}$$

f has derivatives of all orders at $x = 0$:

$$f'(0) = \lim_{h\to 0} \frac{e^{-1/h^2}}{h} = \lim_{h\to 0} \frac{2h}{e^{1/h^2}} = 0$$

Similarly, $f^{(k)}(0) = 0$ for all k. So, $\sum_{n=0}^{\infty} \frac{f^{(n)}(0)}{n!} x^n = 0$, which converges for all x but only agrees with $f(x)$ at $x = 0$. So f has only a power series representation at $x = 0$.

Theorem 2.38 (Abel's Theorem) *Suppose the power series $\sum_n a_n x^n$ has radius of convergence $R \in (0, \infty)$. If $\sum_n a_n x^n$ converges at $x = R$, then $\sum_n a_n x^n$ converges uniformly on $[0, R]$. Similarly, if $\sum_n a_n x^n$ converges at $x = -R$, then $\sum_n a_n x^n$ converges uniformly on $[-R, 0]$.*

Proof By rescaling if necessary, we can assume $R = 1$ (i.e. replace $f(x)$ with $f_R(x) = f(x/R)$) We use the Cauchy criterion for uniform convergence. Fix $\varepsilon > 0$. We have convergence at $x = 1$, so there exists $N \in \mathbb{N}$ such that $m \geq N$ implies $\left|\sum_{n=m}^{m+k} a_n\right| < \varepsilon$ for all $k \in \mathbb{N}$. For $m \geq N$ and for each $j \in \mathbb{N} \cup \{0\}$, define $s_j = \sum_{n=m}^{m+j} a_n$, so $|s_j| < \varepsilon$ for all j. Then write

$$\begin{aligned}
\sum_{n=m}^{m+k} a_n x^n &= s_0 x^m + \sum_{n=1}^{k} (s_n - s_{n-1}) x^{m+n} \\
&= \sum_{n=0}^{k-1} s_n (x^{m+n} - x^{m+n+1}) + s_k x^{m+k} \\
&= x^m \left[\sum_{n=0}^{k-1} s_n (x^n - x^{n+1}) + s_k x^k \right] \\
&= x^m \left[\sum_{n=0}^{k-1} s_n x^n (1 - x) + s_k x^k \right] \\
&= x^m (1 - x) \sum_{n=0}^{k-1} s_n x^n + s_k x^{m+k}
\end{aligned}$$

For $x \in [0, 1]$

$$\left| \sum_{n=m}^{m+k} a_n x^n \right| \leq |x|^m |1 - x| \sum_{n=0}^{k-1} |s_n| |x|^n + |s_k| |x|^{m+k}$$

$$
\begin{aligned}
&< x^m(1-x)[\varepsilon + \varepsilon x + \ldots + \varepsilon x^{k-1}] + \varepsilon x^{m+k} \\
&= \varepsilon x^m(1-x^k) + \varepsilon x^{m+k} \\
&= \varepsilon x^m < \varepsilon
\end{aligned}
$$

as desired. The proof for $[-1, 0]$ is similar. □

We now investigate power series in the complex numbers. First, recall that if $r \in \mathbb{C}$, $|r| < 1$, then we have the geometric series

$$\sum_{k=0}^{\infty} r^k = \frac{1}{1-r}$$

Definition 2.22 A ***power series*** (in $\mathbb{C}$) is a series of functions of the form

$$f(z) = \sum_{k=0}^{\infty} a_k z^k, \quad a_k \in \mathbb{C}, z \in \mathbb{C}$$

Note by definition $f(z) = \lim\limits_{n\to\infty} \sum_{k=0}^{n} a_k z^k$.

Proposition 2.11 *If* $f(z) = \sum_{k=0}^{\infty} a_k z^k$ *is defined for some* $z_0 \in \mathbb{C}, z_0 \neq 0$, *then* $f(z)$ *converges absolutely and uniformly for all* $|z| < |z_0|$.

Proof Note that if $|z| < |z_0|$, $|a_k z^k| < |a_k||z_0|^k$ for all $k \geq 0$. Further, as $\sum_{k=0}^{n} a_k z_0^k$ is convergent, it is Cauchy so $|a_k z^k| \to 0$, and in particular $|a_k z_0^k| \leq C$ for all $k \geq 0$ for some $C \in \mathbb{R}$. Then $0 \leq \frac{|z|}{|z_0|} = r < 1$. So

$$\left|\sum_{k=0}^{n} a_k z^k\right| \leq \sum_{k=0}^{n} |a_k z_0^k| r^k \leq C \sum_{k=0}^{n} r^k \leq \frac{C}{1-r}$$

Thus $\sum_{k=0}^{n} |a_k z^k|$ is a bounded increasing sequence, and so converges. In particular, letting $M_k = Cr^k = C\frac{|z_1|^k}{|z_0|^k}$, we have that for all $|z| < |z_1| < |z_0|$, $|a_k z^k| \leq M_k$ and $\sum_{k=0}^{\infty} M_k < \infty$, so $\sum_{k=0}^{n} a_k z^k$ is uniformly convergent to $f(z)$ on $|z| < |z_1|$ for all $|z_1| < |z_0|$, so in particular it is uniformly convergent on $|z| < |z_0|$. □

From this theorem we have a few cases for $f(z)$:

- $f(z)$ converges for all $z \in \mathbb{C}$
- There exists $R > 0$ such that the series converges absolutely for $|z| < R$, but diverges for $|z| > R$.
- The series does not converge for any $z \neq 0$ ($R = 0$)

Proposition 2.12 *If* $f(z) = \sum_{k=0}^{\infty} a_k z^k$ *converges in* $D_R = \{z \in \mathbb{C} : |z| < R\}$, *then it converges uniformly in any disk* D_S *with* $0 < S < R$. *In particular, this yields* $f(z)$ *is continuous on* D_R.

Proof Let $0 < S < R$ and pick T such that $0 < S < T < R$. Then the series converges for any $|z| = T$. In particular, there exists $C > 0$ such that $|a_k T^k| < C$ for all $k \geq 0$. If $z \in D_S$, then $|a_k z^k| \leq |a_k T^k| \frac{|z|^k}{|T|^k} \leq C r^k$, where $0 \leq r = \frac{|z|}{|T|} < 1$. As $|r| < 1$, $C \sum_{k=0}^{n} r^k$ converges, and so is Cauchy. Thus, for $\varepsilon > 0$, there exists $N \in \mathbb{N}$ such that if $m > n \geq N$,

$$\sum_{k=n}^{m} ||a_k z^k||_{D_S} \leq \sum_{k=n}^{m} C r^k < \varepsilon$$

so $\sum_{k=0}^{n} a_k z^k$ is uniformly Cauchy on D_S. So it converges uniformly to $f(z)$ in D_S for all $S < R$. So $f(z)$ is continuous in D_S for all $S < R$, and in particular f is continuous on D_R. □

Definition 2.23 For $f(z) = \sum_{k=0}^{\infty} (z - z_0)^k$, let the ***radius of convergence*** R be defined by

$$\frac{1}{R} = \limsup_{n \to \infty} |a_n|^{1/n}$$

If the right hand side is 0, $R = \infty$, and if the right hand side is ∞, $R = 0$.

Proposition 2.13 *The series $f(z) = \sum_{k=0}^{\infty} a_k (z - z_0)^k$ converges absolutely for $|z - z_0| < R$, and diverges when $|z - z_0| > R$. Then when $R > 0$, f is a continuous function $f : D_R(z_0) \to \mathbb{C}$.*

Proof If $R' < R$, and $R' \neq 0$, $\frac{1}{R} < \frac{1}{R'}$. Note $\frac{1}{R} = \limsup_{n \to \infty} |a_n|^{1/n}$, and let $\varepsilon = \left(\frac{1}{R'} - \frac{1}{R}\right)$. Then there exists $N \in \mathbb{N}$ such that for $n \geq N$, $\sup_{k \geq n} |a_k|^{1/k} < \frac{1}{R} + \varepsilon$, so $\sup_{k \geq n} |a_k|^{1/k} < \frac{1}{R'}$, so $|a_n|^{1/n} R' < 1$ and $|a_n| R'^n < 1$ for all $n \geq N$. If $|z - z_0| < R'$, then $\frac{|z - z_0|}{R'} = r < 1$, so $|a_n (z - z_0)^n| = |a_n| R'^n r^n < r^n$ for all $n \geq N$. Then $\sum_{n=m}^{'infty} ||a_n (z - z_0)||_{D_{R'}} \leq \sum_{n=m}^{\infty} r^n \to 0$ for all $m \geq N$. So we have absolute uniform convergence in $D_{R'}(z_0)$, for all $0 < R' < R$.

Divergence: Let $R'' > R$, so $\frac{1}{R''} < \frac{1}{R} = \limsup_{n \to \infty} |a_n|^{1/n}$. Then for all n, $\sup_{k \geq n} |a_n|^{1/n} > \frac{1}{R''}$, and in particular there are infinitely many $k \geq n$, for all n, such that $|a_k|^{1/k} > \frac{1}{R''}$, so $|a_k| R''^k > 1$. If $|z - z_0| \geq R''$, then $|a_n (z - z_0)^n| \geq |a_n| R''^n > 1$ for infinitely many $n \geq 0$. Thus, $|a_n (z - z_0)^n|$ does not converge to 0, so the series diverges. □

Definition 2.24 A function defined by $\sum_{k=0}^{\infty} a_k z^k$ with radius of convergence $R > 0$ is said to be ***analytic*** on D_R.

Definition 2.25 In $\mathbb{R}^n$ with $x = (x_1, ..., x_n)$, for any multi-index $\alpha = (\alpha_1, ..., \alpha_n)$ with $\alpha_j \in \mathbb{N}_0$, we define

$$x^\alpha = x_1^{\alpha_1} \cdots x_n^{\alpha_n}$$

and we define a power series in $\mathbb{R}^n$ by

$$f(x) = \sum_{|\alpha| \geq 0} a_\alpha x^\alpha$$

2.4.2 Products of Series

We investigate when we can distribute multiplication of two series:

Proposition 2.14 *If* $A = \sum_{k=0}^{\infty} a_k$ *and* $B = \sum_{k=0}^{\infty} b_k$ *are absolutely convergent, then*

$$AB = \sum_{k=0}^{\infty} c_k, \quad c_k = \sum_{j=0}^{k} a_{k-j} b_j$$

Proof Let $A_k = \sum_{n=0}^{k} a_n$ and $B_k = \sum_{n=0}^{k} b_n$. Then

$$\begin{aligned} A_k B_k &= \sum_{n=0}^{k} \sum_{m=0}^{k} a_n b_m \\ &= \sum_{l=0}^{k} \sum_{j=0}^{l} a_j b_{l-j} + R_l = \sum_{l=0}^{k} c_l + R_k \end{aligned}$$

where $R_k = \sum_{(n,m)\in\sigma(k)} a_n b_m$ where $\sigma(k) := \{(n,m) \in \mathbb{N}_0 \times \mathbb{N}_0 : n, m \leq k, n+m > k\}$. Then

$$\begin{aligned} |R_k| &\leq \sum_{(n,m)\in\sigma(k)} |a_n b_m| \\ &\leq \sum_{k\geq n\geq k/2} \sum_{m\leq k} |a_n||b_m| + \sum_{n\leq k} \sum_{k/2\leq m\leq k} |a_n||b_m| \\ &\leq \sum_{k/2\leq n\leq k} \sum_{m=0}^{\infty} |a_n||b_m| + \sum_{k/2\leq m\leq k} \sum_{n=0}^{\infty} |b_m||a_n| \\ &\leq A \sum_{k/2\leq m} |b_m| + B \sum_{k/2\leq n} |a_n| \end{aligned}$$

where both sums on the right go to zero as k goes to infinity. So for all $\varepsilon > 0$, there exists $N \in \mathbb{N}$ such that $k \geq N$ implies $|R_k| < \varepsilon$. Then

$$A_k B_k = \sum_{l=0}^{k} c_l + R_k$$

so

$$\lim_{k\to\infty} \sum_{l=0}^{k} c_l = \lim_{k\to\infty} (A_k B_k - R_k) = AB + 0 + AB$$

and we conclude

$$AB = \sum_{l=0}^{\infty} c_l$$

as claimed. □

Corollary 2.8 *If the series*

$$f(z) = \sum_{n=0}^{\infty} a_n (z - z_0)^n, \quad g(z) = \sum_{n=0}^{\infty} b_n (z - z_0)^n$$

converge for some $|z - z_0| < R$*, then*

$$f(z)g(z) = \sum_{n=0}^{\infty} c_n (z - z_0)^n$$

converges for $|z - z_0| < R$ *with*

$$c_n = \sum_{j=0}^{n} a_j b_{n-j}$$

It follows that the set of analytic functions on some disk is an algebra over $\mathbb{C}$.

Proposition 2.15 *If* $a_{jk} \in \mathbb{C}$ *and* $\sum_{j,k=0}^{\infty} |a_{j,k}| < \infty$*, then for each* k*,* $\alpha_k = \sum_{j=0}^{\infty} a_{j,k}$ *and for each* j $\beta_j = \sum_{k=0}^{\infty} a_{j,k}$ *are absolutely convergent, where*

$$\lim_{n\to\infty} \sum_{j=0}^{n} \sum_{k=0}^{n} |a_{j,k}| =: \sum_{j,k=0}^{\infty} |a_{j,k}|$$

Then

$$\sum_{j=0}^{\infty} \beta_j = \sum_{k=0}^{\infty} \alpha_k = \sum_{j,k=0}^{\infty} a_{j,k}$$

or equivalently

$$\sum_{j=0}^{\infty} \left(\sum_{k=0}^{\infty} a_{j,k} \right) = \sum_{k=0}^{\infty} \left(\sum_{j=0}^{\infty} a_{j,k} \right) = \sum_{j,k=0}^{\infty} a_{j,k}$$

Proof For all $\varepsilon > 0$ there exists $N \in \mathbb{N}$ such that

$$\sum_{j,k \geq n} |a_{j,k}| < \varepsilon$$

Thus for all $j \in \mathbb{N}$ and all $k \in \mathbb{N}$,

$$\sum_{k=0}^{\infty} |a_{j,k}| < \sum_{j,k} |a_{j,k}| < \infty$$

and

$$\sum_{j=0}^{\infty} |a_{j,k}| < \sum_{j,k} |a_{j,k}| < \infty$$

Then for all $M, K \geq N$,

$$\left| \sum_{k=0}^{M} \sum_{j=0}^{K} a_{j,k} - \sum_{j,k}^{N} a_{j,k} \right| < \sum_{j,k \geq N} |a_{j,k}| < \varepsilon$$

In particular, taking the limit as K goes to infinity,

$$\left|\sum_{k=0}^{M}\sum_{j=0}^{\infty} a_{j,k} - \sum_{j,k}^{N} a_{j,k}\right| \leq \sum_{j,k\geq N} |a_{j,k}| < \varepsilon$$

and

$$\left|\sum_{k=0}^{\infty}\sum_{j=0}^{\infty} a_{j,k} - \sum_{j,k}^{N} a_{j,k}\right| \leq \sum_{j,k\geq N} |a_{j,k}| < \varepsilon$$

Reversing the sums before taking the limits, we obtain the other direction, so

$$\sum_{j=0}^{\infty}\left(\sum_{k=0}^{\infty} a_{j,k}\right) = \sum_{j,k=0}^{\infty} a_{j,k} = \sum_{k=0}^{\infty}\left(\sum_{j=0}^{\infty} a_{j,k}\right)$$

as desired. □

Chapter 3

Integration

3.1 Introduction to Definite Integrals

Definition 3.1 Let $a < b$. A ***partition*** of the interval $I = [a, b]$ is a finite collection of points in $[a, b]$, one of which is a, and one of which is b. Equivalently, a partition of I into N subintervals consists pf endpoints $x_0 = a < x_1 < ... < x_N = b$ with subinterval k being $J_k = [x_k, x_{k+1}]$ for $0 \leq k \leq N - 1$.

The points in a partition can be numbered $t_0, ..., t_n$ so that

$$a = t_0 < t_1 < ... < t_{n-1} < t_n = b \tag{3.1.1}$$

we shall always assume that such a numbering has been assigned.

Definition 3.2 Suppose $f : I \to \mathbb{R}$ is bounded on $I = [a, b]$ so there exists $M \in \mathbb{R}^+$ such that $||f||_I \leq M$. Let $P = \{t_0, ..., t_N\} \in \prod([a, b])$ be a partition of $[a, b]$. The ***lower Riemann sum*** of f for P, denoted $L(f, P)$, is defined as

$$L(f, P) := \sum_{k=0}^{N-1} \inf_{J_k}(f)\ell(J_k) \tag{3.1.2}$$

The ***upper Riemann sum*** of f for P, denoted $U(f, P)$, is defined as

$$U(f, P) = \sum_{i=0}^{N-1} \sup_{J_k}(f)\ell(J_k) \tag{3.1.3}$$

where $\ell(J_k) = t_{k+1} - t_k > 0$. Note

$$-M \leq \sup_{J_k}(f) \leq M$$

and

$$-M \leq \inf_{J_k}(f) \leq M$$

Remark 3.1 If P is any partition, then

$$L(f,P) \leq U(f,P) \tag{3.1.4}$$

because

$$L(f,P) = \sum_{k=0}^{N-1} \inf_{J_k}(f)\ell(J_k)$$
$$U(f,P) = \sum_{i=0}^{N-1} \sup_{J_k}(f)\ell(J_k)$$

and for each i we have $\inf_{J_k}(f)\ell(J_k) \leq \sup_{J_k}(f)\ell(J_k)$.

Definition 3.3 Given two partitions P adn Q of I, we say P is a ***refinement*** of Q if every endpoint in Q belongs to P, and we write

$$P \succ Q$$

Lemma 3.1 *If P is a partition of $[a,b]$ which contains Q, that is $P \succ Q$, then*

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

Proof Consider first the special case in which Q contains just one more point than P;

$$P = \{t_0, ..., t_n\}$$
$$Q = \{t_0, ..., t_{k-1}, u, t_k, ..., t_n\}$$

where

$$a = t_0 < t_1 < ... < t_{k-1} < u < t_k < ... < t_n = b$$

Let

$$m' = \inf\{f(x) : t_{k-1} \leq x \leq u\}$$
$$m'' = \inf\{f(x) : u \leq x \leq t_k\}$$

Then

$$L(f,P) = \sum_{i=1}^{n} m_i(t_i - t_{i-1})$$
$$L(f,Q) = \sum_{i=1}^{k-1} m_i(t_i - t_{i-1}) + m'(u - t_{k-1}) + m''(t_k - u) + \sum_{i=k+1}^{n} m_i(t_i - t_{i-1})$$

To prove that $L(f,P) \leq L(f,Q)$ it therefore suffices to show that

$$m_k(t_k - t_{k-1}) \leq m'(u - t_{k-1}) + m''(t_k - u)$$

Now, the set $\{f(x) : t_{k-1} \leq x \leq t_k\}$ contains all the numbers in $\{f(x) : t_{k-1} \leq x \leq u\}$ and possibly some smaller ones, so the greatest lower bound of the first set is less than or equal to the greatest lower bound of the second; thus

$$m_k \leq m'$$

Similarly,

$$m_k \leq m''$$

Therefore,

$$m_k(t_k - t_{k-1}) = m_k(t_k - u) + m_k(u - t_{k-1}) \leq m''(t_k - u) + m'(u - t_{k-1})$$

This proves, in this special case that $L(f, P) \leq L(f, Q)$. Now, let

$$\begin{aligned} M' &= \sup\{f(x) : t_{k-1} \leq x \leq u\} \\ M'' &= \sup\{f(x) : u \leq x \leq t_k\} \end{aligned}$$

Then

$$\begin{aligned} U(f, P) &= \sum_{i=1}^{n} M_i(t_i - t_{i-1}) \\ U(f, Q) &= \sum_{i=1}^{k-1} M_i(t_i - t_{i-1}) + M'(u - t_{k-1}) + M''(t_k - u) + \sum_{i=k+1}^{n} M_i(t_i - t_{i-1}) \end{aligned}$$

Hence, to prove that $U(f, Q) \leq U(f, P)$ it suffices to show that

$$M'(u - t_{k-1}) + M''(t_k - u) \leq M_k(t_k - t_{k-1})$$

As before, the set $\{f(x) : t_{k-1} \leq x \leq t_k\}$ contains all the numbers in $\{f(x) : t_{k-1} \leq x \leq u\}$ and possibly some larger ones, so the smallest upper bound of the first set is greater than or equal to the smallest upper bound of the second; thus

$$M_k \geq M'$$

Similarly,

$$M_k \geq M''$$

Therefore,

$$M_k(t_k - t_{k-1}) = M_k(t_k - u) + M_k(u - t_{k-1}) \geq M''(t_k - u) + M'(u - t_{k-1})$$

This proves, in this special case that $U(f, P) \geq U(f, Q)$.

The general case can now be deduced quite easily. The partition Q can be obtained from P by adding one point at a time; in otherwords, there is a sequence of partition

$$P = P_1 \subsetneq P_2 \subsetneq P_3 \subsetneq \ldots \subsetneq P_\alpha = Q$$

such that $P_{j+1} = P_j \cup \{u_{j+1}\}$ for some $u_{j+1} \in [a, b] - P_j$. Then

$$L(f, P) = L(f, P_1) \leq L(f, P_2) \leq \ldots \leq L(f, P_\alpha) = L(f, Q)$$

and

$$U(f,P) = U(f,P_1) \geq U(f,P_2) \geq \ldots \geq U(f,P_\alpha) = U(f,Q)$$

completing the proof. □

Theorem 3.1 *Let* P_1 *and* P_2 *be partitions of* $[a,b]$*, and let* f *be a function which is bounded on* $[a,b]$*. Then*

$$L(f,P_1) \leq U(f,P_2) \tag{3.1.5}$$

Proof There is a partition P which contains both P_1 and P_2 (let $P = P_1 \cup P_2$). According to the lemma

$$L(f,P_1) \leq L(f,P) \leq U(f,P) \leq U(f,P_1)$$

Definition 3.4 Let $\prod(I)$ denote the set of all partitions of I. We define the ***upper*** and ***lower Riemann integrals*** by

$$U_I(f) = \inf_{P\in\prod(I)} U(f,P) \geq L_I(f) = \sup_{P\in\prod(I)} L(f,P)$$

Definition 3.5 (Definite Integral) A function f which is bounded on $[a,b]$ is said to be ***Riemann integrable*** on $[a,b]$ if and only if

$$L_I(f) = U_I(f)$$

In this case, this common number is called the ***integral*** of f on $[a,b]$ and is denoted by

$$\int_I f = \int_a^b f(x)dx = L_I(f) = U_I(f) \tag{3.1.6}$$

The integral $\int_I f$ is also called the ***area*** of $R(f,a,b)$ when $f(x) \geq 0$ for all $x \in [a,b]$.

Theorem 3.2 *If* f *is bounded on* $[a,b]$*, then* f *is integrable on* $[a,b]$ *if and only if for every* $\epsilon > 0$ *there is a partition P of* $[a,b]$ *such that*

$$U(f,P) - L(f,P) < \epsilon$$

Proof Suppose first that for every $\epsilon > 0$ there is such a partition P. Since

$$U_I(f) \leq U(f,P)$$
$$L_I(f) \geq L(f,P)$$

it follows that

$$U_I(f) - L_I(f) \leq U(f,P) - L(f,P) < \epsilon$$

Since this is true for all $\epsilon > 0$, it follows that

$$U_I(f) = L_I(f)$$

so by definition, then, f is integrable. Conversely, if f is integrable then

$$U_I(f) = L_I(f)$$

Let M denote the value of this. Then for each $\epsilon > 0$ there exist partitions P' and P'' such that $|U(f,P') - M| < \epsilon/2$ and $|L(f,P'') - M| < \epsilon/2$. Then as $U(f,P') \geq L(f,P'')$ from the previous theorem, we have that

$$U(f,P') - L(f,P'') = |U(f,P') - L(f,P'')| \leq |U(f,P') - M| + |M - L(f,P'')| < \epsilon$$

Let $P = P' \cup P''$ be a common refinement. Then, according to the lemma $U(f,P) \leq U(f,P')$ and $L(f,P) \geq L(f,P'')$ so

$$U(f,P) - L(f,P) \leq U(f,P'') - L(f,P') < \epsilon$$

Example 3.1 Define $f(x) = \begin{cases} 1 & x \in \mathbb{Q} \cap [0,1] \\ 0 & x \notin \mathbb{Q} \cap [0,1] \end{cases}$ Since both $S = \mathbb{Q} \cap [0,1]$ and $[0,1] \backslash S$ are dense in $[0,1]$, for any subinterval J_k we have $\sup_{J_k}(f) = 1$ and $\inf_{J_k}(f) = 0$ so $L_I(f) = 0 < 1 = U_I(f)$. Thus f is not Riemann integrable, but it is Lebesque integrable since $\mathbb{Q}$ is a set of measure 0.

Theorem 3.3 *If $f \in C(I)$ is continuous on $I = [a,b]$, then $f \in \mathcal{R}(I)$ is integrable on $I = [a,b]$.*

Proof Notice, first, that f is bounded on $[a,b]$, because it is continuous on $[a,b]$. To prove that f is integrable on $[a,b]$, we want to use our previous theorem, and show that for every $\epsilon > 0$ there is a partition P of $[a,b]$ such that

$$U(f,P) - L(f,P) < \epsilon$$

Now we know, by our result on uniform continuity, that f is uniformly continuous on $[a,b]$. So there is some $\delta > 0$ such that for all $x, y \in [a,b]$, if $|x - y| < \delta$, then $|f(x) - f(y)| < \epsilon/[2(b-a)]$. We choose a partition $P = \{t_0, ..., t_N\}$ such that each $|t_{k+1} - t_k| < \delta$. Then for each subinterval J_k we have

$$|f(x) - f(y)| < \frac{\epsilon}{2(b-a)}$$

for all $x, y \in J_k$. Then, $f(x) < \frac{\epsilon}{2(b-a)} + f(y)$ for all $x \in J_k$, so $\sup_{J_k}(f) \leq \frac{\epsilon}{2(b-a)} + f(y)$. Further, $f(y) \geq \sup_{J_k}(f) - \frac{\varepsilon}{2(b-a)}$ for all $y \in J_k$, so $\inf_{J_k}(f) \geq \sup_{J_k}(f) - \frac{\varepsilon}{2(b-a)}$, so $\sup_{J_k}(f) - \inf_{J_k}(f) \leq \frac{\varepsilon}{2(b-a)}$. Since this is true for all k, we have that

$$\begin{aligned} U(f,P) - L(f,P) &= \sum_{k=0}^{N-1} (\sup_{J_k}(f) - \inf_{J_k}(f))\ell(J_k) \\ &< \frac{\epsilon}{b-a} \sum_{k=0}^{N-1} \ell(J_k) \\ &= \frac{\epsilon}{b-a} \ell(I) \\ &= \epsilon \end{aligned}$$

Thus, by our previous theorem f is integrable. □

Proposition 3.1 *$\mathcal{R}(I)$ is a $\mathbb{R}$-vector space. If $f, g \in \mathcal{R}(I)$, $a \in \mathbb{R}$, then $af + g \in \mathcal{R}(I)$, and*

$$\int_I (af + g) = a \int_I f + \int_I g$$

Proof First we do additivity. Let $J_k \subseteq I$. Then $\sup_{J_k}(f+g) \leq \sup_{J_k}(f)+\sup_{J_k}(g)$ and $\inf_{J_k}(f+g) \geq \inf_{J_k}(f) + \inf_{J_k}(f)$. So it follows that $U(f+g,P) \leq U(f,P) + U(g,P)$ for any $P \in \prod(I)$ and $L(f+g,P) \geq L(f,P) + L(g,P)$. Then we have

$$L_I(f) + L_I(g) \leq L_I(f+g) \leq U_I(f+g) \leq U_I(f) + U_I(g)$$

where by assumption $L_I(f) = U_I(f)$ and $L_I(g) = U_I(g)$, so $U_I(f+g) = L_I(f+g)$, and $f+g \in \mathcal{R}(I)$. Further, $\int_I (f+g) = U_I(f+g) = U_I(f) + U_I(g) = \int_I f + \int_I g$.

Next, let $a \in \mathbb{R}$. If $a = 0$, then $af = 0$ so $U(af,P) = 0 = L(af,P)$ for all partitions P of $[a,b]$, and so $U_I(af) = 0 = L_I(af)$. Thus, $af \in \mathcal{R}(I)$, and $\int_I af = 0 = a\int_I f$. Next, if $a > 0$, then $U(af,P) = aU(f,p)$ and $L(af,P) = aL(f,P)$. Then $U_I(af) = aU_I(f) = aL_I(f) = L_I(af)$, and we have our desired result. On the other hand, if $a < 0$, $U(af,P) = aL(f,P)$, and $L(af,P) = aU(f,P)$. Further, $U_I(af) = aL_I(f) = aU_I(f) = L_I(af)$, so we again obtain our desired result. □

Theorem 3.4 *Let $a < c < b$. If f is integrable on $[a,b]$, then f is integrable on $[a,c]$ and one $[c,b]$. Conversely, if f is integrable on $[a,c]$ and on $[c,b]$, then f is integrable on $[a.b]$. Finally, if f is integrable on $[a,b]$, then*

$$\int_a^b f = \int_a^c f + \int_c^b f \tag{3.1.7}$$

Proof (1) Suppose f is integrable on $[a,b]$. Then f is bounded on $[a,b]$, so it is bounded on $[a,c]$ and $[c,b]$. Indeed, f being bounded implies that there exists $M \in \mathbb{R}$ such that for all $x \in [a,b]$ $|f(x)| \leq M$. Thus, as this applies for all $x \in [a,b]$ and $[a,c], [c,b] \subset [a,b]$, we have that it holds for all $x \in [a,c]$ and all $x \in [c,b]$. Now fix $\epsilon > 0$. Then there exists a partition P of $[a,b]$ such that

$$U(f,P) - L(f,P) < \epsilon$$

Without loss of generality suppose $c = t_j$ for some $t_j \in P = \{t_0, t_1, ..., t_n\}$. Then we have partitions $P' = \{t_0, ..., t_j\}$ and $P'' = \{t_j, ..., t_n\}$ for $[a,c]$ and $[c,b]$ respectively. Moreover,

$$U(f,P) = U(f,P') + U(f,P'')$$
$$L(f,P) = L(f,P') + L(f,P'')$$

Hence, we have that

$$[U(f,P') - L(f,P')] + [U(f,P'') - L(f,P'')] = U(f,P) - L(f,P) < \epsilon$$

But $U(f,P') \geq L(f,P')$ and $U(f,P'') \geq L(f,P'')$, so

$$U(f,P') - L(f,P') \leq U(f,P) - L(f,P) < \epsilon$$
$$U(f,P'') - L(f,P'') \leq U(f,P) - L(f,P) < \epsilon$$

Therefore, f is integrable on $[a,c]$ and $[c,b]$

(2) Suppose f is integrable on $[a,c]$ and $[c,b]$. Thus, there exists $M_1, M_2 \in \mathbb{R}$ such that for all $x \in [a,c]$ $|f(x)| \leq M_1$ and for all $x \in [c,b]$ $|f(x)| \leq M_2$. Let $M = \max(M_1, M_2)$. Then for all $x \in [a,b]$ we have $|f(x)| \leq M$, so f is bounded on $[a,b]$. Let $\epsilon > 0$. Then there exist partitions P_1, P_2 of $[a,c]$ and $[c,b]$ respectively such that

$$U(f,P_1) - L(f,P_1) < \epsilon/2$$
$$U(f,P_2) - L(f,P_2) < \epsilon/2$$

Let $P = P_1 \cup P_2$, where $P_1 \cap P_2 = \{c\}$. Then we have that

$$U(f,P) - L(f,P) = [U(f,P_1) - L(f,P_1)] + [U(f,P_2) - L(f,P_2)] < \epsilon/2 + \epsilon/2 = \epsilon$$

Therefore, by definition f is integrable on $[a,b]$.

(3) Suppose f is integrable on $[a,b]$, so by the previous results f is integrable on $[a,c]$ and $[c,b]$. Let $\int_a^b f = R$, $\int_a^c f = R_1$, and $\int_c^b f = R_2$. Let P be a partition of $[a,b]$, and without loss of generality suppose $c \in P = \{t_0, ..., t_j = c, ..., t_n\}$. Then let $P_1 = \{t_0, ..., t_j\}$ and $P_2 = \{t_j, ..., t_n\}$ be partitions of $[a,c]$ and $[c,b]$. It then follows that

$$L(f,P_1) \leq R_1 \leq U(f,P_1)$$
$$L(f,P_2) \leq R_2 \leq U(f,P_2)$$

Hence, we have that

$$L(f,P) = L(f,P_1) + L(f,P_2) \leq R_1 + R_2$$
$$U(f,P) = U(f,P_1) + U(f,P_2) \geq R_1 + R_2$$

Thus $L(f,P) \leq R_1 + R_2 \leq U(f,P)$. Note that this holds for all partitions P, as if P' is a partition, then considering the partition $P'_c = P' \cup \{c\}$ we have that

$$L(f,P') \leq L(f,P'_c) \leq R_1 + R_2 \leq U(f,P'_c) \leq U(f,P')$$

Therefore, this holds for all partitions of $[a,b]$, but R is the unique number which does this so we must have that $R = R_1 + R_2$. Thus

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

Definition 3.6 Using the previous theorem, we define

$$\int_a^a f := 0 \quad and \quad \int_a^b f := -\int_b^a f, \ for\ a > b \tag{3.1.8}$$

Theorem 3.5 *Suppose f is integrable on $[a,b]$ and that*

$$m \leq f(x) \leq M \forall x \in [a,b] \tag{3.1.9}$$

Then

$$m(b-a) \leq \int_a^b f \leq M(b-a) \tag{3.1.10}$$

Proof It is clear that $M \geq \sup\{f(x) : x \in [a,b]\}$ and $m \leq \inf\{f(x) : x \in [a,b]\}$, so

$$m(b-a) \leq L(f,P) \ and \ M(b-a) \geq U(f,P)$$

for every partition P. Since $\int_a^b f = \sup\{L(f,P)\} = \inf\{U(f,P)\}$ we have that

$$m(b-a) \leq \sup\{L(f,P)\} = \int_a^b f = \inf\{U(f,P)\} \leq M(b-a)$$

If f is integrable on $[a,b]$, we can define a new function F on $[a,b]$ by

$$F(x) = \int_a^x f = \int_a^x f(t)dt \tag{3.1.11}$$

Theorem 3.6 *If f is integrable on $[a,b]$ and F is defined on $[a,b]$ by*

$$F(x) = \int_a^x f$$

then F is continuous on $[a,b]$.

Proof Suppose $c \in [a,b]$. Since f is integrable on $[a,b]$ it is, by definition, bounded on $[a,b]$; let M be a number such that

$$|f(x)| \leq M, \forall x \in [a,b]$$

If $h > 0$ for $c+h \in [a,b]$, then

$$F(c+h) - F(c) = \int_a^{c+h} f - \int_a^c f = \int_c^{c+h} f$$

Since $-M \leq f(x) \leq M$ for all $x \in [a,b]$ it follows from the Theorem 3.5 that

$$-M \cdot h \leq \int_c^{c+h} f \leq M \cdot h$$

In other words

$$-M \cdot h \leq F(c+h) - F(c) \leq M \cdot h$$

If $h < 0$, with $c+h \in [a,b]$ we find the inequality

$$-M \cdot h \geq F(c+h) - F(c) \geq M \cdot h$$

In either case we have that

$$|F(c+h) - F(c)| \leq M \cdot |h|$$

Therefore, if $\epsilon > 0$, we have

$$|F(c+h) - F(c)| < \epsilon$$

provided that $|h| < \epsilon/M$. This proves that

$$\lim_{h \to 0} F(c+h) = F(c)$$

so F is continuous at c. □

3.2 Reimann Sums

Definition 3.7 Suppose $P = \{t_0, ..., t_N\}$ is a partition of $[a, b]$, and for each k choose $\xi_k \in J_k$. Then we clearly have that

$$L(f, P) \leq \sum_{k=0}^{N-1} f(\xi_k)\ell(J_k) \leq U(f, P) \tag{3.2.1}$$

Any sum of the form

$$\sum_{k=0}^{N-1} f(\xi_k)\ell(J_k) \tag{3.2.2}$$

is called a ***Reimann sum*** of f for P.

Before proving an important result related to Riemann sums and definite integrals, we handle the following proposition:

Proposition 3.2 *Suppose $f : [a, b] \longrightarrow \mathbb{R}$ is bounded by $M \in \mathbb{R}^+$. If P and Q are partitions of $[a, b]$ such that for some $k > 1$*

$$maxsize(P) \leq \frac{minsize(Q)}{k}$$

then

$$U(f, P) \leq U(f, Q) + \frac{2M\ell([a, b])}{k}$$

and

$$L(f, P) \geq L(f, Q) - \frac{2M\ell([a, b])}{k}$$

Proof Let $P_1 = P \cup Q$ be the common refinement of P and Q. The intervals in P, denoted $\tilde{J}_i$, can be separated into two classes,

(i) $\tilde{J}_i \subseteq J_k$ for some k, where J_k are the intervals in Q

(ii) $\tilde{J}_i \not\subseteq J_k$ for all k

Note that in the case of (i), $\tilde{J}_i$ also belongs to P_1. In the second case, $\tilde{J}_i$ is split into two intervals in P_1. Then

$$|U(f, P) - U(f, P_1)| \leq \sum_{\text{intervals in (ii)}} |\sup_{\tilde{J}_i}(f)\ell(\tilde{J}_i) - (\sup_{\tilde{J}_i^+}(f)\ell(\tilde{J}_i^+) + \sup_{\tilde{J}_i^-}(f)\ell(\tilde{J}_i^-))|$$

where the intervals in class (i) cancel since they belong to both partitions. If $\tilde{J}_i$ is in class (ii), then $\tilde{J}_i = \tilde{J}_i^+ \cup \tilde{J}_i^-$ in P_1. Then proceeding with the inequality

$$\begin{aligned} |U(f, P) - U(f, P_1)| &\leq M \sum_{\in(ii)} \ell(\tilde{J}_i) + \ell(\tilde{J}_i^+) + \ell(\tilde{J}_i^-) \\ &= 2M \sum_{\in(ii)} \ell(\tilde{J}_i) \end{aligned}$$

If $\tilde{J}_i$ is in class (ii), then there exists a unique endpoint x_l in Q such that $x_l \in \tilde{J}_i^\circ$ (the interior of the interval) so

$$\ell(\tilde{J}_i) \leq \frac{\ell(J_l)}{k}$$

Now

$$|U(f,P) - U(f,P_1)| \leq 2M \sum_{\in(ii)} \ell(\tilde{J}_i) \leq 2M \sum_{\in(ii)} \frac{\ell(J_l)}{k} \leq \frac{2M\ell(I)}{k}$$

As P_1 refines Q, $U(f,P_1) \leq U(f,Q)$, so

$$U(f,P_1) \leq U(f,P) \leq U(f,P_1) + \frac{2M\ell(I)}{k} \leq U(f,Q) + \frac{2M\ell(I)}{k}$$

Similarly, $L(f,P_1)$ refines Q so $L(f,P_1) \geq L(f,Q)$, and by the same chain as above $|L(f,P) - L(f,P_1)| \leq \frac{2M\ell(I)}{k}$, so

$$L(f,P_1) \geq L(f,P) \geq L(f,P_1) - \frac{2M\ell(I)}{k} \geq L(f,Q) - \frac{2M\ell(I)}{k}$$

as desired. □

Now we move on to our fundamental result:

Theorem 3.7 (Darboux's Theorem) *Let $f : [a,b] \to \mathbb{R}$ be bounded. Let $P_\nu \in \prod([a,b])$ be a sequence of partitions of $I = [a,b]$ such that*

$$\lim_{\nu\to\infty} maxsize(P_\nu) = 0$$

Then

$$U_I(f) = \lim_{\nu\to\infty} U(f,P_\nu)$$

and

$$L_I(f) = \lim_{\nu\to\infty} L(f,P_\nu)$$

In particular, $f \in \mathbb{R}([a,b])$ if and only if

$$\int_I f = \lim_{\nu\to\infty} \sum_{k=0}^{N_\nu - 1} f(\xi_{\nu,k})\ell(J_{\nu,k})$$

where $P_\nu = (a = x_{\nu,0} < \ldots < x_{\nu,N_\nu} = b)$, $J_{\nu,k} = [x_{\nu,k}, x_{\nu,k+1}]$, and where $\xi_{\nu,k}$ is an arbitrary point of $J_{\nu,k}$.

Proof Let $P_\nu \in \prod([a,b])$ be a sequence of partition such that $\lim_{\nu\to\infty} \text{maxsize} P_\nu = 0$. Note $U_I(f) = \inf_{P\in\prod([a,b])} U(f,P)$, which exists and is well defined if f is bounded, and similarly for $L_I(f) = \sup_{P\in\prod([a,b])} L(f,P)$. Given $\varepsilon > 0$, there exists $Q \in \prod(I)$ such that $U_I(f) + \varepsilon \geq U(f,Q) \geq U_I(f)$ and taking a refinement if needed, $L_I(f) - \varepsilon \leq L(f,Q) \leq L_I(f)$. Let $\nu, \mu \geq N$ such that

$$\text{maxsize} P_\nu \leq \varepsilon \text{minsize} Q, \ \forall \nu \geq N$$

In particular, we can choose $\varepsilon = \frac{1}{k}$, $k \in \mathbb{N}$, $k > 1$. By the previous proposition

$$U(f, P_\nu) \leq U(f, Q) + \varepsilon 2M\ell(I)$$
$$L(f, P_\nu) \geq L(f, Q) - \varepsilon 2M\ell(I)$$

Then $U(f, P_\nu) \leq U_I(f) + \varepsilon(2M\ell(I) + 1)$ and $L(f, P_\nu) \geq L_I(f) - \varepsilon(2M\ell(I) + 1)$. Then

$$|U(f, P_\nu) - U_I(f)| \leq \varepsilon(2M\ell(I) + 1)$$

and

$$|L_I(f) - L(f, P_\nu)| \leq \varepsilon(2M\ell(I) + 1)$$

for all $\nu \geq N$. As we can make ε as small as we wish,

$$\lim_{\nu\to\infty} U(f, P_\nu) = U_I(f) \text{ and } \lim_{\nu\to\infty} L(f, P_\nu) = L_I(f)$$

Then $f \in \mathcal{R}(I)$ if and only if $U_I(f) = L_I(f)$, which occurs if and only if $\lim\limits_{\nu\to\infty} U(f, P_\nu) = \lim\limits_{\nu\to\infty} L(f, P_\nu)$. We observe that these last limits are equal if and only if $\lim\limits_{\nu\to\infty} \sum_{k=0}^{N_\nu - 1} f(\xi_{\nu,k})\ell(J_{\nu,k})$ exists whenever maxsize$P_\nu \to 0$ for any choices of $\xi_{\nu,k} \in J_{\nu,k}$, as

$$L(f, P_\nu) \leq \sum_{k=0}^{N_\nu - 1} f(\xi_{\nu,k})\ell(J_{\nu,k}) \leq U(f, P_\nu)$$

3.3 The Fundamental Theorem of Calculus

Theorem 3.8 (The Fundamental Theorem of Calculus Part 1) *Let $f \in \mathcal{R}([a, b])$ be integrable on $[a, b]$, and define F on $[a, b]$ by*

$$F(x) = \int_a^x f = \int_a^x f(t)dt \tag{3.3.1}$$

If f is continuous at c in $[a, b]$, then F is differentiable at c, and

$$F'(c) = f(c) \tag{3.3.2}$$

(if $c = a$ or b, then $F'(c)$ is understood to mean the right or left hand derivative of F).

Proof First, consider $c \in (a, b)$. By definition,

$$F'(c) = \lim_{h\to 0} \frac{F(c + h) - F(c)}{h}$$

Suppose first that $h > 0$. Then

$$F(c + h) - F(c) = \int_c^{c+h} f$$

Define m_h and M_h as follows:

$$m_h = \inf\{f(x) : c \leq x \leq c+h\}$$
$$M_h = \sup\{f(x) : c \leq x \leq c+h\}$$

It follows from Theorem 3.5 that

$$m_h \cdot h \leq \int_c^{c+h} f \leq M_h \cdot h$$

Therefore,

$$m_h \leq \frac{F(c+h)-F(c)}{h} \leq M_h$$

If $h < 0$, let

$$m_h = \inf\{f(x) : c+h \leq x \leq c\}$$
$$M_h = \sup\{f(x) : c+h \leq x \leq c\}$$

It follows from Theorem 3.5 that

$$m_h \cdot (-h) \leq \int_{c+h}^{c} f \leq M_h \cdot (-h)$$

Since

$$F(c+h)-F(c) = \int_c^{c+h} f = -\int_{c+h}^{c} f$$

this yields

$$m_h \cdot h \geq F(c+h)-F(c) \geq M_h \cdot h$$

Since $h < 0$, we have that

$$m_h \leq \frac{F(c+h)-F(c)}{h} \leq M_h$$

This inequality is true for any integrable function, continuous or not. Since f is continuous at c, however,

$$\lim_{h\to 0} m_h = \lim_{h\to 0} M_h = f(c)$$

and this proves that

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

Now, if $c = a$ we need only look at when $h > 0$, and in this case we still have

$$m_h \leq \frac{F(a+h)-F(a)}{h} \leq M_h$$

and from our previous limits,

$$\lim_{h\to 0^+} m_h = \lim_{h\to 0^+} m_h = f(a)$$

thus we have that

$$F'(a) = \lim_{h \to 0^+} \frac{F(a+h) - F(a)}{h} = f(a)$$

Similarly, if $c = b$ we need only look at $h < 0$, so we have that

$$\lim_{h \to 0^-} m_h = \lim_{h \to 0^-} m_h = f(b)$$

and

$$F'(b) = \lim_{h \to 0^-} \frac{F(b+h) - F(b)}{h} = f(b)$$

completing the proof. □

We may consider

$$F(x) = \int_a^x f \tag{3.3.3}$$

when $x < a$. In this case we have that

$$F(x) = -\int_x^a f = -\left(\int_b^a f - \int_b^x f\right) \tag{3.3.4}$$

so for $c \in [a, b]$,

$$F'(c) = -(-f(c)) = f(c) \tag{3.3.5}$$

as before.

Theorem 3.9 (Fundamental Theorem of Calculus Part 2) *Suppose* G *is continuous in* $[a, b]$ *and differentiable in* (a, b)*, with* $G' \in \mathcal{R}([a, b])$*. Then*

$$\int_a^b G'(t)dt = G(b) - G(a)$$

Proof Let G be as described. Let P_n be a partition with endpoints $a + \frac{b-a}{n}k$ for $0 \leq k \leq n$. By Darboux's theorem, for arbitrary $\xi_{n,k} \in J_{n,k} = \left[a + \frac{b-a}{n}k, a + \frac{b-a}{n}(k+1)\right]$,

$$\int_a^b G'(t)dt = \lim_{n \to \infty} \sum_{k=0}^{n-1} G'(\xi_{n,k})\ell(J_{n,k}) = \lim_{n \to \infty} \sum_{k=0}^{n-1} G'(\xi_{n,k})\frac{b-a}{n}$$

Now observe we have the telescopic sum

$$G(b) - G(a) = G(x_n) - G(x_0) = \sum_{k=0}^{n-1} G(x_{k+1}) - G(x_k)$$

As G is continuous on $[a, b]$ and differentiable on (a, b), we have by the mean value theorem that there exists $\xi^*_{n,k} \in (x_k, x_{k+1})$ such that

$$G(x_{k+1}) - G(x_k) = G'(\xi^*_{n,k})(x_{k+1} - x_k)$$

Then let the arbitrary $\xi_{n,k}$ be the $\xi^*_{n,k}$, so

$$\begin{aligned}\int_a^b G'(t)dt &= \lim_{n\to\infty}\sum_{k=0}^{n-1} G'(\xi^*_{n,k})\frac{b-a}{n}\\ &= \lim_{n\to\infty}\sum_{k=0}^{n-1} G(x_{k+1})-G(x_k)\\ &= \lim_{n\to\infty}(G(b)-G(a)) = G(b)-G(a)\end{aligned}$$

as desired. □

It is important to note that this is merely a useful result for certain functions f, ***not*** a definition.

If f is any bounded function on $I=[a,b]$, then

$$L_I(f) \;\; and \;\; U_I(f) \tag{3.3.6}$$

will both exist. These numbers are called the ***lower integral*** of f on $[a,b]$ and the ***upper integral*** of f on $[a,b]$, respectively, and will sometimes be denoted by

$$\mathbf{L}\int_a^b f \;\; and \;\; \mathbf{U}\int_a^b f \tag{3.3.7}$$

If $a<c<b$, then

$$\mathbf{L}\int_a^b f = \mathbf{L}\int_a^c f + \mathbf{L}\int_c^b f \;\; and \;\; \mathbf{U}\int_a^b f = \mathbf{U}\int_a^c f + \mathbf{U}\int_c^b f \tag{3.3.8}$$

and if $m \le f(x) \le M$ for all $x \in [a,b]$, then

$$m(b-a) \le \mathbf{L}\int_a^b f \le \mathbf{U}\int_a^b f \le M(b-a) \tag{3.3.9}$$

f is integrable precisely when

$$\mathbf{L}\int_a^b f = \mathbf{U}\int_a^b f \tag{3.3.10}$$

We shall now demonstrate an alternate proof for the following theorem stated previously.

Theorem 3.10 *If f is continuous on $[a,b]$, then f is integrable on $[a,b]$.*

Proof Define function L and U on $[a,b]$ by

$$L(x) = \mathbf{L}\int_a^x f \;\; and \;\; U(x) = \mathbf{U}\int_a^x f$$

Let $x \in (a,b)$. If $h>0$ and

$$\begin{aligned} m_h &= \inf\{f(t) : x \le t \le x+h\}\\ M_h &= \sup\{f(t) : x \le t \le x+h\}\end{aligned}$$

then

$$m_h \cdot h \leq \mathbf{L}\int_x^{x+h} f \leq \mathbf{U}\int_x^{x+h} f \leq M_h \cdot h$$

so

$$m_h \cdot h \leq L(x+h) - L(x) \leq U(x+h) - U(x) \leq M_h \cdot h$$

or

$$m_h \leq \frac{L(x+h) - L(x)}{h} \leq \frac{U(x+h) - U(x)}{h} \leq M_h$$

If $h < 0$ and

$$\begin{aligned} m_h &= \inf\{f(t) : x + h \leq t \leq x\} \\ M_h &= \sup\{f(t) : x + h \leq t \leq x\} \end{aligned}$$

one obtains the same inequality, precisely as in the proof of 3.8.

Since f is continuous at x, we have

$$\lim_{h\to 0} m_h = \lim_{h\to 0} M_h = f(x)$$

and this proves that

$$L'(x) = U'(x) = f(x), \forall x \in (a, b)$$

THis means that there is a number c such that

$$U(x) = L(x) + c, \forall x \in [a, b]$$

Since $U(a) = L(a) = 0$, the number c must be equal to 0, so

$$U(x) = L(x) \forall x \in [a, b]$$

In particular,

$$\mathbf{U}\int_a^b f = U(b) = L(b) = \mathbf{L}\int_a^b f$$

so f is integrable on $[a, b]$. □

3.4 Content of Sets

Using properties of sets, we can investigate the collection of Riemann integrable functions more carefully.

Definition 3.8 Given $S \subseteq I$ we define the ***characteristic function for S*** by

$$\chi_S(x) = \begin{cases} 1 & x \in S \\ 0 & x \notin S \end{cases}$$

and the ***upper content of*** S and the ***lower content of*** S by

$$\mathrm{cont}^+(S) := U_I(\chi_S) \text{ and } \mathrm{cont}^-(S) = L_I(\chi_S)$$

If $\mathrm{cont}^+(S) = \mathrm{cont}^-(S)$ we say that S has ***content***

$$m(S) = \mathrm{cont}^+(S) = \mathrm{cont}^-(S)$$

and say S is ***contented***.

We observe by Darboux's theorem $\mathrm{cont}^+(S) = U_I(\chi_S) = \lim_{\nu\to\infty} U(\chi_S, P_\nu)$ if $\mathrm{maxsize} P_\nu \to 0$. Since $\sup_J \chi_S = 1$ if $S \cap J \neq \emptyset$, $\sup_J \chi_S = 0$ if $S \cap J = \emptyset$, $\inf_J(\chi_S) = 1$ if $S \supseteq J$, and $\inf_J(\chi_S) = 0$ if $S \not\supseteq J$, we can formulate the upper and lower contents by

$$\mathrm{cont}^+(S) = \inf\left\{\sum_{k=1}^{N} \ell(J_k) : S \subseteq \bigcup_{k=1}^{N} J_k\right\}$$

and

$$\mathrm{cont}^-(S) = \sup\left\{\sum_{k=1}^{N} \ell(J_k) : S \supseteq \sqcup_{k=1}^{N} J_k\right\}$$

(Note we need disjoint sets for the lower content)

We now define the Lebesque measure, which extends the upper content to what is known as an outer measure.

Definition 3.9 We define the ***Lebesque measure*** by

$$m^*(S) = \inf\left\{\sum_{k\geq 1} \ell(J_k) : S \subseteq \bigcup_{k\geq 1} J_k\right\}$$

which is an ***outer measure***, and we can define the associated ***inner measure*** by

$$m_*(S) = \sup\left\{\sum_{k\geq 1} \ell(J_k) : S \supseteq \sqcup_{k\geq 1} J_k\right\}$$

Example 3.2 We observe that $\mathrm{cont}^+(\mathbb{Q}\cap[0,1]) = 1$ and $\mathrm{cont}^-(\mathbb{Q}\cap[0,1]) = 0$, but $m^*(\mathbb{Q}\cap[0,1]) = 0$. Indeed, as $\mathbb{Q}\cap[0,1] = \{q_1, q_2, ...\}$ is countable, if $\varepsilon > 0$ we can choose $J_k = \left[q_k - \frac{\varepsilon}{2^{k+2}}, q_k + \frac{\varepsilon}{2^{k+2}}\right] \cap [0,1]$. Then $\mathbb{Q} \cap [0,1] \subseteq \bigcup_{k\geq 1} J_k$ and

$$\sum_{k\geq 1} \ell(J_k) \leq \sum_{k\geq 1} \frac{\varepsilon}{2^{k+1}} \leq \frac{\varepsilon}{2} < \varepsilon$$

In general, if S is countable then $m^*(S) = 0$. We can now investigate a necessary and sufficient condition for Riemann integrability:

Proposition 3.3 *If $f : I = [a,b] \to \mathbb{R}$ is bounded, and S is the set of points of discontinuity of f in $[a,b]$, then $m^*(S)0$ implies $f \in \mathcal{R}(I)$, where*

$$m^*(S) = \inf \left\{ \sum_{k\geq 1} \ell(J_k) : S \subseteq \bigcup_{k\geq 1} J_k \right\}$$

Proof As f is bounded, there exists $M > 0$ such that $||f||_\infty \leq M$. Let $\varepsilon > 0$. As $m^*(S) = 0$, there exists J_k, $k \geq 1$, such that

$$S \subseteq \bigcup_{k\geq 1} J_k \text{ and } \sum_{k=1}^{\infty} \ell(J_k) < \varepsilon$$

If $x \in I \backslash S$, then there exists an open interval K_x containing x such that

$$0 \leq \sup_{K_x}(f) - \inf_{K_x}(f) < \varepsilon$$

as f is continuous at x. Then $I \subseteq \left(\bigcup_{k\geq 1} J_k\right) \cup \left(\bigcup_{x\in I\backslash S} K_x\right)$. But I is compact so there exists a finite covering

$$I \subseteq \left(\bigcup_{k=1}^{N} J_k\right) \cup \left(\bigcup_{i=1}^{M} K_i\right)$$

Let $P \in \prod(I)$ be the partition conformed by all the endpoints of intervals $J_l, 1 \leq l \leq N$, and $K_i, 1 \leq i \leq M$. Write $P = \{L_1, ..., L_\mu\}$. We have two cases:

1. $L_j \subseteq K_i$ for some i or
2. $L_j \subseteq J_l$ for some l

Let $A = \{1 \leq j \leq \mu : \exists i; L_j \subseteq K_i\}$ and $B = \{1 \leq j \leq \mu : \exists l; L_j \subseteq J_l\}$. Note $\sum_{j\in B} \ell(L_j) \leq \sum_{k=1}^{N} \ell(J_k) < \varepsilon$ and

$$\begin{aligned} 0 \leq U(f,P) - L(f,P) &= \sum_{j\in A} \left(\sup_{L_j}(f) - \inf_{L_j}(f)\right) \ell(L_j) + \sum_{j\in B} \left(\sup_{L_j}(f) - \inf_{L_j}(f)\right) \ell(L_j) \\ &< \varepsilon \sum_{j\in A} \ell(L_j) + 2M \sum_{j\in B} \ell(L_j) \\ &\leq \varepsilon \ell(I) + 2M\varepsilon \\ &= \varepsilon(\ell(I) + 2M) \end{aligned}$$

which goes to 0 as we can make ε arbitrarily small. Thus $f \in \mathcal{R}([a,b])$ as desired. □

Appendix A: Trigonometric Functions

Definition 3.10 We define the mathematical constant π as the area of the unit circle, or in this case, twice the area of a semi-circle:

$$\pi := 2 \cdot \int_{-1}^{1} \sqrt{1-x^2}dx \tag{3.4.1}$$

Definition 3.11 If $-1 \leq x \leq 1$, then the area of the sector bounded between the upper unit circle from $[x, 1]$ and the x-axis and radial arm is

$$A(x) := \frac{x\sqrt{1-x^2}}{2} + \int_{x}^{1} \sqrt{1-t^2}dt \tag{3.4.2}$$

Remark 3.2 For $-1 < x < 1$, A is differentiable at x and

$$\begin{aligned} A'(x) &= \frac{1}{2}\left[\sqrt{1-x^2} + x \cdot \frac{-2x}{2\sqrt{1-x^2}}\right] - \sqrt{1-x^2} \\ &= \frac{1}{2}\frac{1-x^2-x^2}{\sqrt{1-x^2}} - \frac{1-x^2}{\sqrt{1-x^2}} \\ &= \frac{1}{2}\frac{-1}{\sqrt{1-x^2}} \\ &= \frac{-1}{2\sqrt{1-x^2}} \end{aligned}$$

Note that on $[-1, 1]$, the function A decreases from $A(-1) = \frac{\pi}{2}$ to $A(1) = 0$.

Definition 3.12 If $0 \leq x \leq \pi$, then $\cos x$ is the unique number in $[-1, 1]$ such that

$$A(\cos x) = \frac{x}{2} \tag{3.4.3}$$

and

$$\sin x := \sqrt{1-(\cos x)^2} \tag{3.4.4}$$

Note that such a $\cos x$ exists as A is continuous on $[-1, 1]$, and $A(-1) = \frac{\pi}{2}$ while $A(1) = 0$. Hence, by **??** there exists $y \in [-1, 1]$ such that $A(y) = \frac{x}{2}$ for all $x \in [0, \pi]$.

Theorem 3.11 *If* $0 < x < \pi$, *then*

$$\begin{aligned} \cos'(x) &= -\sin x \\ \sin'(x) &= \cos x \end{aligned}$$

Proof If $B = 2A$, then the definition $A(\cos x) = x/2$ can be written

$$B(\cos x) = x$$

in other words, cos is just the inverse of B. We have already computed taht

$$A'(x) = -\frac{1}{2\sqrt{1-x^2}}$$

from which we conclude

$$B'(x) = -\frac{1}{\sqrt{1-x^2}}$$

Consequently we have that

$$\begin{aligned}\cos'(x) &= (B^{-1})'(x)\\ &= \frac{1}{B'(B^{-1}(x))}\\ &= \frac{1}{-\frac{1}{\sqrt{1-[B^{-1}(x)]^2}}}\\ &= -\sqrt{1-(\cos x)^2}\\ &= -\sin x\end{aligned}$$

Then, since $\sin x = \sqrt{1-(\cos x)^2}$ we also obtain

$$\begin{aligned}\sin'(x) &= \frac{1}{2}\cdot\frac{-2\cos x\cdot\cos'(x)}{\sqrt{1-(\cos x)^2}}\\ &= \frac{-\cos x\cdot(-\sin x)}{\sin x}\\ &= \cos x\end{aligned}$$

Definition 3.13 Now, to define sin and cos on $\mathbb{R}$, we proceed as follows

1. If $\pi \leq x \leq 2\pi$, the

$$\begin{aligned}\sin x &= -\sin(2\pi - x)\\ \cos x &= \cos(2\pi - x)\end{aligned}$$

2. If $x = 2\pi k + x'$ for some integer k and some $x' \in [0, 2\pi]$, then

$$\begin{aligned}\sin x &= \sin x'\\ \cos x &= \cos x'\end{aligned}$$

Lemma 3.2 *Suppose f has a second derivative everywhere and that*

$$\begin{aligned}f'' + f &= 0\\ f(0) &= 0\\ f'(0) &= 0\end{aligned}$$

Then $f = 0$

Proof Multiplying both sides of the first equation by f' yields

$$f'f'' + ff' = 0$$

Thus

$$[(f')^2 + f^2]' = 2(f'f'' + ff') = 0$$

so $(f')^2 + f^2$ is a constant function. From $f(0) = 0$ and $f'(0) = 0$ it follows that the constant is 0; thus

$$[f'(x)]^2 + [f(x)]^2 = 0 \forall x$$

This implies that

$$f(x) = 0 \forall x$$

Theorem 3.12 *If f has a second derivative everywhere and*

$$\begin{aligned} f'' + f &= 0 \\ f(0) &= a \\ f'(0) &= b \end{aligned}$$

then

$$f = b \cdot \sin + a \cdot \cos$$

Proof Let

$$g(x) = f(x) - b \sin x - a \cos x$$

Then

$$\begin{aligned} g'(x) &= f'(x) - b \cos x + a \sin x \\ g''(x) &= f''(x) + b \sin x + a \cos x \end{aligned}$$

Consequently,

$$\begin{aligned} g'' + g &= 0 \\ g(0) &= 0 \\ g'(0) &= 0 \end{aligned}$$

which shows by the previous lemma that

$$0 = g(x) = f(x) - b \sin x - a \cos x, \forall x$$

Theorem 3.13 *If x and y are any two numbers, then*

$$\begin{aligned} \sin(x + y) &= \sin x \cos y + \cos x \sin y \\ \cos(x + y) &= \cos x \cos y - \sin x \sin y \end{aligned}$$

Proof For any particular $y \in \mathbb{R}$, we can define a function f by

$$f(x) = \sin(x + y)$$

Then $f'(x) = \cos(x + y)$ and $f''(x) = -\sin(x + y)$. Consequently,

$$\begin{aligned} f'' + f &= 0 \\ f(0) &= \sin y \\ f'(0) &= \cos y \end{aligned}$$

It follows from the previous theorem that

$$f = (\cos y) \cdot \sin + (\sin y) \cdot \cos$$

that is

$$\sin(x + y) = \cos y \sin x + \sin y \cos x, \forall x$$

Since any number y could have been chosen to begin with, this proves the first formula for x and y.

Similarly, for any $y \in \mathbb{R}$ define $f(x) = \cos(x+y)$, so $f'(x) = -\sin(x+y)$ and $f''(x) = -\cos(x+y)$. Then $f'' + f = 0$, $f(0) = \cos y$ and $f'(0) = -\sin y$. Then we have that

$$\cos(x + y) = \cos y \cos x - \sin y \sin x$$

proving the second formula. □

Remark 3.3 Since

$$\arcsin'(x) = \frac{1}{\sqrt{1 - x^2}}, -1 < x < 1$$

it follows from 3.9 that

$$\arcsin x = \arcsin x - \arcsin 0 = \int_0^x \frac{1}{\sqrt{1 - t^2}} dt$$

Using this definition of arcsin we could define sin as $\arcsin^{-1}$, and the formula for the derivative of an inverse function would show that

$$\sin'(x) = \sqrt{1 - \sin^2 x}$$

which could be defined as $\cos x$.

Appendix B: The Logarithm and Exponential Functions

Definition 3.14 If $x > 0$, then define

$$\log x := \int_1^x \frac{1}{t} dt \tag{3.4.5}$$

Theorem 3.14 *If* $x.y > 0$*, then*

$$\log(xy) = \log x + \log y \tag{3.4.6}$$

Proof Notice first that $\log'(x) = 1/x$, by 3.8. Now, choose a number $y > 0$ and let

$$f(x) = \log(xy)$$

Then

$$f'(x) = \log'(xy) \cdot y = \frac{1}{xy} \cdot y = \frac{1}{x}$$

Thus, $f' = \log'$. This means that there is a number c such that $f(x) = \log(x) + c$ for all $x > 0$, that is,

$$\log(xy) = \log x + c, \ \forall x > 0$$

The number c can be evaluated by noting that $\log(1) = 0$, so $\log(1 \cdot y) = c$. Thus

$$\log(xy) = \log x + \log y$$

for all x. Since this is true for all $y > 0$, the theorem is proved. □

Corollary 3.1 *If* n *is a natural number and* $x > 0$*, then*

$$\log(x^n) = n \log x \tag{3.4.7}$$

Proof We proceed by induction on $n \in \mathbb{N}$. If $n = 1$ we simply have $\log(x^1) = 1 \cdot \log x$, so the base case holds. Now suppose inductively that there exists $k \geq 1$ such that if $n = k$,

$$\log(x^k) = k \log x$$

Then, observe that by the previous theorem

$$\begin{aligned} \log(x^{k+1}) &= \log(x^k x) \\ &= \log(x^k) + \log x \\ &= k \log x + \log x && \text{(by the Induction Hypothesis)} \\ &= (k+1) \log x \end{aligned}$$

as desired. Thus by mathematical induction we conclude that for all $n \geq 1$, $\log(x^n) = n \log x$. □

Corollary 3.2 *If* $x, y > 0$*, then*

$$\log\left(\frac{x}{y}\right) = \log x - \log y$$

Proof This result follows from the equation

$$\log x = \log\left(\frac{x}{y} \cdot y\right) = \log\left(\frac{x}{y}\right) + \log y$$

Definition 3.15 The ***exponential function***, exp, is defined as $\log^{-1}$.

Theorem 3.15 *For all numbers x,*

$$\exp'(x) = \exp(x)$$

Proof Observe that

$$\begin{aligned}\exp'(x) &= (\log^{-1})'(x)\\ &= \frac{1}{\log'(\log^{-1}(x))}\\ &= \frac{1}{\frac{1}{\log^{-1}(x)}}\\ &= \log^{-1}(x) = \exp(x)\end{aligned}$$

Theorem 3.16 *If x and y are any two numbers, then*

$$\exp(x+y) = \exp(x)\cdot\exp(y)$$

Proof Let $x' = \exp(x)$ and $y' = \exp(y)$, so that $x = \log x'$ and $y = \log y'$. Then

$$x + y = \log x' + \log y' = \log(x'y')$$

This means that

$$\exp(x+y) = x'y' = \exp(x)\cdot\exp(y)$$

Definition 3.16 We define

$$e := \exp(1) \tag{3.4.8}$$

and this is equivalent to the equation

$$1 = \log e = \int_1^e \frac{1}{t}dt \tag{3.4.9}$$

Then, we note that $\exp(x) = [\exp(1)]^x = e^x$ for rational x, so we define for any $x \in \mathbb{R}$,

$$e^x = \exp(x) \tag{3.4.10}$$

Definition 3.17 If $a > 0$, then, for any real number x,

$$a^x := e^{x\log a} \tag{3.4.11}$$

If $a = e$ this definition agrees with our previous one.

Theorem 3.17 *If $a > 0$, then*

$$(1)\;\; (a^b)^c = a^{bc},\;\; \forall a, b \in \mathbb{R}$$

and

$$(2)\ \ a^1 = a \textit{ and } a^{x+y} = a^x \cdot a^y,\ \forall x, y \in \mathbb{R}$$

Proof First, observe that

$$\begin{aligned}(a^b)^c &= e^{c \log a^b} \\ &= e^{c \log e^{b \log a}} \\ &= e^{cb \log a} \\ &= a^{bc}\end{aligned}$$

Next, observe that

$$a^1 = e^{1 \log a} = e^{\log a} = a$$

and

$$\begin{aligned}a^{x+y} &= e^{(x+y) \log a} \\ &= e^{x \log a + y \log a} \\ &= e^{x \log a} \cdot e^{y \log a} \\ &= a^x \cdot a^y\end{aligned}$$

Theorem 3.18 *If f is differentiable and*

$$f'(x) = f(x),\ \forall x \in \mathbb{R}$$

then there is a number c such that

$$f(x) = ce^x,\ \forall x \in \mathbb{R}$$

Proof Let $g(x) = f(x)/e^x$, which is possible as $e^x \neq 0$ for all x. Then

$$g'(x) = \frac{e^x f'(x) - f(x)e^x}{(e^x)^2} = 0$$

THerefore, there is a number c such that

$$g(x) = \frac{f(x)}{e^x} = c,\ \forall x$$

Theorem 3.19 *For any natural number n,*

$$\lim_{x\to\infty} \frac{e^x}{x^n} = \infty \tag{3.4.12}$$

Proof Step 1. We claim that $e^x > x$ for all x, and consequently $\lim\limits_{x\to\infty} e^x = \infty$.

For $x \leq 0$ this is immediate. Now, it suffices to show $x > \log x$ for all $x > 0$. If $x < 1$ this clearly holds since $\log x < 0$. If $x > 1$, then $x - 1$ is an upper sum for $f(t) = \frac{1}{t}$ on $[1, x]$, so $\log x < x - 1 < x$.

Step 2. We claim $\lim\limits_{x\to\infty} \frac{e^x}{x} = \infty$. First, note that

$$\frac{e^x}{x} = \frac{e^{x/2} \cdot e^{x/2}}{\frac{x}{2} \cdot 2} = \frac{1}{2} \left(\frac{e^{x/2}}{\frac{x}{2}} \right) \cdot e^{x/2}$$

By Step 1. the expression in parentheses is greater than 1, and $\lim\limits_{x\to\infty} e^{x/2} = \infty$; this shows that $\lim\limits_{x\to\infty} e^x/x = \infty$.

Step 3. To prove the main claim note that

$$\frac{e^x}{x^n} = \frac{(e^{x/n})^x}{\left(\frac{x}{n}\right)^n \cdot n^n} = \frac{1}{n^n} \cdot \left(\frac{e^{x/n}}{\frac{x}{n}} \right)^n$$

The expression in parentheses becomes arbitrarily large, by Step 2., so the nth power certainly becomes arbitrarily large. $\square$

Part II

Higher Dimensional Analysis

Chapter 4

Metric Spaces

4.1 Euclidean Spaces

We begin our study with one of the most well studied metric spaces, Euclidean n-space.

Definition 4.1 ***Euclidean*** n-space is defined as the product

$$\mathbb{R}^n := \mathbb{R} \times \cdots \times \mathbb{R} = \{(x_1, ..., x_n) : x_1, ..., x_n \in \mathbb{R}\}$$

We define $+ : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n$ component-wise, and we define a module action $\cdot : \mathbb{R} \times \mathbb{R}^n \to \mathbb{R}^n$ also component-wise, turning $\mathbb{R}^n$ into an $\mathbb{R}$-linear space.

Euclidean space is a special type of vector space which satisfies certain additional structures:

Definition 4.2 A ***real-inner product*** on a real vector space V is a map $\langle , \rangle : V \times V \to \mathbb{R}$ such that

1. $\langle v, v \rangle \geq 0$ for all $v \in V$, and $\langle v, v \rangle = 0$ if and only if $v = 0_V$
2. $\langle v, w \rangle = \langle w, v \rangle$ for all $v, w \in V$
3. $\langle av + bu, w \rangle = a\langle v, w \rangle + b\langle u, w \rangle$ for all $v, u, w \in V$ and all $a, b \in \mathbb{R}$.

Definition 4.3 A ***norm*** on a real vector space V is a map $|| \cdot || : V \to \mathbb{R}$ such that

1. $||v|| \geq 0$ for all $v \in V$, and $||v|| = 0$ if and only if $v = 0_V$
2. $||av|| = |a| \cdot ||v||$ for all $v \in V$ and all $a \in \mathbb{R}$
3. $||v + w|| \leq ||v|| + ||w||$ for all $v, w \in V$ (triangle inequality)

Then $(V, || \cdot ||)$ is called a ***normed linear space***.

Definition 4.4 A ***metric*** on a set X is a map $d : X \times X \to \mathbb{R}$ such that

1. $d(x, y) \geq 0$ for all $x, y \in X$, and $d(x, y) = 0$ if and only if $x = y$
2. $d(x, y) = d(y, x)$ for all $x, y \in X$
3. $d(x, z) \leq d(x, y) + d(y, z)$ for all $x, y, z \in X$ (triangle inequality)

In this case we call (X, d) a ***metric space***.

Proposition 4.1 *If* $\langle , \rangle : V \times V \to \mathbb{R}$ *is an inner product,* $|| \cdot || : V \to \mathbb{R}$ *defined by* $|| \cdot || = \sqrt{\langle \cdot, \cdot \rangle}$ *is a norm, and* $d : V \times V \to \mathbb{R}$ *defined by* $d(a, b) = ||a - b||$ *is a metric.*

We now define the inner product on Euclidean n-space:

Definition 4.5 The Euclidean inner product on $\mathbb{R}^n$ is given by

$$\langle x, y \rangle = x \cdot y = \sum_{i=1}^{n} x_i y_i$$

for all $x = (x_1, ..., x_n), y = (y_1, ..., y_n) \in \mathbb{R}^n$.

Proposition 4.2 (Cauchy-Schwarz Inequality) *For any inner product* $\langle , \rangle : V \times V \to \mathbb{R}$

$$|\langle v, w \rangle| \leq |\langle v, v \rangle||\langle w, w \rangle|$$

Proof Let $t > 0$, and consider $0 \leq \langle tx - t^{-1}y, tx - t^{-1}y \rangle$, so $0 \leq t^2||x||^2 - 2\langle x, y \rangle + t^{-2}||y||^2$. Without loss of generality suppose $x, y \neq 0$. Then let $t^2 = \frac{||y||}{||x||}$, so we have $2\langle x, y \rangle \leq 2||y|| \cdot ||x||$, so we have our desired result. Replacing x with $-x$ we obtain $-\langle x, y \rangle \leq ||y|| \cdot ||x||$, so $|\langle x, y \rangle| \leq ||x|| \cdot ||y||$. □

The triangle inequality follows from this Cauchy-Schwarz inequality. Thus, $\mathbb{R}^n$ is a metric space, and in particular it is a normed linear space.

4.1.1 Sequences and Convergence

We can now use the metric on $\mathbb{R}^n$ to define notions of convergence for sequences.

Definition 4.6 If $(\mathbf{x}_j)_{j=1}^{\infty} \subset \mathbb{R}^n$, then $\mathbf{x}_j \to \mathbf{x} \in \mathbb{R}^n$ if and only if $||\mathbf{x}_j - \mathbf{x}||$ converges to 0 in $\mathbb{R}$. Further, we say such a sequence is ***Cauchy*** if and only if for all $\varepsilon > 0$, there exists $N \in \mathbb{N}$ such that for $j, k \geq N$, $||\mathbf{x}_j - \mathbf{x}_k|| < \varepsilon$.

We note that we can consider $\mathbb{C} \cong \mathbb{R}^2$ as normed linear spaces over $\mathbb{R}$.

Proposition 4.3 $\mathbb{R}^n$ *is a complete metric space.*

Proof By completeness of $\mathbb{R}$, and the fact that $(x_j)_{j=1}^\infty = (\langle x_{j,1}, ..., x_{j,n}\rangle)_{j=1}^\infty$ is Cauchy if and only if the $(x_{j,i})_{j=1}^\infty$ are Cauchy for all i. □

4.1.2 Topological Properties of Euclidean Space

First we define the notions of open and closed sets in $\mathbb{R}^n$:

Definition 4.7 We say that $S \subseteq \mathbb{R}^n$ is ***closed*** if and only if for all sequences $(x_j)_{j=1}^\infty \subseteq S$ such that x_j converges to $x \in \mathbb{R}^n$, $x \in S$.

Definition 4.8 We say that a set $U \subseteq \mathbb{R}^n$ is ***open*** if and only if $\mathbb{R}^n \backslash U = U^C$ is closed. This holds if and only if for all $x \in U$, there exists $\varepsilon > 0$ such that $B_\varepsilon(x) \subseteq U$, where

$$B_\varepsilon(x) := \{y \in \mathbb{R}^n : d(x, y) < \varepsilon\}$$

is the ε-ball centered at $x \in \mathbb{R}^n$.

Definition 4.9 A set $K \subseteq \mathbb{R}^n$ is ***sequentially compact*** if and only if for all $(p_j)_{j=1}^\infty \subseteq K$, there exists a subsequence $(p_{j_k})_{k=1}^\infty \subseteq K$ and a $p \in K$ such that p_{j_k} converges to p.

That is every sequence in a sequentially compact set has a convergent subsequence.

Definition 4.10 A set $S \subseteq T \subseteq \mathbb{R}^n$ is said to be ***dense*** in T if $\overline{S} \supseteq T$, where $\overline{S}$ is the ***closure*** of S:

$$\overline{S} := \bigcap \{C \subseteq \mathbb{R}^n : S \subseteq C \text{ and } C \text{ is closed}\}$$

Note that the closure of a set S is the smallest closed set containing S.

Definition 4.11 A topological space (X, τ) is said to be ***separable*** if it has a ***countable dense subset***.

We note that $\mathbb{R}^n$ is separable with countable dense subset $\mathbb{Q}^n$. Indeed, for all $U \subseteq \mathbb{R}^n$, U is open if and only if for all $p \in U$, there exists $q \in \mathbb{Q}^n$ and $r \in \mathbb{Q}^+$ such that $p \in B_r(q) \subseteq U$. Hence,

$$\mathcal{B} := \{B_r(q) : r \in \mathbb{Q}^+, q \in \mathbb{Q}^n\}$$

is a countable base for the topology on $\mathbb{R}^n$, so $\mathbb{R}^n$ is a ***second countable space***. We have the following useful results about compactness in Euclidean spaces:

Corollary 4.1 *If* $K \subseteq \mathbb{R}^n$ *is sequentially compact then it is topologically compact.*

To prove this corollary we first prove some intermediate results.

Proposition 4.4 *If $K \subseteq \mathbb{R}^n$ is sequentially compact, and $X_1 \supseteq X_2 \supseteq X_3 \supseteq \ldots$ is a chain of nonempty closed subsets of K, then $\bigcap_{j\geq 1} X_j \neq \emptyset$.*

Proof Let $x_j \in X_j \subseteq K$ for each j. Then $(x_j)_{j=1}^\infty \subseteq K$, so as K is compact we have a subsequence $(x_{j_k})_{k=1}^\infty \subseteq K$ such that $x_{j_k} \to x \in K$. But, for each $m \in \mathbb{N}$, $\{x_{j_k} : k \geq m\} \subseteq X_m$, so as X_m is closed it follows that $x \in X_m$. Thus, $x \in \bigcap_{j\geq 1} X_j$, so $\bigcap_{j\geq 1} X_j \neq \emptyset$. □

Proposition 4.5 *If $K \subseteq \mathbb{R}^n$ is compact and $U_1 \subseteq U_2 \subseteq \ldots$ is a chain of open sets which cover K, $K \subseteq \bigcup_{j\geq 1} U_j$, then there exists $M \in \mathbb{N}$ such that $K \subseteq U_M$.*

Proof Consider $X_m = K \backslash U_m$, which gives a chain of open sets. We have that $\bigcap_{j\geq 1} X_j = \emptyset$, so by the contrapositive of the previous proposition there exists $M \in \mathbb{N}$ such that $X_M = \emptyset$. Then $K \subseteq U_M$, as desired. □

Now we proceed to the proof of the corollary:

Proof Let $\{U_\alpha\}_{\alpha\in J}$ be an open cover of K, which is sequentially compact in $\mathbb{R}^n$. For each α we have $U_\alpha = \bigcup_{j\geq 1} B_{r_{\alpha_j}}(q_{\alpha_j})$ is the union of a countable number of open sets in our countable base. Then, $K \subseteq \bigcup_{\alpha\in J} \bigcup_{j\geq 1} B_{r_{\alpha_j}}(q_{\alpha_j})$, which is a union over a countable collection of open sets since $\mathcal{B}$ is countable. We claim any countable cover has a finite subcover for K. Indeed, consider $U_m = \bigcup_{j=1}^m B_j$ for a countable cover $\bigcup_{j\geq 1} B_j$. Then by the previous proposition there exists $M \in \mathbb{N}$ such that $K \subseteq \bigcup_{j=1}^M B_j$. Thus, there exist $\alpha_1, \ldots, \alpha_N \in J$, and M_j such that $K \subseteq \bigcup_{j=1}^N \bigcup_{k=1}^{M_j} B_{r_{\alpha_{j,k}}}(q_{\alpha_{j,k}}) \subseteq \bigcup_{j=1}^N U_{\alpha_j}$, so K it topologically compact. □

4.2 Metric Space Properties

Definition 4.12 (Metric Space) A set X together with a function $d : X \times X \to [0, \infty)$ is called a ***metric space*** if the function d, called a ***metric*** satisfies the following properties for all $x, y, z \in X$:

(1)$d(x, y) = 0$ if and only if $x = y$ (Positive definiteness)

(2)$d(x, y) = d(y, x)$ (Symmetry)

(3)$d(x, y) \leq d(x, z) + d(z, y)$ (Triangle inequality)

Example 4.1 $\mathbb{R}$ and $\mathbb{C}$ with the usual modulus are metric spaces. In this case $d(x, y) = |x - y|$. Additionally, as seen in the last section $\mathbb{R}^n$ and $\mathbb{C}^n$ are metric spaces, with the Euclidean metric

$$d(\mathbf{x}, \mathbf{y}) = \sqrt{\sum_{j=1}^n |x_i - y_i|^2}$$

where the triangle inequality follows by the Cauchy-Schwartz inequality.

Proposition 4.6 *Let V be a real inner product space. Then for all $u, v \in V$,*

$$|\langle u, v\rangle|^2 \leq \langle u, u\rangle \cdot \langle v, v\rangle$$

Proof Let $t \in \mathbb{R}\backslash\{0\}$, and consider

$$0 \leq \langle tx - t^{-1}y, tx - t^{-1}y\rangle = t^2|x|^2 - 2\langle x, y\rangle + t^{-2}|y|^2$$

Now, if x or y is the zero vector, then the result holds trivially. Thus, suppose $x, y \neq 0$, and let $t^2 = \frac{|y|}{|x|}$. It follows that

$$2\langle x, y\rangle \leq |y||x| + |x||y| = 2|x||y|$$

so $\langle x, y\rangle \leq |x||y|$. Exchanging x with $-x$ we obtain $-\langle x, y\rangle \leq |x||y|$, so $|\langle x, y\rangle| \leq |x||y|$. Squaring both sides we obtain the desired inequality. □

Example 4.2 (Discrete Metric) Let X be any set. Then we can define the ***discrete metric*** on X by

$$d(x, y) = \begin{cases} 1 & x \neq y \\ 0 & x = y \end{cases}$$

which gives X the discrete topology, in which $\tau_X = \mathcal{P}(X)$.

Example 4.3 (Subsets of Metric Spaces) If $Y \subseteq X$ and (X, d) is a metric space, then $(Y, d|_Y)$ is also a metric space, with topology corresponding to the subspace topology.

4.2.1 Normed Linear Spaces

Definition 4.13 (Normed Linear Space) Let V be a vector space over $\mathbb{R}$ or $\mathbb{C}$. We say that V is a ***normed linear space*** if there is a function $|| \cdot || : V \longrightarrow [0, \infty)$ called a ***norm*** satisfying the following properties for all $v, w \in V$ and constants c:

(1)$||v|| = 0$ if and only if $v = 0$ (Positive definitiness)

(2)$||cv|| = |c|||v||$, (Absolute homegeneity)

(3)$||v + w|| \leq ||v|| + ||w||$ (triangle inequality)

Proposition 4.7 *A normed linear space $(V, ||\cdot||)$ is also a metric space with metric $d(v, w) = ||v-w||$.*

Proof First, observe that $d(v, v) = ||v-v|| = ||0|| = 0$, and if $d(v, w) = ||v-w|| = 0$ then $v-w = 0_V$, so $v = w$. Next,

$$d(v, w) = ||v - w|| = || - (w - v)|| = | - 1|||w - v|| = d(w, v)$$

so we have symmetry. Finally, for all $u, v, w \in V$,

$$d(v,w) = ||v-w|| = ||v-u+u-w|| \leq ||v-u|| + ||u-w|| = d(v,u) + d(u,w)$$

so the triangle inequality is satisfied. □

This leads to the question:

? Question

Does every metric on a vector space arise from a norm?

The answer is negative. Let V be a vector space and assume $\dim V \geq 1$ (so it isn't the zero vector space). Equip V with the discrete metric. For the sake of contradiction assume there is a norm $||\cdot||$ on V with $d(v,w) = ||v-w||$. If $v \neq 0$, then

$$1 = d(v,0) = ||v-0|| = ||v||$$

But, using absolute homogeneity,

$$1 = d(2v,0) = ||2v|| = 2||v||$$

which is a contradiction.

Example 4.4 We define the Euclidean norm on $\mathbb{R}^n$ and $\mathbb{C}^n$ by

$$||\mathbf{x}||_2 := \sqrt{\sum_{j=1}^{n} |x_j|^2}$$

This norm induces the Euclidean metric.

Example 4.5 (Spaces of Continuous Functions) Let $S \subseteq \mathbb{R}$. Define the following:

(i) Continuous functions on S:

$$C(S) = \{f : S \to \mathbb{R} | f \text{ is continuous}\}$$

This is an $\mathbb{R}$-vector space since $f+g$ and cf are continuous on S whenever f, g are and $c \in \mathbb{R}$

(ii)Continuous bounded functions on S:

$$C_b(S) = \{f : S \to \mathbb{R} | f \text{ is continuous and } ||f||_S < \infty\}$$

This is a subspace of $C(S)$ and is also a normed linear space with norm $||f||_S = \sup_{x\in S} |f(x)|$. Indeed, $||f||_S = 0$ if and only if $|f(x)| = 0$ for all $x \in S$, if and only if $f \equiv 0$. Further,

$|cf(x)| = |c||f(x)|$, and taking the supremum of both sides $||cf||_S = |c|||f||S$. Finally, if $f, g \in C_b(S)$, then

$$|f(x) + g(x)| \leq |f(x)| + |g(x)| \leq ||f||_S + ||g||_S$$

for all $x \in S$, so $||f + g||_S \leq ||f||_S + ||g||_S$.

(iii)Continuous functions vanishing at ∞:

$$C_0(S) = \{f : S \to \mathbb{R} | f \text{ is continuous and } \forall \varepsilon > 0, \exists K \subseteq S \text{ compact; } \sup_{x \in S \setminus K} |f(x)| < \varepsilon\}$$

That is to say f is “small" outside of a compact set. $C_0(S)$ is a subspace of $C_b(S)$ with norm $|| \cdot ||_\infty$. Note that

$$C_0(\mathbb{R}) = \{f : \mathbb{R} \to \mathbb{R} | \lim_{x \to \infty} f(x) = 0 = \lim_{x \to -\infty} f(x)\}$$

(iv)Continuous functions with compact suppose on S:

$$C_c(S) = \{f : S \to \mathbb{R} | \exists K \subseteq S \text{ with } \overline{K} \text{ compact and } f(x) = 0 \forall x \in S \setminus K\}$$

K is called the ***support*** of f. $C_c(S)$ is a subspace of $C_0(S)$.

Remark 4.1 We note that

(i) If S is compact, $C(S) = C_b(S) = C_0(S) = C_c(S)$

(ii)All of these may be defined for $\mathbb{C}$ valued functions as well

(iii)If $S = [a, b]$, then there are many familiar norms on $C([a, b])$, one such being

$$||f||_p := \left[\int_a^b |f(x)|^p dx\right]^{1/p}$$

for any $p \geq 1$.

Proposition 4.8 *Suppose $n \in \mathbb{N}$. The quantity*

$$||x||_\infty = \max\{|x_k| : 1 \leq k \leq n\}$$

defines a norm called the **∞-norm** *on $\mathbb{R}^n$ and $\mathbb{C}^n$.*

Proof First, observe that $||x||_\infty = 0$ if $x = \mathbf{0}$. Further, if $||x||_\infty = 0$, then $0 \leq |x_k| \leq 0$ for all $1 \leq k \leq n$, so $x_k = 0$ for all k and hence $x = (0, 0, ..., 0)$. Then, if $c \in \mathbb{C}$, it follows that

$$\begin{aligned} ||cx||_\infty &= \max\{|cx_k| : 1 \leq k \leq n\} \\ &= \max\{|c||x_k| : 1 \leq k \leq n\} \\ &= |c| \max\{|x_k| : 1 \leq k \leq n\} = |c|||x||_\infty \end{aligned}$$

Finally, if $x, y \in \mathbb{C}^n$, then

$$|x_k + y_k| \leq |x_k| + |y_k| \leq \max\{|x_k| : 1 \leq k \leq n\} + \max\{|x_k| : 1 \leq k \leq n\} = ||x||_\infty + ||y||_\infty$$

for all k, since $|\cdot|$ is a norm on $\mathbb{C}$, so

$$||x+y||_\infty = \max\{|x_k + y_k| : 1 \le k \le n\} \le ||x||_\infty + ||y||_\infty$$

Proposition 4.9 *Suppose $n \in \mathbb{N}$. The quantity*

$$||x_1|| = \sum_{k=1}^{n} |x_k|$$

defines a norm called the **1-norm** *on $\mathbb{R}^n$ and $\mathbb{C}^n$.*

Proof First, $||x||_1 = 0$ if and only if $\sum_{k=1}^n |x_k| = 0$ if and only if $|x_k| = 0$ for all k if and only if $x_k = 0$ for all k, if and only if $x = \mathbf{0}$. Next,

$$||cx||_1 = \sum_{k=1}^{n} |cx_k| = |c| \sum_{k=1}^{n} |x_k| = |c|||x||_1$$

and

$$||x+y||_1 = \sum_{k=1}^{n} |x_k + y_k| \le \sum_{k=1}^{n} |x_k| + |y_k| = ||x||_1 + ||y||_1$$

Definition 4.14 (ℓ_n^p norm) Suppose $p \ge 1$ and $n \in \mathbb{N}$. For any vector $x = (x_1, ..., x_n) \in \mathbb{R}^n$ or $\mathbb{C}^n$, the quantity

$$||x||_p = \left(\sum_{k=1}^{n} |x_k|^p\right)^{1/p}$$

is called the ***p*-norm**. When $p = \infty$ we use

$$||x||_\infty = \max\{|x_k| : 1 \le k \le n\}$$

We now aim to verify this is a norm for the case of p not equal to 1 or ∞.

Lemma 4.1 (Young's Inequality) *Suppose $p > 1$ and $q > 1$ satisfy $\frac{1}{p} + \frac{1}{q} = 1$. If $a > 0$ and $b > 0$ then*

$$ab \le \frac{a^p}{p} + \frac{b^q}{q}$$

Equality holds if and only if $a^p = b^q$.

Proof If $p = q = 2$, then $(a-b)^2 \ge 0$ so $a^2 - 2ab + b^2 \ge 0$, so $\frac{a^2+b^2}{2} \ge ab$. Further, note $\frac{p-1}{p} = 1 - \frac{1}{p} = \frac{1}{q}$, and so $q = \frac{p}{p-1}$ for $p \ne 1$.

Now, define $y = x^{p-1}$ for $x > 0$. Then y is invertible as a function of x and $x = y^{q-1}$ since $\frac{1}{p-1} = \frac{q}{p} = q - 1$. Then

$$\int_0^a x^{p-1} dx + \int_0^b y^{q-1} dy = \frac{x^p}{p}\bigg|_{x=0}^{a} + \frac{y^q}{q}\bigg|_{y=0}^{b} = \frac{a^p}{p} + \frac{b^q}{q}$$

The area of the box corresponding to sidelengths a and b is $ab \leq \frac{a^p}{p} + \frac{b^q}{q}$, and equality holds if and only if (a, b) is on the graph of $y = x^{p-1}$, or in other words $b = a^{p-1}$, which happens if and only if $b^q = a^{(p-1)q} = a^p$. □

Lemma 4.2 (Hölder's Inequality) *Suppose* $p > 1$ *and* $q > 1$ *satisfy* $\frac{1}{p} + \frac{1}{q} = 1$. *For any vectors* $x = (x_1, ..., x_n)$ *and* $y = (y_1, ..., y_n)$ *in* $\mathbb{R}^n$ *or* $\mathbb{C}^n$ *we have*

$$\sum_{k=1}^{n} |x_k y_k| \leq ||x||_p ||y||_q$$

Note taht the case of $p = q = 2$ is the Cauchy-Schwartz Inequality.

Proof If $\mathbf{x} = 0$ or $\mathbf{y} = 0$, we get 0 on both sides so we may assume $||\mathbf{x}||_p > 0$ and $||\mathbf{y}||_q > 0$. Define $a_j = \frac{|x_j|}{||\mathbf{x}||_p}$ and $b_j = \frac{|y_j|}{||\mathbf{y}||_q}$. Then

$$\begin{aligned}
\sum_{j=1}^{n} \frac{|x_j y_j|}{||\mathbf{x}||_p ||\mathbf{y}||_q} &= \sum_{j=1}^{n} a_j b_j \\
&\leq \sum_{j=1}^{n} \frac{a_j^p}{p} + \frac{b_j^q}{q} && \text{(by Young's Inequality)} \\
&= \sum_{j=1}^{n} \frac{|x_j|^p}{p||\mathbf{x}||_p^p} + \sum_{j=1}^{n} \frac{|y_j|^q}{q||\mathbf{y}||_q^q} \\
&= \frac{1}{p||\mathbf{x}||_p^p} ||\mathbf{x}||_p^p + \frac{1}{q||\mathbf{y}||_q^q} ||\mathbf{y}||_q^q \\
&= \frac{1}{p} + \frac{1}{q} = 1
\end{aligned}$$

Multiplying both sides by $||\mathbf{x}||_p ||\mathbf{y}||_q$ we obtain the desired inequality

$$\sum_{j=1}^{n} |x_j y_j| \leq ||\mathbf{x}||_p ||\mathbf{y}||_q$$

Theorem 4.1 *For any* $p \geq 1$, *including* $p = \infty$, *and* $n \in \mathbb{N}$, *the vector space* $\mathbb{R}^n$ *or* $\mathbb{C}^n$ *together with the norm* $|| \cdot ||_p$ *is a normed linear space.*

Proof Without loss of generality we may assume $1 < p < \infty$. Then $||\mathbf{x}||_p = \left(\sum_{j=1}^{n} |x_j|^p\right)^{1/p} = 0$ if and only if $|x_j|^p = 0$ for all j, which occurs if and only if $x_j = 0$ for all j, or $\mathbf{x} = \mathbf{0}$. Secondly, for all $c \in \mathbb{C}$,

$$||c\mathbf{x}||_p = \left(\sum_{j=1}^{n} |cx_j|^p\right)^{1/p} = \left(|c|^p \sum_{j=1}^{n} |x_j|^p\right)^{1/p} = |c| \left(\sum_{j=1}^{n} |x_j|^p\right)^{1/p} = |c|||\mathbf{x}||_p$$

Finally, write $\mathbf{x} = (x_1, ..., x_n)^T$ and $\mathbf{y} = (y_1, ..., y_n)^T$. Then

$$\begin{aligned}
||\mathbf{x}+\mathbf{y}||_p^p = \sum_{j=1}^{n} |x_j + y_j|^p &= \sum_{j=1}^{n} |x_j + y_j||x_j + y_j|^{p-1} && (p-1>0)\\
&\le \sum_{j=1}^{n} |x_j||x_j + y_j|^{p-1} + \sum_{j=1}^{n} |y_j||x_j + y_j|^{p-1}\\
&\le \left(\sum_{j=1}^{n} |x_j|^p\right)^{1/p}\left(\sum_{j=1}^{n} |x_j + y_j|^{(p-1)q}\right)^{1/q} + \left(\sum_{j=1}^{n} |y_j|^p\right)^{1/p}\left(\sum_{j=1}^{n} |x_j + y_j|^{(p-1)q}\right)^{1/q}\\
& && \text{(by Hölder's inequality)}\\
&= (||\mathbf{x}||_p + ||\mathbf{y}||_p||)\left(\sum_{j=1}^{n} |x_j + y_j|^p\right)^{(p-1)/p}\\
& && \left(\text{as } \tfrac{1}{p} + \tfrac{1}{q} = 1, \text{ so } p = q(p-1)\right)\\
&= (||\mathbf{x}||_p + ||\mathbf{y}||_p)||\mathbf{x}+\mathbf{y}||_p^{p-1}
\end{aligned}$$

Note that if $\mathbf{x}+\mathbf{y} = 0$, the triangle inequality follows by definition of the p-norm. Assuming $||\mathbf{x}+\mathbf{y}||_p \neq 0$ and dividing the above inequality by $||\mathbf{x}+\mathbf{y}||_p^{p-1} > 0$ we obtain

$$||\mathbf{x}+\mathbf{y}||_p \le ||\mathbf{x}||_p + ||\mathbf{y}||_p$$

known as ***Minkowski's Inequality*** □

Definition 4.15 We denote $\mathbb{R}^n$ or $\mathbb{C}^n$ together with the p-norm as ℓ_n^p.

Example 4.6 Let (X, d) be a metric space. Then X is also a metric space with the metric $\tilde{d} = \frac{d}{1+d}$. The first two axioms are immediate from the fact that d is a metric. Now, if $0 \le a \le b$, then $\frac{a}{1+a} \le \frac{b}{1+b}$ since this is true if and only if $a + ab \le b + ab$, which holds if and only if $a \le b$. Then

$$\begin{aligned}
\tilde{d}(x,y) = \frac{d(x,y)}{1+d(x,y)} &\le \frac{d(x,z)+d(z,y)}{1+d(x,z)+d(z,y)}\\
&= \frac{d(x,z)}{1+d(x,z)+d(z,y)} + \frac{d(z,y)}{1+d(x,z)+d(z,y)}\\
&\le \frac{d(x,z)}{1+d(x,z)} + \frac{d(z,y)}{1+d(z,y)}\\
&= \tilde{d}(x,z) + \tilde{d}(z,y)
\end{aligned}$$

Even if V is a normed linear space, and d is induced by a norm, the metric $\tilde{d}$ is never induced by a norm, unless V is zero dimensional. Indeed, suppose $d(v,w) = ||v-w||$ and $\dim V > 0$, then $\tilde{d}(v,w) = \frac{||v-w||}{1+||v-w||}$. Assume for the sake of contradiction that $\tilde{d}(v,w) = |||v-w|||$ for some norm $||| \cdot |||$. If $v \neq 0$, $\tilde{d}(v,0) = |||v||| = \frac{||v||}{1+||v||}$. But

$$|||2v||| = \frac{||2v||}{1+||2v||} = \frac{2||v||}{1+2||v||} \neq \frac{2||v||}{1+||v||} = 2|||v|||$$

4.2.2 Sequences and Limits

Throughout this section let (X, d) be a metric space. Then we can define convergence of sequences as follows:

Definition 4.16 A sequence, that is a denumerable list of elements, (x_n) in a metric space (X, d) is said to converge if there is an $x \in X$ such that for all $\varepsilon > 0$ there is an $N \in \mathbb{N}$ such that $n \geq N$ implies $d(x_n, x) < \varepsilon$. Equivalently, if $\lim_{n\to\infty} d(x_n, x) = 0$.

Example 4.7 Vectors $\mathbf{x}_n$ in ℓ_n^p converge to $\mathbf{x}$ if and only if $||\mathbf{x}_n - \mathbf{x}||_p \to 0$.

This is identical for all normed linear spaces

Example 4.8 f_n in $C_b(S)$ for $S \subseteq \mathbb{R}$ converge to $f \in C_b(S)$ if and only if $||f_n - f||_\infty \to 0$. That is, convergence in this space is uniform convergence.

Definition 4.17 A sequence $(x_j)_{j=1}^\infty$ in a metric space X is said to be ***Cauchy*** (with respect to d) if and only if $d(x_j, x_k)$ converges to 0 in $\mathbb{R}$ as k and j go to infinity. Equivalently, if and only if for all $\varepsilon > 0$ there is an $N \in \mathbb{N}$ such that $m, n \geq N$ implies

$$d(x_n, x_m) < \varepsilon$$

Proposition 4.10 *In any metric space (X, d), convergent sequences are Cauchy.*

Proof Suppose $x_n \to x$ in X. Then for $\varepsilon > 0$ there exists $N \in \mathbb{N}$ such that for $n \geq N$, $d(x_n, x) < \varepsilon/2$. It follows that for $m, n \geq N$,

$$d(x_n, x_m) \leq d(x_n, x) + d(x, x_m) < \varepsilon/2 + \varepsilon/2 = \varepsilon$$

so the sequence is Cauchy. □

Definition 4.18 X is said to be a ***complete metric space***, or a ***Fréchet space***, if and only if for all Cauchy sequences $(x_j)_{j=1}^\infty \subseteq X$, there exists $x \in X$ such that x_j converges to x.

Example 4.9 With respect to the usual absolute value metric, $\mathbb{R}$ is complete. With this same metric, $\mathbb{Q}$ is not complete. For $\mathbb{Q}$, let x_n = the nth decimal truncation of $\sqrt{2}$. In $\mathbb{R}$ $x_n \to \sqrt{2}$, and limits are unique. Since $\mathbb{Q}$ is a metric subspace of $\mathbb{R}$ under the same metric, it must have the same limit. But $\sqrt{2} \notin \mathbb{Q}$.

Example 4.10 Every discrete space is complete. Indeed, if (X, d) is a discrete space, then x_n converges to $x \in X$ if and only if there exists $N \in \mathbb{N}$ such that $x_n = x$ for all $n \geq N$, and equivalently, x_n is Cauchy if and only if there exists $N \in \mathbb{N}$ such that $x_n = x_m$ for all $n, m \geq N$.

Example 4.11 The space $C_b(S)$ for $S \subseteq \mathbb{R}$ is complete. The norm here is the supremum norm, $||\cdot||_\infty$. Recall the Cauchy criterion for uniform convergence: $f_n \to_u f$ if and only if f_n is Cauchy with respect to $||\cdot||_\infty$. Additionally, as the uniform limit of a sequence of bounded functions is bounded, the limit function will also be in $C_b(S)$. But, $C_c(S)$ is not generally complete. Consider $f_n \in C_c(\mathbb{R})$, with

$$f_n(x) = \begin{cases} \frac{1}{1+x^2} & x \in [-n, n] \\ \text{“continuous linear piecewise"} & x \in [-n-1, -n] \cup [n, n+1] \\ 0 & x < -(n+1) \text{ or } x > n+1 \end{cases}$$

$f_n \in C_c(\mathbb{R})$ since $(-n-1, n+1) = \{x | f_n(x) \neq 0\}$ has compact closure $[-n-1, n+1]$. First, for $\varepsilon > 0$, there exists $N \in \mathbb{N}$ such that $\frac{1}{1+N^2} < \varepsilon$. Then for $n \geq N$ we have that $||f_n - f||_\infty \leq \frac{1}{1+N^2} < \varepsilon$, so $f_n \to_u f$. But, $f(x) = \frac{1}{1+x^2} \notin C_c(\mathbb{R})$. So $C_c(\mathbb{R})$ is not complete.

Example 4.12 The continuous function on $[0, 1]$ with norm $||f||_1 = \int_0^1 |f(x)|dx$ is not complete. Let f_n be the piecewise continuous function

$$f_n(x) = \begin{cases} 1 & 0 \leq x \leq 1/2 \\ -\frac{2}{n}\left(x - \frac{n+2}{2n}\right) & \frac{1}{2} \leq x \leq \frac{1}{2} + \frac{1}{n} \\ 0 & 1 \geq x \geq \frac{1}{2} + \frac{1}{n} \end{cases}$$

Let $n \geq m \in \mathbb{N}$. Then for $0 \leq x \leq 1/2$, $|f_n(x) - f_m(x)| = 0 \leq \frac{1}{n} + \frac{1}{m}$. If $1/2 \leq x \leq 1/2 + 1/n$, then

$$\begin{aligned} ||f_n - f_m||_1 &= \int_0^1 |f_n(x) - f_m(x)|dx \\ &= \int_{1/2}^{1/2+1/n} |f_n(x) - f_m(x)|dx + \int_{1/2+1/n}^{1/2+1/m} |f_n(x) - f_m(x)|dx \\ &\leq \frac{1}{n} + \frac{1}{m} - \frac{1}{n} \\ &\leq \frac{1}{n} + \frac{1}{m} \end{aligned}$$

Now, fix $\varepsilon > 0$. Find $N \in \mathbb{N}$ such that $\frac{2}{N} < \varepsilon$. Then if $m, n \geq N$, it follows that

$$||f_n - f_m||_1 \leq \frac{1}{n} + \frac{1}{m} \leq \frac{1}{N} + \frac{1}{N} < \varepsilon$$

So (f_n) is uniformly Cauchy. Then under the $||\cdot||_1$ norm, the limit is $f(x) = 1$, for $0 \leq x \leq 1/2$, and 0 for $1/2 < x \leq 1$. Note that f is integrable and

$$\begin{aligned} ||f - f_n||_1 &= \int_0^1 |f_n - f|dx \\ &= \int_{1/2}^{1/2+1/n} |-n(x - (n+2)/2n)|dx \\ &= \frac{n+2}{2}x - \frac{n}{2}x^2|_{1/2}^{1/2+1/n} \\ &= \frac{n+2}{2n} - \frac{n}{2}\left(\frac{1}{n^2} + \frac{1}{n}\right) \end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{2} + \frac{1}{n} - \frac{1}{2n} - \frac{1}{2} \\
&= \frac{1}{2n} \to 0
\end{aligned}
$$

But f is not continuous, and so not in $C([0,1])$. The limit f does however reside in the larger space $L^1([0,1])$, where it is the limit of f_n. In the subspace $C([0,1])$, there is no limit.

Proposition 4.11 *Limits of convergent sequences are unique in any metric space.*

Proof Suppose $x_n \to x$ and $x_n \to y$. Then $d(x,y) \leq d(x,x_n) + d(x_n,y) \to 0+0=0$, so $d(x,y)=0$ which holds if and only if $x=y$ by properties of metrics. □

Theorem 4.2 *Let $\{\mathbf{x}_m\}$ be a sequence of vectors in $\mathbb{R}^n$. Write $\mathbf{x}_m = (x_{m,1}, ..., x_{m,n})$ so that each $\{x_{m,i}\}$ is a sequence of real numbers. Then $\mathbf{x}_m$ converges in any p-norm if and only if each $x_{m,i}$ converges.*

Proof First, if $p=\infty$, then $||\mathbf{x}_k - \mathbf{x}||_\infty \to 0$ if and only if $\max_{1\leq k\leq n} |x_{k,i} - x_i| \to 0$, which occurs if and only if $|x_{k,i} - x_i| \to 0$ for all $1 \leq i \leq n$, as desired.

Next, suppose $1 \leq p < \infty$. Then

$$
\begin{aligned}
|x_{k,i} - x_i| &= (|x_{k,i} - x_i|^p)^{1/p} \\
&\leq \left(\sum_{i=1}^n |x_{k,i} - x_i|^p \right)^{1/p} \\
&= ||\mathbf{x}_k - \mathbf{x}||_p
\end{aligned}
$$

so if $||\mathbf{x}_k - \mathbf{x}||_p \to 0$, then $|x_{k,i} - x_i| \to 0$ for all $1 \leq i \leq n$. Conversely, if each $x_{k,i} \to x_i$ for all $1 \leq i \leq n$, then

$$
||\mathbf{x}_k - \mathbf{x}||_p^p = \sum_{i=1}^n |x_{k,i} - x_i|^p \to 0 + ... + 0
$$

so$||\mathbf{x}_k - \mathbf{x}||_p \to 0$ □

Corollary 4.2 *The space ℓ_n^p is complete for any $p \geq 1$, including $p=\infty$, and any $n \geq 1$.*

Proof Using the same notation and argument in the previous proposition, we can show $(\mathbf{x}_k)$ in ℓ_n^p is Cauchy if and only if $\{x_{k,i}\}$ is Cauchy in $\mathbb{R}$ for all $1 \leq i \leq n$. But $\{x_{k,i}\}$ is Cauchy in $\mathbb{R}$ if and only if it converges since $\mathbb{R}$ is a cmoplete metric space. Thus, by the proposition $(\mathbf{x}_k)$ converges in ℓ_n^p □

Example 4.13 (Doubly recursive sequence in $\mathbb{R}^2$) In ℓ_2^2 let $\mathbf{v}_0 = [0\ 0]^T$ and $\mathbf{v}_n = [x_n\ y_n]$ where

$$
x_{n+1} = \frac{x_n + y_n + 1}{2}
$$

and

$$
y_{n+1} = \frac{x_n - y_n + 1}{2}
$$

So for instance $\mathbf{v}_1 = [1/2\ 1/2]^T, \mathbf{v}_2 = [1\ 1/2]^T, \mathbf{v}_3 = [5/4\ 3/4]^T$, and $\mathbf{v}_4 = [3/2\ 3/4]^T$. I claim that $\mathbf{v}_n \to [2\ 1]^T$. For now, assume it converges to $[x\ y]^T$. Then

$$[x\ y]^T = \lim_{n\to\infty} \frac{1}{2}[x_n + y_n + 1\ x_n - y_n + 1]^T = \frac{1}{2}[x + y + 1\ x - y + 1]^T$$

using our previous result. Then $x = y + 1$ and $3y = x + 1$, so $3y - 1 = y + 1$, and $2y = 2$. It follows that $y = 1$ and $x = 2$. So now we show $[2\ 1]$ is the limit. Compute

$$\begin{aligned}
||\mathbf{v}_{n+1} - [2\ 1]^T||_2^2 &= \left\|[x_{n+1}\ y_{n+1}]^T - [2\ 1]^T\right\|_2^2 \\
&= \left\|[x_n + y_n + 1\ x_n - y_n + 1]^T/2 - [2\ 1]^T\right\|_2^2 \\
&= \left(\frac{x_n + y_n + 1}{2} - 2\right)^2 + \left(\frac{x_n - y_n + 1}{2} - 1\right)^2 \\
&= \frac{(x_n + y_n - 3)^2}{4} + \frac{(x_n - y_n - 1)^2}{4} \\
&= \frac{x_n^2 + y_n^2 + 2x_ny_n - 6x_n - 6y_n + 9 + x_n^2 + y_n^2 - 2x_ny_n - 2x_n + 2y_n + 1}{4} \\
&= \frac{2x_n^2 - 8x_n + 2y_n^2 - 4y_n + 10}{4} \\
&= \frac{2(x_n - 2)^2 + 2(y_n - 1)^2}{4} \\
&= \frac{(x_n - 2)^2}{2} + \frac{(y_n - 1)^2}{2} \\
&= \frac{1}{2}\left\|[x_n - 2\ y_n - 1]^T\right\|_2^2 \\
&= \frac{1}{2}||\mathbf{v}_n - [2\ 1]^T||_2^2 && \text{(iterate this argument)} \\
&= \frac{1}{2^{n+1}}||\mathbf{v}_0 - [2\ 1]^T||_2^2 \\
&= \frac{5}{2^{n+1}} \to 0
\end{aligned}$$

as claimed.

4.3 Topology of Metric Spaces

We now define the natural topology on metric spaces defined by their metrics:

Definition 4.19 Let (X, d) be a metric space. For $x_0 \in X$ and $r > 0$, define the ***open ball centred at*** ***x_0 of radius r*** by

$$B_r(x_0) = \{x \in X | d(x, x_0) < r\}$$

The ***closed ball centred at x_0 of radius r*** is

$$\overline{B}_r(x_0) = \{x \in X | d(x, x_0) \leq r\}$$

If V is a normed linear space, the ***unit ball*** is the set $B_1(0)$ and the ***closed unit ball*** is the set $\overline{B}_1(0)$.

Example 4.14 We draw the unit balls for $\ell_2^1, \ell_2^2, \ell_2^4$, and ℓ_2^∞. Note for $p = 1$

$$B_1(\mathbf{0}) = \{[x\ y]^T \,||x| + |y| < 1\}$$

for $p = 2$,

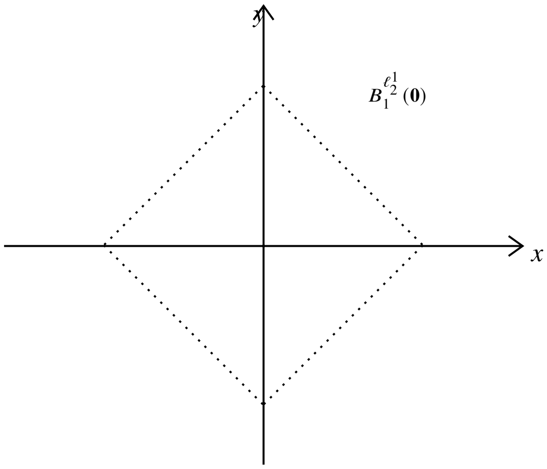
$$B_1(\mathbf{0}) = \{[x\ y]^T \,|x^2 + y^2 < 1\}$$

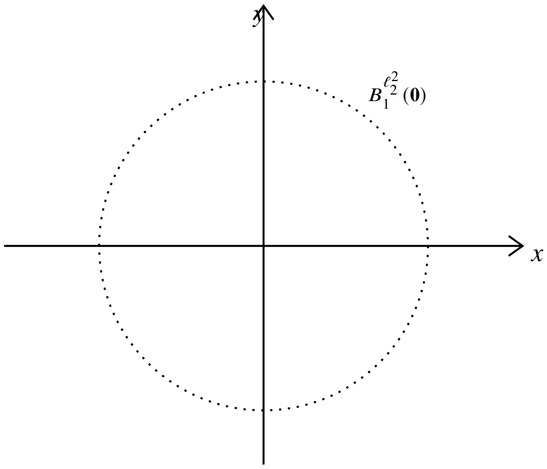
for $p = 4$

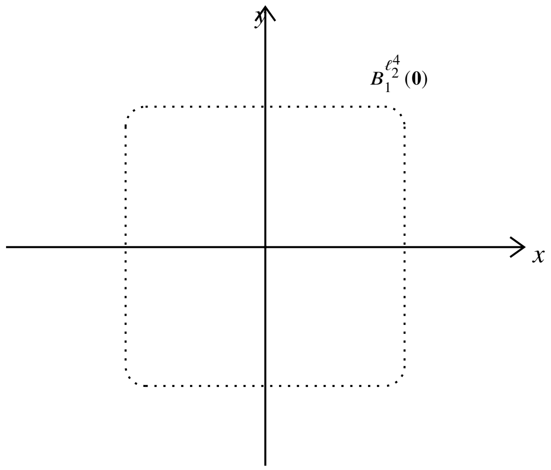
$$B_1(\mathbf{0}) = \{[x\ y]^T \,|x^4 + y^4 < 1\}$$

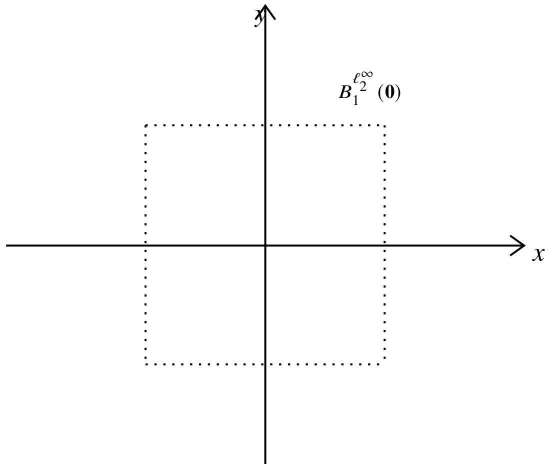
and for $p = \infty$

$$B_1(\mathbf{0}) = \{[x\ y]^T \,|\max\{|x|, |y|\} < 1\}$$


y
$B_1^{\ell_2^1}(\mathbf{0})$
x


y
$B_1^{\ell_2^2}(\mathbf{0})$
x





Example 4.15 The closed unit ball for $C([0,1])$ is given by

$$\overline{B}_1(0) = \{f|||f||_\infty \leq 1\}$$

which is any function $f \in C([0,1])$ contained in the unit rectangle.

Definition 4.20 (Interior) Suppose (X,d) is a metric space and $x_0 \in S \subseteq X$. We say that x_0 is an ***interior point for S*** if there is an $r > 0$ such that $B_r(x_0) \subseteq S$. The ***interior of S***, denoted Int$S$ of S° is the set of all interior points for S. We say that S is ***open in X*** if $S = S^\circ$.

Example 4.16 In any metric space X, both X and $\emptyset$ are always open. $\emptyset$ is open since it contains only interior points, vacuously. If $x_0 \in X$, then $B_r(x_0) \subseteq X$ for any $r > 0$ by definition.

Example 4.17 Any open ball is open. Fix $x_0 \in X$ and $r > 0$. Let $y \in B_r(x_0)$. Let $\delta = r - d(y,x_0) > 0$. Then for all $z \in B_\delta(y)$, $d(z,x) \leq d(z,y) + d(y,x_0) < r - d(y,x_0) + d(y,x_0) = r$, so $z \in B_r(x_0)$. Thus $B_\delta(y) \subseteq B_r(x_0)$, so the ball is indeed open.

Example 4.18 Suppose X is any set and d the discrete metric. Then every subset of X is open. Indeed, for any x_0, $B_{1/2}(x_0) = \{x_0\}$. Thus, if $S \subseteq X$, for all $x \in S$, $B_{1/2}(x) \subseteq S$ so S is open.

Example 4.19 The set $S = \{(x, y)|y > x^2\}$ is open in ℓ_2^p for any $p \geq 1$. Let $f(x, y) = y - x^2$, so $S = f^{-1}((0, \infty)) = \{(x, y)|f(x, y) > 0\}$. Then f is continuous and as we will show later, this implies the inverse image of an open set is open in X.

Proposition 4.12 *Let I be an arbitrary index set and $\{V_i\}_{i \in I}$ a collection of open sets in a metric space X. Then $\bigcup_{i \in I} V_i$ is open in X. If $U_1, ..., U_n$ are open sets in X, then $\bigcap_{i=1}^n U_i$ is open in X.*

Proof Let $V = \bigcup_{i \in I} U_i$, and let $x \in V$. Then there exists $i \in I$ such that $x \in U_i$. As U_i is open there exists $r > 0$ such that $B_r(x) \subseteq U_i \subseteq V$, so V is also open. Let $U = \bigcap_{i=1}^n U_i$ and let $x \in U$. Then for each i, $x \in U_i$ so as U_i is open there exists $r_i > 0$ such that $B_{r_i}(x) \subseteq U_i$. Let $r = \min\{r_1, ..., r_n\} > 0$. Then $B_r(x) \subseteq B_{r_i}(x) \subseteq U_i$ for all i, so $B_r(x) \subseteq U$. Thus, U is indeed open as claimed. □

Example 4.20 The intersection of infinitely many open sets may not be open. Using the usual metric on $\mathbb{R}$, $\bigcap_{n=1}^{\infty}(-1/n, 1/n) = \{0\}$, which is not open in $\mathbb{R}$.

Definition 4.21 Let X be a metric space. We say that $S \subseteq X$ is ***closed in** X* if $X \backslash S$ is open in X.

Example 4.21 In any metric space X, the subsets X and $\emptyset$ are always closed.

Example 4.22 Closed balls are closed:

$$\overline{B}_r(x_0) = \{y \in X|d(x_0, y) \leq r\}$$

So $X \backslash \overline{B}_r(x_0) = \{y \in X|d(x_0, y) > r\}$ This is open by the same argument used for $B_r(x_0)$.

Example 4.23 In any discrete space, all subsets are closed. Every $X \backslash Y$ is open, so $Y = X \backslash (X \backslash Y)$ is closed.

Proposition 4.13 *Let I be an arbitrary index set and $\{S_i\}_{i \in I}$ a collection of closed sets in a metric space X. Then $\bigcap_{i \in I} S_i$ is closed in X. If $T_1, ..., T_n$ are closed sets in X, then $\bigcup_{i=1}^n T_i$ is closed in X.*

The proof of these statements follow by the corresponding statements for open sets and deMorgan's laws.

Example 4.24 In $\mathbb{R}$ the intervals (a, b), $(-\infty, b)$, (a, ∞), and $\mathbb{R}$ are open, the intervals $[a, b]$, $[a, \infty)$, $(-\infty, b]$ and $\mathbb{R}$ are closed, and $(a, b]$ and $[a, b)$ are neither.

With all of this discussion of open and closed, we not that this rely's heavily on what our parent space is. For example, if $Z \subseteq Y \subseteq X$ are all metric spaces with the same metric, questions like: is Z open in Y? Open in X? Do these coincide? are answered depending on the exact sets X and

Y, with the answer to the third being in general no. Consider $Y = [0, 1)$ as a metric space under $d(x, y) = |x - y|$. Y is both open and closed in itself, but neither in $\mathbb{R}$. A ball in Y is given by $B_r(y_0) = \{y \in Y | d(y_0, y) < r\}$. For instance, $[1/2, 1)$ is closed in Y since $Y \backslash [1/2, 1) = [0, 1/2)$ which is open. Why? Well $(0, 1/2)$ are all interior points evidently, but $\{0\}$ is also an interior point since $B_{1/4}(0) = \{y \in [0, 1) || y| < 1/4\} = [0, 1/4) \subseteq [0, 1/2)$.

Proposition 4.14 *Suppose X is a metric space, and $Y \subseteq X$ is the induced metric space taken by restricting d. Then $V \subseteq Y$ is open in Y if and only if $V = Y \cap U$ for some U which is open in X.*

Proof Suppose $V \subseteq Y$ is open in Y. Then, for all $x \in V$, there exists $r_x > 0$ such that $B^Y_{r_x}(x) = \{y \in Y | d(x, y) < r_x\} \subseteq V$. But,

$$B^Y_{r_x}(x) = \{y \in X | d(x, y) < r_x\} \cap Y = B_{r_x}(x) \cap Y$$

Then, observe that

$$V = \bigcup_{x \in V} B^Y_{r_x}(x) = \bigcup_{x \in V} B_{r_x}(x) \cap Y = Y \cap \bigcup_{x \in V} B_{r_x}(x)$$

But, $B_{r_x}(x)$ is open in X for all x, so the union is also open in X, and $V = Y \cap U$ for $U = \bigcup_{x \in V} B_{r_x}(x)$ open in X.

Conversely, suppose U is open in X and let $V = U \cap Y$. Let $x \in V$. Then there exists $r > 0$ such that $B_r(x) \subseteq U$ since U is open. Then

$$B^Y_r(x) = B_r(x) \cap Y \subseteq U \cap Y = V$$

so V is open in Y, as claimed. □

Example 4.25 (Products of open sets) Suppose U_1 and U_2 are open in $(\mathbb{R}, |\cdot|)$. Show that $U_1 \times U_2 = \{(x, y) | x \in U_1, y \in U_2\}$ is open in ℓ^2_2. Let $(x_0, y_0) \in U_1 \times U_2$. Then there exists $r_i > 0$ such that $B_{r_1}(x_0) \subseteq U_1$ and $B_{r_2}(y_0) \subseteq U_2$. Let $r = \min\{r_1, r_2\}/2$. We claim $B_r((x_0, y_0)) \subseteq U_1 \times U_2$. If $(x, y) \in B_r((x_0, y_0))$, then

$$|x - x_0| \leq \sqrt{(x - x_0)^2 + (y - y_0)^2} < r \leq r_1/2 < r_1$$

and identially for $|y - y_0|$, so $x \in U_1$ and $y \in U_2$, so $(x, y) \in U_1 \times U_2$. Thus $B_r((x_0, y_0)) \subseteq U_1 \times U_2$ so $U_1 \times U_2$ is open.

If $X_1, ..., X_n$ are metric spaces with metrics $d_1, ..., d_n$, then for any $p \geq 1$ the set $X_1 \times ... \times X_n$ is a metric space under the metric

$$d\left((x_1, ..., x_n), (y_1, ..., y_n)\right) := \left[\sum_{i=1}^{n} d_i(x_i, y_i)^p\right]^{1/p}$$

4.3.1 Boundary and Limit Points

Definition 4.22 Suppose X is a metric space and $S \subseteq X$. We say that $x_0 \in X$ is a ***boundary point for S*** if for all $r > 0$, the intersections $B_r(x_0) \cap S$ and $B_r(x_0) \cap (X \backslash S)$ are non-empty. The ***boundary of S***, denoted BdS or ∂S, is the set of all boundary points for S.

x_0 is said to be an ***accumulation point for S*** (or limit point, or cluster point) if for all $r > 0$, the intersection $(B_r(x_0) \backslash \{x_0\}) \cap S$ is non-empty. The set of accumulation points for S is denoted S', and called the ***derived set***.

$x_0 \in S$ is said to be an ***isolated point for S*** if x_0 is not in S'. Equivalently, there exists $r > 0$ such that $(B_r(x_0) \backslash \{x_0\}) \cap S = \emptyset$.

The interior and isolated points for S belong to S, by definition. But boundary points and accumulation points for S may or may not belong to S.

Example 4.26 (Subsets of $\mathbb{R}$*)* Let $S = (-\infty, -3) \cup [1, 2) \cup \{3\} \cup \{\pi\}$. Then $S^\circ = (-\infty, -3) \cup (1, 2)$, $\partial S = \{-3, 1, 2, 3, \pi\}$, $S' = (-\infty, -3] \cup [1, 2]$, and the isolated points are $S \backslash S' = \{3, \pi\}$. Note that $S \cup S' = S \cup \partial S$ (this isn't a coincidence).

Example 4.27 Consider $S = \{(x, y) \in \ell_2^2 | y^2 - x^2 > 1\}$. S is open, so $S^\circ = S$. $\partial S\{(x, y) | y^2 - x^2 = 1\}$, and in this case $S^\circ \subseteq S'$, so $S' = S \cup \partial S = \{(x, y) | y^2 - x^2 \geq 1\}$. There are no isolated points for this set.

Example 4.28 Consider $\mathbb{Q}^n \subseteq \ell_n^p$. Then $\partial \mathbb{Q}^n = \mathbb{R}^n$, $(\mathbb{Q}^n)' = \mathbb{R}^n$ ($\mathbb{Q}^n$ is dense in $\mathbb{R}^n$), $\mathbb{Q}^{n\circ} = \emptyset$, and there are no isolated points. This follows from $\mathbb{Q}' = \partial \mathbb{Q} = \mathbb{R}$ and $\mathbb{Q}^\circ = \emptyset$ in $\mathbb{R}$ under the usual metric.

Example 4.29 Let V be a normed linear space, $r > 0$, and $v_0 \in V$. Then $B_r(v_0)$ is open so $B_r^\circ(v_0) = B_r(v_0)$. Further, $\partial(B_r(v_0)) = \{v \in V | ||v - v_0|| = r\}$ (the sphere of radius r).

Example 4.30 If X is any set under the discrete metric, we've already seen that any $S \subseteq X$ is both open and closed. Every point in S is isolated since $B_{1/2}(x_0) = \{x_0\} \subseteq S$. But, $\partial S = \emptyset$ since $B_{1/2}(x_0) = \{x_0\}$, so either $B_r(x_0) \cap S$ or $B_r(x_0) \cap (X \backslash S)$ is empty.

Proposition 4.15 *A subset S of a metric space X is closed in X if and only if* $S' \subseteq S$.

Proof Suppose S is closed in X and $x_0 \in S'$. For the sake of contradiction suppose $x_0 \in X \backslash S = U$, which is open. Then there exists $r > 0$ such that $B_r(x_0) \subseteq U$. But then $B_r(x_0) \cap S = \emptyset$, which implies $(B_r(x_0) \backslash \{x_0\}) \cap S = \emptyset$, contradicting the fact that $x_0 \in S'$. Thus, $S' \subseteq S$.

Conversely, suppose $S' \subseteq S$, and towards a contradiction suppose S is not closed, so $U = X \backslash S$ is not open. Then there exists $x \in U$ such that for all $r > 0$, $B_r(x) \nsubseteq U$, and in particular $(B_r(x) \cap \{x\}) \cap S \neq \emptyset$ for all $r > 0$. Thus, by definition $x \in S'$. But then $x \in S$ as $S' \subseteq S$, and $x \in X \backslash S$ which implies $x \notin S$, which is a contradiction. Consequently we conclude that S must be closed. □

Proposition 4.16 *Suppose X is a metric space with $x_0 \in X$ and $S \subseteq X$. Then $x_0 \in S'$ if and only if there is a sequence $\{x_n\}$ in $S\backslash\{x_0\}$ with x_n converging to x_0.*

Proof If $x_0 \in S'$, then for all $n \in \mathbb{N}$, $B^*_{1/n}(x_0) \cap S \neq \emptyset$, where $B^*_{1/n}(x_0) = (B_{1/n}(x_0)\backslash\{x_0\})$. Then by the axiom od choice, for each$n \in \mathbb{N}$ we can choose $x_n \in (B_{1/n}(x_0)\backslash\{x_0\}) \cap S$. Then $0 < d(x_0, x_n) < 1/n$ for each n. Thus $\lim\limits_{n\to\infty} d(x_0, d_n) = 0$, so $\lim\limits_{n\to\infty} x_n = x_0$.

Conversely, suppose there exists $(x_n) \subseteq S\backslash\{x_0\}$ with $d(x_n, x_0) \to 0$. If $r > 0$, find $N \in \mathbb{N}$ such that if $n \geq N$, $d(x_n, x_0) < r$. Then $x_N \in B^*_r(x_0) \cap S$, so as r was arbitrary $x_0 \in S'$. □

Example 4.31 The ***general linear group*** over the reals of order 2 is given by

$$\mathbf{GL}_2(\mathbb{R}) := \{A \in M_2(\mathbb{R}) | \det A \neq 0\}$$

Think of $\mathbf{GL}_2(\mathbb{R})$ as a subset of ℓ^2_4 (as a metric space). That is,

$$d\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix}, \begin{bmatrix} x & y \\ z & t \end{bmatrix}\right) := \sqrt{(a-x)^2 + (b-y)^2 + (c-z)^2 + (d-t)^2}$$

Then $\mathbf{GL}_2(\mathbb{R})$ is open in $M_2(\mathbb{R})$. We will show $M_2(\mathbb{R})\backslash GL_2(\mathbb{R})$ is closed. Combining the above two results, it suffices to show that if $A_n = \begin{bmatrix} a_n & b_n \\ c_n & d_n \end{bmatrix} \in M_2(\mathbb{R})\backslash\mathbf{GL}_2(\mathbb{R})$ and $A_n \to A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in M_2(\mathbb{R})$, then $A \in M_2(\mathbb{R})\backslash\mathbf{GL}_2(\mathbb{R})$. We know in ℓ^2_4 that $\begin{bmatrix} a_n & b_n \\ c_n & d_n \end{bmatrix} \to \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ if and only if $a_n \to a, b_n \to b, c_n \to c$, and $d_n \to d$. Then $0 = a_n d_n - b_n c_n \to ad - bc = \det(A)$, so $\det(A) = 0$ and $A \in M_2(\mathbb{R})\backslash\mathbf{GL}_2(\mathbb{R})$ as desired.

Definition 4.23 (Closures) Suppose X is a metric space and $S \subseteq X$. The ***closure of*** S ***in*** X is the set

$$\overline{S} = \bigcap\{B \subseteq X | S \subseteq B \text{ and } B \text{ is closed}\}$$

Note that $\overline{S}$ is an intersection of closed sets, and hence closed. Additionally, $S \subseteq \overline{S}$ by construction. Finally, $\overline{S}$ is the smallest closed set containing S. If $S \subseteq C$ and C is closed, then C is a part of the collection being intersected so $C \supseteq \overline{S}$.

Proposition 4.17 *Suppose X is a metric space and $S \subseteq X$. Then S is closed in X if and only if $\overline{S} = S$.*

Proof If $S = \overline{S}$, then S is closed from the discussion above. If S is closed then it is in the collection being intersected so $\overline{S} \subseteq S$. But $S \subseteq \overline{S}$, by construction, so $S = \overline{S}$. □

Proposition 4.18 *Suppose X is a metric space and $S \subseteq X$. Then S is closed in X if and only if whenever x_n is a convergent sequence in S then its limit also belongs to S.*

Proof If S is closed, we know $S' \subseteq S$. Suppose $(x_n) \subseteq S$, with $x_n \to x \in X$. If $x \in S'$, then $x \in S$ and we are done. Otherwise $x \notin S'$. So there exists $r > 0$ such that $B^*_r(x) \cap S = \emptyset$. But, there exists $N \in \mathbb{N}$ such that $x_n \in B_r(x)$ for $n \geq N$, while $x_n \notin B^*_r(x)$, so $x_n \in \{x\}$ for all $n \geq N$. Thus, $x = x_N \in S$.

Conversely, suppose S has the property that whenever $(x_n) \subseteq S$ with $x_n \to x \in X$, then $x \in S$. We know if $x_0 \in S'$, there exists $(x_n) \subseteq S$ with $x_n \to x_0$, so $x_0 \in S$ by assumption. Hence $S' \subseteq S$ and S is closed. □

Example 4.32 Suppose $A \subseteq B \subseteq X$. We claim $A^\circ \subseteq B^\circ$. Indeed, if $x_0 \in A^\circ$, then there exists $r > 0$ such that $B_r(x_0) \subseteq A \subseteq B$, and hence $x_0 \in B^\circ$. However, both statements $\partial A \subseteq \partial B$ and $\partial B \subseteq \partial A$ are false in general. Take $A = [0, 1]$, $B = [-1, 2]$. Then $\partial A = \{0, 1\}$ and $\partial B = \{-1, 2\}$.

Example 4.33 Consider the balls $B_r(x_0) = \{x \in X | d(x, x_0) < r\}$ and $\overline{B}_r(x_0) = \{x \in X | d(x, x_0) \leq r\}$. What is $\overline{B_r(x_0)}$? First, we have $B_r(x_0) \subseteq \overline{B}_r(x_0)$, which is closed, so $\overline{B_r(x_0)} \subseteq \overline{B}_r(x_0)$. This inclusion can be in some cases strict:

Let X be any set with at least two elements equipped with the discrete metric. Then $\overline{B}_1(x_0) = X$, but $B_1(x_0) = \{x_0\}$ is closed so $\overline{B_1(x_0)} = \{x_0\} \neq X$.

In a normed linear space, we always have $\overline{B_r(v_0)} = \overline{B}_r(v_0)$. First look at $v_0 = 0$. Then $B_r(0) = \{v | ||v|| < r\}$ and if $||v|| = r$ then let $v_n = (1 - 1/n)v$; Then $||v_n|| = (1 - 1/n)r < r$, so $v_n \in B_r(0)$ for all n and $||v_n - v|| \to 0$. This implies $v \in B_r(0)'$, so $v \in \overline{B_r(0)}$ and $\overline{B_r(0)} = \overline{B}_r(0)$. If $v_0 \neq 0$, we use the identities

$$B_r(v_0) = \{v | ||v - v_0|| < r\} = \{v + v_0 | ||v|| < r\}$$

and $\overline{B}_r(v_0) = \{v + v_0 | ||v|| \leq r\}$. Then, take $v_n = v_0 + (1 - 1/n)v$ for $||v|| = r$, and the same result holds.

Proposition 4.19 *Suppose X is a metric space and $S \subseteq X$. Then*

(1)$S^\circ = \bigcup\{U \subseteq S | U$ is open in $X\}$

(2)S is open if and only if $S^\circ = S$.

Proof (1): Suppose $x_0 \in S^\circ$. Then there exists $r > 0$ such that $B_r(x_0) \subseteq S$. But $B_r(x_0)$ is open, so $x_0 \in$ union. If $x_0 \in$ union, then $x_0 \in U$ for some open set $U \subseteq S$. U being open implies there exists $r > 0$ such that $B_r(x_0) \subseteq U \subseteq S$, so $x_0 \in S^\circ$ and we have set equality.

(2): S° is open, so evidently if $S^\circ = S$ then S is open. Conversely, if S is open then $x_0 \in S^\circ$ for all $x_0 \in S$, so $S^\circ \subseteq S \subseteq S^\circ$. Hence $S = S^\circ$. □

Example 4.34 Note that $B_r(x_0) = B_r(x_0)^\circ$ as they are open. As $B_r(x_0) \subseteq \overline{B}_r(x_0)$, we also have $B_r(x_0) = B_r(x_0)^\circ \subseteq [\overline{B}_r(x_0)]^\circ$. In a normed linear space we have equality, but in a discrete space, equality may fail.

? Question

If S is closed, is it true that $\overline{S^\circ} = S$? If S is open, is it true that $\overline{S}^\circ = S$?

The answer is no to both! In $\mathbb{R}$, let $S = \mathbb{Z}$. Then $S^\circ = \emptyset$, and $\overline{\emptyset} = \emptyset \neq \mathbb{Z}$. For the second, $\overline{S} = \mathbb{Z}$ since the space has no accumulation points, and $S^\circ = \emptyset \neq \mathbb{Z}$.

Definition 4.24 (Bounded Set) A subset S of a metric space X is said to be ***bounded*** if there is an $x_0 \in X$ and $r > 0$ such that $S \subseteq B_r(x_0)$.

Example 4.35 The usual intervals $(a, b), [a, b), (a, b], [a, b]$ are all bounded. All of which are contained in $B_{2\max\{|a|,|b|\}}(0)$.

Example 4.36 Bounded subsets of a normed linear space are of the form $S \subseteq B_r(v)$, which implies $||v - v_0|| < r$ for all $v \in S$, or $||v|| < r + ||v_0|| = M$ for all $v \in S$, which implies $S \subseteq B_M(0)$. Thus, S is bounded if and only if $||v|| \leq M$ for all $v \in S$ and for some $M > 0$.

Theorem 4.3 (Bolzano-Weierstrass) *Suppose $p \in [1, \infty]$ and $n \in \mathbb{N}$. Any bounded sequence in ℓ_n^p has a convergent subsequence.*

Proof Let $(\mathbf{x}_k) = ((x_{k,1}, ..., x_{k,n}))$ be a bounded sequence in ℓ_2^p, so there exists $M > 0$ such that

$$||\mathbf{x}_k||_p \leq M$$

for all $k \in \mathbb{N}$. If $p < \infty$, we have $|x_{k,j}| = (|x_{k,j}|^p)^{1/p} \leq M$, and for $p = \infty$ we have directly $|x_{k,j}| \leq M$, for all k and all $1 \leq j \leq n$. So by the Bolzano-Weierstrass theorem for $\mathbb{R}$, there exists a convergent subsequence $x_{s_1(k),1}$ converging to $x \in \mathbb{R}$. Then $x_{s_1(k),2}$ also has a convergent subsequence $x_{s_2(k),2}$ converging to some $x_2 \in \mathbb{R}$. Continuing in this manner we have that for all $1 \leq j \leq n - 1$ there exists a subsequence $x_{s_{j+1}(k),j+1}$ of $x_{s_j(k),j+1}$ converging to some $x_{j+1} \in \mathbb{R}$. Then $(\mathbf{x}_{s_n(k)})$ is a subsequence in which for all $1 \leq j \leq n$, $x_{s_n(k),j}$ converges to $x_j \in \mathbb{R}$, so $\mathbf{x}_{s_n(k)} \to \mathbf{x} = (x_1, ..., x_n) \in \ell_n^p$. □

Corollary 4.3 *Any bounded and infinite subset S of ℓ_n^p has an accumulation point.*

Proof Choose a sequence of distinct points $(\mathbf{x}_k)$ in S, which is possible as S is infinite. This sequence is bounded and hence has a convergent subsequence. Let $\mathbf{x} \in X$ be the limit. If x occurs in the sequence at $N \in \mathbb{N}$, then $\mathbf{x} \notin (\mathbf{x}_{k_j})_{j>N}$ as the points are distinct, for our convergent subsequence. Thus, $(\mathbf{x}_{k_j})_{j>N} \subseteq S \backslash \{\mathbf{x}\}$, so by definition $\mathbf{x} \in S'$. □

4.4 Completion of a Metric Space

As in the case of $\mathbb{Q}$ to $\mathbb{R}$, we can form the completion of a general metric space.

Definition 4.25 If (X, d) is not complete, we can define its completion $(\hat{X}, \hat{d})$ by taking $\hat{X} = \{[(x_j)_{j=1}^\infty] : (x_j)_{j=1}^\infty \subseteq X \text{ is Cauchy}\}$, where $[(x_j)]$ is the the equivalence class defined by $(x_j) \sim (x'_j)$ if and only if $d(x_j, x'_j) \to 0$ in $\mathbb{R}$. For $\xi = [(x_j)]$ and $\eta = [(y_j)]$, we define

$$\hat{d}(\xi, \eta) := \lim_{j\to\infty} d(x_j, y_j)$$

It remains to show that $\hat{d}$ is a well defined metric. Note by the triangle inequality $d(x,y) \leq d(x,z) + d(z,y)$, so $d(x,y) - d(x,z) \leq d(y,z)$ and using $d(x,z) \leq d(x,y) + d(y,z)$, $d(x,z) - d(x,y) \leq d(y,z)$. Thus

$$|d(x,z) - d(x,y)| \leq d(y,z)$$

Then suppose $(x_j) \sim (x'_j)$ and $(y_j) \sim (y'_j)$. It follows that

$$\begin{aligned} |d(x_j,y_j) - d(x'_j,y'_j)| &= |d(x_j,y_j) - d(x_j,y'_j) + d(x_j,y'_j) - d(x'_j,y'_j)| \\ &\leq |d(x_j,y_j) - d(x_j,y'_j)| + |d(x_j,y'_j) - d(x'_j,y'_j)| \\ &\leq d(y_j,y'_j) + d(x_j,x'_j) \end{aligned}$$

which goes to 0 as $(x_j) \sim (x'_j)$ and $(y_j) \sim (y'_j)$. This proves that if the limit exists, then it is independent of representative. Now, as $\mathbb{R}$ is complete, to show $\hat{d}(\xi,\eta) = \lim\limits_{j\to\infty} d(x_j,y_j)$ exists it is sufficient to show $d(x_j,y_j)$ is Cauchy. Then

$$\begin{aligned} |d(x_j,y_j) - d(x_k,y_k)| &= |d(x_j,y_j) - d(x_j,y_k) + d(x_j,y_k) - d(x_k,y_k)| \\ &\leq |d(x_j,y_j) - d(x_j,y_k)| + |d(x_j,y_k) - d(x_k,y_k)| \\ &\leq d(y_j,y_k) + d(x_j,x_k) \end{aligned}$$

which goes to 0 as j and k go to ∞ as (y_j) and (x_j) are Cauchy, so $d(x_j,y_j)$ is Cauchy. Thus $\hat{d}$ is well defined. Further, $\hat{d}(\xi,\eta) = \lim\limits_{j\to\infty} d(x_j,y_j) \geq 0$, $\hat{d}(\xi,\eta) = 0$ if and only if $d(x_j,y_j) \to 0$, which holds if and only if $\xi = \eta$ by definition of $\sim$,

$$\hat{d}(\xi,\eta) = \lim_{j\to\infty} d(x_j,y_j) = \lim_{j\to\infty} d(y_j,x_j) = \hat{d}(\eta,\xi)$$

and

$$\hat{d}(\xi,\eta) = \lim_{j\to\infty} d(x_j,y_j) \leq \lim_{j\to\infty} (d(x_j,z_j) + d(z_j,y_j) = \hat{d}(\xi,\mu) + \hat{d}(\mu,\eta)$$

for any $\mu = [(z_j)] \in \hat{X}$. Thus $(\hat{X},\hat{d})$ is indeed a metric space.

Example 4.37 Consider $\mathbb{Q}[x]$, the set of polynomials over $\mathbb{Q}$, and define $d(p,q) = \max_{x\in[0,1]} |p(x) - q(x)|$. The max exists as $[0,1]$ is a compact set in $\mathbb{R}$ and all polynomials are continuous. Then $d(p,q) = 0$ if and only if $p = q$, as polynomials of degree > 0 have only a finite number of roots, $d(p,q) = d(q,p)$, and $d(p,q) \leq d(p,r) + d(r,q)$ using the triangle inequality for $|\cdot|$ on $\mathbb{R}$. Note this is a countable metric space. But, this is not complete as $\mathbb{Q}$ is not complete. Then $(\hat{X},\hat{d}) = \{[p_j] : (p_j) \text{ is Cauchy in } (\mathbb{Q}[x],d)\}$. We will see that $\hat{X} = C([0,1])$, all continuous functions in $[0,1]$. This says if f is continuous in $[0,1]$, for all $\varepsilon > 0$ there exists $p \in \mathbb{Q}[x]$ such that $\hat{d}(p,f) < \varepsilon$. Suppose now we define $X = \mathbb{Q}[x]$ with the distance

$$d_1(p,q) = \max_{x\in[0,1]} |p(x) - q(x)| + \max_{x\in[0,1]} |p'(x) - q'(x)|$$

Note $X \subseteq \mathbb{C}^\infty([0,1])$. Upon completion we obtain $(\hat{X},\hat{d}_1) = C^1([0,1])$, the space of all continuous functions with continuous first derivative.

We claim that $\hat{X}$ is indeed a complete metric space:

Lemma 4.3 *X is dense in $\hat{X}$.*

and

Proposition 4.20 *$(\hat{X}, \hat{d})$ is complete.*

which follow similarly to the case of $\mathbb{R}$.

We now proceed with a full derivation:

Theorem 4.4 (Completion of a Metric Space) *Suppose (X, d) is a metric space. There exists a complete metric space $(\hat{X}, \hat{d})$ and an isometry $\varphi : X \to \hat{X}$ so that $\varphi(X)$ is dense in $\hat{X}$ (i.e. that $\overline{\varphi(X)} = \hat{X}$) Moreover, $\hat{X}$ is uniquely determined by this property up to homeomorphism.*

Definition 4.26 The space $\hat{X}$ above is called the ***completion*** of X.

What do we mean when we say unique? Suppose (Y, d_Y) is a metric space with an isometry $\varphi : X \to Y$ such that $\overline{\varphi(X)} = Y$, with Y complete. Then there is a bijective isometry $\Gamma : Y \to \hat{X}$ such that

$$\begin{tikzcd}
X \arrow[r, "\psi"] \arrow[d, "\varphi"'] & Y \arrow[dl, dashed, "\exists!\Gamma"] \\
\hat{X} &
\end{tikzcd}$$

the diagram commutes. Defin $\gamma : \psi(X) \to \hat{X}$ by $\gamma(\psi(x)) := \varphi(x)$ for all $x \in X$. Then γ is an isometry:

$$\hat{d}(\gamma(\psi(x)), \gamma(\psi(y)) := \hat{d}(\varphi(x), \varphi(y)) = d(x, y)$$

since φ is an isometry. Let's extend γ to a map $\Gamma : Y \to \hat{X}$ by

$$\Gamma(y) = \lim_{n\to\infty} \gamma(\psi(x_n)) = \lim_{n\to\infty} \varphi(x_n)$$

where $\psi(x_n) \to y$, where (x_n) exists since $\overline{\psi(X)} = Y$. Then the following hold:

- Γ is well-defined and isometric
- Γ is onto

Note that if $S \subseteq X$ and X is complete, then $\hat{S} \cong \overline{S}$. In this case, Γ is just the identity $\mathrm{Id} : S \to \overline{S}$, and note $\overline{S}$ is complete since X is complete.

Example 4.38 (The Hardy space of the disk) Let $P = \{a_0 + a_1 z + ... + z_n z^n | n \in \mathbb{N} \cup \{0\}, a_i \in \mathbb{C}\}$ denote the polynomials with complex variable z. Define a norm $|| \cdot ||_2$ on P as follows

$$||a_0 + a_1 z + ... + a_n x^n||_2 = \sqrt{\sum_{i=0}^{n} |a_i|^2}$$

P is infinite dimensional with basis $\{1, z, z^2, z^3, ...\}$, but P is not complete. Let $p_n(z) = \sum_{k=0}^{n} \frac{z^k}{k!}$. Then $\{p_n\}$ is Cauchy, but cannot converge to any $g(z) = b_0 + b_1 z + ... + b_m z^m$ with respect to $|| \cdot ||_2$, since for $n >> m$,

$$||p_n - g||_2^2 = \sum_{k=0}^{m} \left| \frac{1}{k!} - b_k \right|^2 + \sum_{k=m+1}^{n} \frac{1}{(k!)^2} \geq \frac{1}{[(m+1)!]^2} > 0$$

So as $n \longrightarrow \infty$, there is no way the right hand side goes to 0. The completion of P is called $H^2(\mathbb{D})$, the ***Hardy space of the unit disk***. We define

$$H^2(\mathbb{D}) := \{f : \mathbb{D} \longrightarrow \mathbb{C} | f \text{ is analytic and } f(z) = \sum_{n=0}^{\infty} a_n z^n \text{ with } \sum_{n=0}^{\infty} |a_n|^2 < \infty\}$$

where $\mathbb{D} = \{z \in \mathbb{C} | |z| < 1\}$ with norm $||f||_2 = \sqrt{\sum_{n=0}^{\infty} |a_n|^2}$.

We know prove the completion theorem:

Proof Let (X, d) be a metric space. Let X' denote the set of Cauchy sequences in X. Define a relation on X' as follows: $(x_n) \sim (y_n)$ if and only if $d(x_n, y_n) \longrightarrow 0$. $\sim$ is an equivalence relation. Indeed, $d(x_n, x_n) = 0 \longrightarrow 0$, $d(y_n, x_n) = d(x_n, y_n) \longrightarrow 0$, and if $(x_n) \sim (y_n) \sim (z_n)$, then

$$d(x_n, z_n) \leq d(x_n, y_n) + d(y_n, z_n) \longrightarrow 0$$

Let $\hat{X}$ denote the set of aequivalence classes arising from $\sim$. Given $[(x_n)]$ and $[(y_n)]$ in $\hat{X}$, define $a_n := d(x_n, y_n)$. (a_n) is Cauchy in $\mathbb{R}$ since

$$|a_n - a_m| = |d(x_n, y_n) - d(x_m, y_m)| \leq d(x_n, x_m) + d(y_n, y_m) \longrightarrow 0$$

since (x_n) and (y_n) are Cauchy in X. $\mathbb{R}$ is complete with respect to $| \cdot |$, so a_n has a limit, say a. Define $\hat{d}$ on $\hat{X} \times \hat{X}$ by

$$\hat{d}([(x_n)], [(y_n)]) := \lim_{n\to\infty} a_n = \lim_{n\to\infty} d(x_n, y_n)$$

We must show $\hat{d}$ is well defined, so suppose $(x_n) \sim (x'_n)$ and $(y_n) \sim (y'_n)$. Then

$$|d(x_n, y_n) - d(x'_n, y'_n)| \leq d(x_n, x'_n) + d(y_n, y'_n) \longrightarrow 0$$

so $\lim\limits_{n\to\infty} d(x_n, y_n) = \lim\limits_{n\to\infty} d(x'_n, y'_n)$.

Since d is a metric, $\hat{d}$ is a metric as it is the limit of this metric and limits are additive and satisfy order preservation properties. Define $\varphi : X \longrightarrow \hat{X}$ by $\varphi(x) = [(x, x, x, ...)]$, which is the equivalence class of all Cauchy sequences (y_n) with $d(y_n, x) \longrightarrow 0$, i.e. all sequences converging to x. φ is an isometry since

$$\hat{d}(\varphi(x), \varphi(y)) = \lim_{n\to\infty} d(x, y) = d(x, y)$$

Next, to show $\overline{\varphi(X)} = \hat{X}$, let $\hat{x} = [(x_n)] \in \hat{X}$ and fix $\varepsilon > 0$. (x_n) is Cauchy in X by construction. So there exists $N \in \mathbb{N}$ such that for $m \geq N$, $d(x_N, x_m) < \varepsilon$. Then it follows that

$$\hat{d}(\hat{x}, \varphi(x_N)) = \lim_{n\to\infty} d(x_n, x_N) \leq \varepsilon$$

Finally, we show $\hat{X}$ is complete. Suppose $(\hat{x}_k)$ is a Cauchy sequence in $\hat{X}$. For $k \in \mathbb{N}$, choose $\hat{z}_k$ in $\varphi(X)$ such that

$$\hat{d}(\hat{x}_k, \hat{z}_k) < \frac{1}{k}$$

Then

$$\hat{d}(\hat{z}_k, \hat{z}_m) \leq \hat{d}(\hat{z}_k, \hat{x}_k) + \hat{d}(\hat{x}_k, \hat{x}_m) + \hat{d}(\hat{x}_m, \hat{z}_m) \leq \frac{1}{k} + \hat{d}(\hat{x}_k, \hat{x}_m) + \frac{1}{m}$$

which goes to zero as m and k go to infinity. So $(\hat{z}_k)$ is Cauchy in $\varphi(X)$. Noting that $\varphi : X \longrightarrow \varphi(X)$ is invertible since φ is Cauchy, let $y_k = \varphi^{-1}(\hat{z}_k) \in X$. Then (y_k) is Cauchy in X since $(\hat{z}_k)$ is Cauchy in $\varphi(X)$, and φ^{-1} is an isometry. Let $\hat{y}$ be the equivalence class in $\hat{X}$ which includes (y_k). Then

$$\hat{d}(\hat{x}_k, \hat{y}) \leq \hat{d}(\hat{x}_k, \hat{z}_k) + \hat{d}(\hat{z}_k, \hat{y}) < \frac{1}{k} + \lim_{n\to\infty} d(y_k, y_n)$$

which goes to zero as (y_k) is Cauchy, so $\hat{y}$ is the limit of $\hat{x}_k$ in $\hat{X}$, and $\hat{X}$ is complete! □

4.5 Contractive Maps and Fixed Points

Definition 4.27 (Contractive Map) Suppose X is a metric space. A function $f : X \longrightarrow X$ is a ***contractive map*** (distance shrinking) if there is a constant $r \in (0, 1)$ with $d(f(x), f(y)) \leq r d(x, y)$ for all $x, y \in X$.

Contractive maps are ***uniformly continuous***, as they are Lipschitz.

Example 4.39 A bounded operator $T : V \longrightarrow V$ is contractive if $||T|| \leq r < 1$ for some r, so

$$||Tv - Tw|| \leq ||T||||v - w|| \leq r||v - w||$$

Theorem 4.5 (Banach Fixed Point Theorem (or the Contraction Mapping Principle)) *Suppose X is a* **complete metric space** *and $f : X \longrightarrow X$ is a contractive map. Then f has a* **unique** *fixed point. That is there is a unique $x_0 \in X$ such that $f(x_0) = x_0$.*

Proof For uniqueness suppose x_1 and x_2 are both fixed points, $f(x_1) = x_1$ and $f(x_2) = x_2$. Then

$$d(x_1, x_2) = d(f(x_1), f(x_2)) \leq r d(x_1, x_2)$$

for some $0 < r < 1$. This holds if and only if $d(x_1, x_2) = 0$, which is to say $x_1 = x_2$.

For existence suppose $x_1 \in X$. Define $x_2 = f(x_1)$, $x_2 = f(x_2)$, and in general $x_n = f^{n-1}(x_1)$. For $n \in \mathbb{N}$ and $k \in \mathbb{N}$,

$$\begin{aligned} d(x_{n+k}, x_n) &= d(f^{n+k-1}(x_1), f^{n-1}(x_1)) \\ &\leq r^{n-1} d(f^k(x_1), x_1) = r^{n-1} d(x_{k+1}, x_1) \\ &\leq r^{n-1}[d(x_{k+1}, x_k) + d(x_k, x_{k-1}) + \ldots + d(x_2, x_1)] \\ &= r^{n-1} \sum_{j=0}^{k-1} r^j d(x_2, x_1) \\ &\leq r^{n-1} \frac{1}{1-r} d(x_2, x_1) \end{aligned}$$

This implies (x_n) is Cauchy in X, and so there exists x_0 with $x_n \longrightarrow x_0$ by completeness. Then

$$f(x_0) = f(\lim_{n\to\infty} x_n) = \underbrace{\lim_{n\to\infty} f(x_n)}_{\text{continuity}} = \lim_{n\to\infty} x_{n+1} = x_0$$

so $f(x_0) = x_0$. □

Theorem 4.6 (Picard's Theorem for First Order Differential Equations) *Suppose* $(x_0, y_0) \in \ell_2^2$*,* $r > 0$*, and let* $B = B_r(x_0, y_0)$*. If* $f : B \longrightarrow \mathbb{R}$ *is continuous and bounded and there is a* $C > 0$ *such that*

$$|f(x, y_1) - f(x, y_2)| \leq C|y_1 - y_2|$$

for all pairs (x, y_1) *and* (x, y_2) *in B, then there exists a unique solution y to the initial value problem*

$$\frac{dy}{dx} = f(x, y); y(x_0) = y_0$$

on the interval $(x_0 - \delta, x_0 + \delta)$ *for some* $\delta > 0$*.*

Proof Let $M = \sup_{(x,y)\in B} |f(x, y)|$. Find $\delta > 0$ such that

(i) $\delta < \frac{1}{M}$

(ii)$\{(x, y) | |x - x_0| < \delta, |y - y_0| < M\delta\} \subseteq B$

(iii)$\delta < \frac{1}{C}$

Let $X = \{\varphi \in C([x_0 - \delta, x_0 + \delta]) | |\varphi(x) - y_0| \leq M\delta\}$. X consists of continuous functions whose graph is in the open rectangle described in (ii). Note $C([x_0 - \delta, x_0 + \delta])$ is a complete space, with uniform metric. X is a closed subset of $C([x_0 - \delta, x_0 + \delta])$. Indeed, if $f_n \in X$ and $f_n \longrightarrow_u f$, then $\lim_{n\to\infty} |f_n(x) - y_0| \leq M\delta$, so $|f(x) - y_0| \leq M\delta$. Thus, X is complete itself. Define a map $T : X \longrightarrow X$ by

$$T\varphi(x) = y_0 + \int_{x_0}^{x} f(t, \varphi(t))dt$$

and this process is called the ***Picard iteration***. Note $T\varphi$ is continuous on $[x_0 - \delta, x_0 + \delta]$. Further,

$$|T\varphi(x) - y_0| \leq \left| \int_{x_0}^{x} |f(t, \varphi(t))| dt \right| \leq \delta M$$

so $T\varphi \in X$, and T is well-defined. Further, T is contractive since

$$\begin{aligned} |T\varphi_1(x) - T\varphi_2(x)| &\leq \int_{x_0}^{x} |f(t, \varphi_1(t)) - f(t, \varphi_2(t))| dt \\ &\leq \int_{x_0}^{x} C|\varphi_1(t) - \varphi_2(t)| dt \qquad \leq \int_{x_0}^{x} C \sup_{x_0 - \delta < t < x_0 + \delta} |\varphi_1(t) - \varphi_2(t)| dt \\ &= |x - x_0| C \sup |\varphi_1(t) - \varphi_2(t)| \\ &< \delta C ||\varphi_1 - \varphi_2||_{[x_0 - \delta, x_0 + \delta]} \end{aligned}$$

Taking the supremum

$$||T\varphi_1 - T\varphi_2||_\infty \leq \delta C ||\varphi_1 - \varphi_2||_\infty < ||\varphi_1 - \varphi_2||_\infty$$

since we assumed $\delta < \frac{1}{C}$. So T is contractive and hence has a fixed point. In particular, there exists $\varphi_0 \in X$ with $\varphi_0(x) = T\varphi_0(x) = y_0 + \int_{x_0}^{x} f(t, \varphi_0(t)) dt$. Then $\varphi_0(x_0) = y_0 + 0 = y_0$. Finally, by the FTOC1,

$$\frac{d}{dx}\varphi_0(x) = f(x, \varphi_0(x))$$

which is to say $y = \varphi_0(x)$ solves the DE $\frac{dy}{dx} = f(x, y)$ with initial condition $y(x_0) = y_0$. □

4.6 Compactness for Metric Spaces

Recall that $\emptyset \neq S \subseteq \mathbb{R}$ is ***compact*** if every sequence in S has a convergent subsequence with limit in S. Further, the Heine-Borel theorem told us that $\emptyset \neq S$ is compact if and only if it is closed and bounded. This is not the case for a general metric space, as we shall now show:

Example 4.40 Let $S = \overline{B}_1(0) \subseteq C([0, 1])$ with the supremum norm. S is closed and bounded, but the sequence $f_n(x) = x^n$ has no uniformly convergent subsequences (it converges pointwise to a discontinuous function).

We now introduce a more topologicall useful definition.

Definition 4.28 (Open Covers) Let X be a metric space and A a subset of X. A collection $\{U_i\}_{i \in I}$ of open subsets of X is said to be an ***open cover*** for A if

$$A \subseteq \bigcup_{i \in I} U_i$$

Given such a cover for A, a ***finite subcover for A*** is a finite subcollection $U_{i_1}, ..., U_{i_n}$ of $\{U_i\}_{i \in I}$ such that

$$A \subseteq \bigcup_{j=1}^{n} U_{i_j}$$

Definition 4.29 (Compactness) A non-empty susbet A of a metric space X is ***compact in X*** if every open cover of A in X has a finite subcover. A non-empty subset A is ***sequentially compact in X*** if every sequence in A has a convergent subsequence whose limit is in A.

In other words:

Definition 4.30 X is ***sequentially compact*** if every sequence $(x_j) \subseteq X$ has a convergent subsequence.

And we also have:

Definition 4.31 X is ***limit point compact*** if every infinite subset of X has an accumulation point in X.

These are equivalent for metric spaces by the axiom of choice.

Example 4.41 $\mathbb{R}$ is not compact. Indeed, we have the open cover $\mathbb{R} = \bigcup_{n=1}^{\infty}(-n, n)$ which has no finite subcover, as $\mathbb{R} \not\subseteq \bigcup_{k=1}^{N}(-n_k, n_k) = (-K, K)$ for some $K > 0$. Similarly, $(0, 1)$ is not compact since

$$(0, 1) = \bigcup_{n=1}^{\infty}(0, 1 - 1/n)$$

has no finite subcover.

Example 4.42 Let X be any set with the discrete metric. Then $A \subseteq X$ is compact if and only if A is finite. If A is finite and $A \subseteq \bigcup_{i \in I} U_i$, then write $A = \{a_1, ..., a_n\}$, and for each $k \in \{1, ..., n\}$ there exists U_{i_k} such that $a_k \in U_{i_k}$. Then $A \subseteq \bigcup_{k=1}^{n} U_{i_k}$, which gives a finite subcover. If A is infinite, then $A = \bigcup_{a \in A}\{a\}$ is an open cover with no finite subcover since singletons are open in the discrete metric.

Example 4.43 I claim that open balls in normed linear spaces are not compact. Indeed,

$$B_r(0) = \{v \in V || |v|| < r\} = \bigcup_{0<s<r} B_s(0)$$

has no finite subcover since

$$\bigcup_{j=1}^{n} B_{s_j}(0) = B_s(0) \subsetneq B_r(0)$$

as $s = \max_{1 \le j \le n} s_j < r$.

Although not all closed and bounded sets are compact in metric spaces in general, we do have the converse:

Proposition 4.21 *Compact subsets of a metric space are closed and bounded.*

Proof Let $a \in \mathcal{A}$ with $\mathcal{A}$ compact in X. Then $\{B_n(a) | n \in \mathbb{N}\}$ is an open cover for $\mathcal{A}$. By compactness there exist $n_1, ..., n_k$ such that

$$\mathcal{A} \subseteq \bigcup_{j=1}^{k} B_{n_j}(a) = B_{\max_{1 \leq j \leq k} n_j}(a)$$

so $\mathcal{A}$ is bounded.

To show $\mathcal{A}$ is closed, suppose $x_0 \in \mathcal{A}'$. Assume $x_0 \notin \mathcal{A}$. Define $r_a := \frac{d(a,x_0)}{2}$ for each $a \in \mathcal{A}$, and note $r_a > 0$ as $x_0 \notin \mathcal{A}$. Then $\mathcal{A} \subseteq \bigcup_{a \in \mathcal{A}} B_{r_a}(a)$, so $\{B_{r_a}(a) | a \in \mathcal{A}\}$ is an open cover for $\mathcal{A}$. By compactness, there exist $a_1, ..., a_n \in \mathcal{A}$ with $\mathcal{A} \subseteq \bigcup_{i=1}^{n} B_{r_{a_i}}(a_i)$. Let $r = \min\{r_{a_1}, ..., r_{a_n}\} > 0$. Then $B_r(x_0) \cap B_{r_{a_i}}(a) = \emptyset$ for all $1 \leq i \leq n$. In particular, $B_r(x_0) \cap \mathcal{A} = \emptyset$. But $x_0 \in \mathcal{A}'$, a contradiction. □

We will now show compactness implies sequential compactness - this result will later be repeated with a similar proof.

Proposition 4.22 *Every infintie subset of a compact subset A of a metric space X has an accumulation point in A. In particular, if A contains a sequence of distinct points, then that sequence must contain a convergent subsequence. Consequently, compactness implies sequential compactness.*

Proof Suppose $C \subseteq A$ with C infinite; if A is finite every sequence has a convergent subseqeunce since every sequence must repeat some point an infinite number of times (pigeon hole principle). We want to show $C' \cap A \neq \emptyset$. Suppose $C' \cap A = \emptyset$. Then for all $a \in A$, there exists $r_a > 0$ with

$$(B_{r_a}(a) \backslash \{a\}) \cap C = \emptyset$$

The collection $\{B_{r_a}(a) : a \in A\}$ is an open cover for A. By compactness there exist $a_1, ..., a_n$ with $C \subseteq A \subseteq \bigcup_{i=1}^{n} B_{r_{a_i}}(a_i)$. But then C intersects only the a_i, at most, so $C \subseteq \{a_1, ..., a_n\}$, contradicting the assumption that C is infinite. □

Proposition 4.23 *Suppose A is a compact subset in a metric space X and C is a closed subset of A (or X). Then C is compact in A (in X).*

Proof Let $\{U_i\}$ be an open cover for C. C is closed in A so $A \backslash C$ is open in A. But then

$$A \subseteq \left(\bigcup_{i \in I} U_i \right) \cup (A \backslash C)$$

so $\{U_i\} \cup \{A \backslash C\}$ is an open cover of A. But A is compact, so there exist $i_1, ..., i_n$ such that

$$A \subseteq U_{i_1} \cup ... \cup U_{i_n} \cup (A \backslash C)$$

and $C \subseteq U_{i_1} \cup ... \cup U_{i_n}$, so C has a finite subcover and hence is compact. □

Example 4.44 As we show in our function section, if $f : X \to Y$ is continuous and if X is compact, then $f(X)$ is compact in Y. Indeed, suppose $\{U_i\}_{i \in I}$ is an open cover for $f(X)$, so $f(X) \subseteq \bigcup_{i \in I} U_i$. Then

$$X \subseteq f^{-1}\left(\bigcup_{i\in I} U_i\right) = \bigcup_{i\in I} f^{-1}(U_i)$$

which is an open cover as f is continuous. So $X \subseteq f^{-1}(U_{i_1}) \cup \ldots \cup f^{-1}(U_{i_n})$ for some finite subcover, which implies

$$f(x) \subseteq U_{i_1} \cup \ldots \cup U_{i_n}$$

Example 4.45 As we shall show later, continuous invertible maps on compact spaces are homeomorphisms. Suppose $f : X \to Y$ is continuous and invertible with X compact. Then f^{-1} is continuous and hence f is a homeomorphism. To see this we will show $f(C) = (f^{-1})^{-1}(C)$ is closed whenever C is closed in X. But X is compact, so C is compact as it is closed. Then $f(C)$ is compact in Y by the previous result, so $f(C)$ is closed in Y, completing the claim.

Note that, as stated above, the property that every infinite subset of A has an accumulation point, often coined ***limit point compactness***, is equivalent to sequential compactness for A in metric spaces.

Definition 4.32 A subset A of a metric space X is ***totally bounded*** if for every $\varepsilon > 0$ there exists a finite set $\{x_1, ..., x_N\} \subseteq X$ such that

$$A \subseteq \cup_{j=1}^{N} B_\varepsilon(x_j)$$

Example 4.46 Balls may not be totally bounded. Indeed, $\overline{B}_1(0)$ in $C([0,1])$ is not totally bounded. For example, take $f_n(x) = 0$ for $0 \le x \le 1/2^n$, $f_n(x) = 2^{n+1}(x - 1/2^n)$ for $1/2^n \le x \le 1/2^n + 1/2^{n+1}$, $f_m(x) = -2^{n+1}(x - 1/2^{n-1})$ for $1/2^n + 1/2^{n+1} \le x \le 1/2^{n-1}$, and $f_n(x) = 0$ for $1/2^{n-1} \le x \le 1$ (disjoint triangles of height 1)

Proposition 4.24 *If X is a sequentially compact metric space, then X is totally bounded.*

Proof Let $\varepsilon > 0$. If X is empty, there is nothing to prove, so suppose $X \neq \emptyset$. Let $x_1 \in X$. If $X = B_\varepsilon(x_1)$ we're done. Otherwise, choose $x_2 \in X \backslash B_\varepsilon(x_1)$. If $X = B_\varepsilon(x_1) \cup B_\varepsilon(x_2)$, we're done. Then either this process terminates at a finite step, in which case we're done, or we obtain a sequence $x_1, x_2, \ldots$ such that $d(x_j, x_i) \ge \varepsilon$ for all $i \ne j$. But then (x_j) has no convergent subsequence, contradicting the assumption that X is compact. Thus, this process must terminate at a finite step, so there exists $N \in \mathbb{N}$ such that

$$X = \bigcup_{j=1}^{N} B_\varepsilon(x_j)$$

Corollary 4.4 *If X is a sequentially compact metric space, then X has a countable dense subset, which is to say X is separable.*

Proof By Proposition 4.24 X is totally bounded. Let $S_n = \{x_{n1}, \ldots, x_{nm_n}\}$ such that

$$X = \bigcup_{j=1}^{m_n} B_{2^{-n}}(x_{nj}),\ C := \bigcup_{n=1}^{\infty} S_n$$

Then C is countable being the countable union of finite sets. Let $x \in X$ and $\varepsilon > 0$. Let $n \in \mathbb{N}$ such that $2^{-n} < \varepsilon$. Then $X = \bigcup_{j=1}^{m_n} B_{2^{-n}}(x_{nj})$ so $x \in B_{2^{-n}}(x_{nj})$ for some $1 \leq j \leq m_n$. Thus $d(x, x_{nj}) < 2^{-n} < \varepsilon$, so $x_{nj} \in B_\varepsilon(x)$ and $B_\varepsilon(x) \cap C \neq \emptyset$. Thus C is dense in X. □

This implies a relation between sequentially compact metric spaces and its size. In particular, metric plus compact implies separable.

Proposition 4.25 *If X is a sequentially compact metric space and $K_1 \supseteq K_2 \supseteq \ldots$ is a chain of non-empty closed subsets, then $\bigcap_{j=1}^{\infty} K_j \neq \emptyset$.*

The proof of this result is identical to that of the case for Euclidean Spaces, and the same holds for its corollary:

Corollary 4.5 *If X is a sequentially compact metric space and $U_1 \subseteq U_2 \subseteq \ldots$ is a chain of open sets such that $X = \bigcup_{j=1}^{\infty} U_j$, then there exists $M \in \mathbb{N}$ such that $X = U_M$.*

Proposition 4.26 *If X is a sequentially compact metric space, then X is topologically compact.*

Proof Let $\{U_\alpha\}_{\alpha \in J}$ be an open cover of X. By Proposition **??** X is separable, so we have a countable dense subset C. Let $\mathcal{R} = \{B_q(x) : x \in C, q \in \mathbb{Q}^+\}$. Then $\mathcal{R}$ is countable. Further, if $U \subseteq X$ is open, for all $p \in U$ there exists $\varepsilon > 0$ such that $B_\varepsilon(p) \subseteq U$. Then as C is dense, $C \cap B_{\varepsilon/3}(p) \neq \emptyset$, so there exists $c \in C$ such that $c \in B_{\varepsilon/3}(p)$. Then as $\mathbb{Q}$ is dense in $\mathbb{R}$ there exists $q \in \mathbb{Q}$ such that $\varepsilon/3 < q < 2\varepsilon/3$. It follows that $p \in B_q(c) \subseteq B_\varepsilon(p) \subseteq U$. Thus $U = \bigcup\{B \in \mathcal{R} : B \subseteq U\}$, so every open set can be written as a countable union of open sets in $\mathcal{R}$. Then $X = \bigcup_{\alpha \in J} U_\alpha = U_{\alpha \in J} \bigcup_{j \geq 1} B_{q_{\alpha,j}}(c_{\alpha,j})$. Suppose $\{B_1, B_2, \ldots\}$ is a countable cover of X. Define $U_m = B_1 \cup \ldots \cup B_m$. Then by Corollary 4.5 there exists $M \in \mathbb{N}$ such that $X = \bigcup_{j=1}^{M} B_j$. Thus as $\bigcup_{\alpha \in J} \bigcup_{j \geq 1} B_{q_{\alpha,j}}(c_{\alpha,j})$ is countable, there exists $M \in \mathbb{N}$ such that $X = B_{q_{\alpha_1,i_1}}(c_{\alpha_1,i_1}) \cup \ldots \cup B_{q_{\alpha_M,i_M}}(c_{\alpha_M,i_M})$, so $X = U_{\alpha_1} \cup \ldots \cup U_{\alpha_M}$ is a finite subcover. □

Theorem 4.7 *If X is a metric space, then X is sequentially compact if and only if it is compact in terms of open covers.*

Proof Proposition 4.26 is the forward implication, so suppose X is topologically compact. We show the equivalence with limit point compactness. We argue by contrapositive and suppose $S \subseteq X$ has no accumulation points. Then in particular S is closed, as $\overline{S} = S \cup S' = S$, since $S' = \emptyset$. Then $S_x = S \backslash \{x\}$ for all $x \in S$ is also closed, also having no accumulation points. Then $U_x \in X \backslash S_x$ is a cover for S, with $U_x \cap S = \{x\}$. But then $\{U_x\}_{x \in S} \cup \{X \backslash S\}$ is an open cover for X. As X is open cover compact, we have $x_1, \ldots, x_n$ such that $U_{x_1}, \ldots, U_{x_n}, X \backslash S$ covers X. But then $U_{x_1}, \ldots, U_{x_n}$ covers S, so $S = \bigcup_{i=1}^{n} S \cap U_{x_i} = \{x_1, \ldots, x_n\}$ so S is finite. Thus, X is limit point compact, completing the proof. □

Note that a metric space which is compact is totally bounded and complete. Now we have the converse:

Proposition 4.27 *If X is a complete metric space which is totally bounded, then X is compact.*

Proof Let $S \subseteq X$ be infinite. Because X is totally bounded, there exist $x_1, ..., x_N \in X$ such that $X \subseteq \bigcup_{j=1}^{N} B_{1/2}(x_j)$. Since $S \subseteq X$ is infinite, by the pigeon hole principle there exists $x_j =: x^1$ such that $B_{1/2}(x^1) \cap S$ is infinite. Then there exists $\{x_{2,1}, ..., x_{2,N_2}\} \subseteq X$ such that $X \subseteq \bigcup_{j=1}^{N_2} B_{1/2^2}(x_{2,j})$ and again there exists $x_{2j} =: x^2$ such that $B_{1/2^2}(x^2) \cap (B_{1/2}(x^1) \cap S)$ is infinite. Continuing in this way there exists $x^j \in X$ such that

$$B_{1/2^j}(x^j) \cap ... \cap B_{1/2}(x^1) \cap S$$

is infinite. Let X_j be the closure of this jth set, so we have a decreasing chain $X_1 \supseteq X_2 \supseteq ...$ of non-empty closed sets, such that $X_j \cap S$ is infinite for all j. Pick $z_1 \in X_1 \cap S$, and $z_{j+1} \in X_{j+1} \cap S \backslash \{z_1, ..., z_j\}$, which is possible using the axiom of choice as each set is infinite. Then $(z_j) \subseteq X_1 \cap S$ is a Cauchy sequence. By completeness there exists $z \in X$ such that z_j converges to z. But $(z_j) \subseteq S$, and the z_j are distinct, so $z \in X$ is an accumulation point of S. □

We prove this result again, with its converse, as follows

Theorem 4.8 (General Heine-Borel Theorem) *A non-empty subset A of a metric space X is compact in X if and only if A is complete and totally bounded.*

Proof Suppose A is compact and $r > 0$. Then $\{B_r(x) | x \in A\}$ is an open cover for A. So there exist $x_1, ..., x_n \in X$ with $A \subseteq B_r(x_1) \cup ... \cup B_r(x_n)$. So A is totally bounded. If $\{x_n\} \subseteq A$ is Cauchy, then there exists a subsequence $x_{n_k} \longrightarrow x \in A$ by sequential compactness. Then as the sequence is Cauchy and a subsequence converges, $x_n \longrightarrow x$. So A is complete.

Conversely, suppose A is complete and totally bounded. If A is finite it is compact and we're done. So assume A is infinite. Let $\{x_n\}$ be any sequence of distinct points in A. Cover A by finitely many balls of radius 1 by total boundedness. At least one of these, say $B_1(y_1)$, contains infinitely many x_i. $A \cap B_1(y_1)$ is also totally bounded, and hence can be covered with finitely many balls of radius 1/2. At least one of these, say $B_{1/2}(y_2)$, contains infinitely many of the x_i. Thus, for each k, there exists $x_{n_k} \in B_{1/2^k}(y_k)$. These can be chosen such that $\{x_{n_k}\}_{k \geq l} \subseteq B_{1/2^l}(y_l)$. This implies $d(x_{n_k}, x_{n_m}) < 1/2^{l-1}$ for $k, m \geq l$. Thus $\{x_{n_k}\}$ is Cauchy. By completeness, there exists $x \in A$ with $x_{n_k} \longrightarrow x$. Thus A is sequentially compact, and the result follows. □

Proposition 4.28 *If X is a compact metric space, then diam*$(X) < \infty$*, where*

$$diam(X) = \sup\{d(x, y) : x, y \in X\}$$

Proof As X is compact there exist $x_1, ..., x_N \in X$ such that $X \subseteq \bigcup_{j=1}^{N} B_1(x_j)$. Then, let $M = \max\{d(x_i, x_j) : 1 \leq i, j \leq N\}$. Now, let $x, y \in X$. Then there exist i, j such that $x \in B_1(x_i)$ and $y \in B_1(x_j)$. It follows that

$$d(x, y) \leq d(x, x_i) + d(x_i, x_j) + d(x_j, y) < 1 + M + 1 = M + 2$$

Thus, we have that $\text{diam}(X) \leq M + 2 < \infty$, as desired. □

We now prove this result ones more using a slightly different technique, and slightly different results.

Definition 4.33 Let X be a metric space and let $\{U_n\}_{n\in\mathbb{N}}$ be a countable collection of open sets in X. If for any open set V of X, and any $x \in V$, there exists i such that $x \in U_i \subseteq V$, then $\{U_n\}_{n\in\mathbb{N}}$ is a countable bases for the topology on X, and is called a ***countable base for A***.

Lemma 4.4 *Suppose A is a non-empty subset of a metric space X with the property that every infinite subset of A has an accumulation point (i.e. that $\overline{A}$ is sequentially compact). Then there are open sets $\{U_n\}_{n\in\mathbb{N}}$ of X with the property that if V is any open set in X and $x \in A \cap V$, then there is a U_i with $x \in U_i \subseteq V$.*

That is A has a countable base.

Proof We claim for each $n \in \mathbb{N}$ there is a finite set $\{x_1, ..., x_{N(n)}\} \subseteq X$ such that

$$A \subseteq \bigcup_{j=1}^{N(n)} B_{1/n}(x_j)$$

We show this by contradiction, assuming the inclusion does not hold. Thus, there is no finite collection of balls of radius $1/n$ covering A, for some n. Let $y_1 \in A$ with $A \not\subseteq B_{1/n}(y_1)$, $y_2 \in A \backslash B_{1/n}(y_1)$ with $d(y_2, y_1) \geq 1/n$, and inductively take $y_{k+1} \in A \backslash \bigcup_{j=1}^{k} B_{1/n}(y_j)$ with $d(y_j, y_{k+1}) \geq 1/n$ for all $1 \leq j \leq k$. By construction $\{y_k\}$ consists of infinitely many points. But it has no accumulation points since $d(y_i, y_j) \geq 1/n$ for all i and j. This contradicts the assumption that any infinite set in $\overline{A}$ has an accumulation point. So we take the countable collection

$$\{B_{1/n}(x_i) | n \in \mathbb{N}, 1 \leq i \leq N(n)\}$$

This collection satisfies the conclusion of the theorem. ◻

Theorem 4.9 *A non-empty subset A of a metric space X is compact in X if and only if it is sequentially compact in X.*

Proof We already have compactness implies sequential compactness. Suppose A is sequentially compact. Suppose $A \subseteq \bigcup_{i\in I} U_i$ is an open cover. We first prove A has a countable subcover. Let $x \in A$ so $x \in U_i$ for some i. $\overline{A} = A$ satisfies the hypotheses of the result above. Let $\{V_n\}_{n\in\mathbb{N}}$ be the collection in the conclusion. There exists V_j with $x \in V_j \subseteq U_i$. Thus,

$$A \subseteq \bigcup_{j=1}^{\infty} V_j \subseteq \bigcup_{countable} U_i$$

taking the countable subcover of U_i's associated to the V_j's. We can write $A \subseteq \bigcup_{j=1}^{\infty} U_{i_j}$, and we want a finite subcover. Towards a contradiction suppose there is no finite subcollection covering A. Then there exists $x_1 \in A \backslash U_{i_1}$, $x_2 \in A \backslash U_{i_1} \cup U_{i_2}$, and in general $x_k \in A \backslash \bigcup_{j=1}^{k} U_{i_j}$. By assumption there exists a subsequence $x_{n_k} \longrightarrow x \in A$. The U_{i_j} cover A. So there exists $N \in \mathbb{N}$ such that $x \in U_{i_N}$. But U_{i_N} is open. So there exists M such that for $n_k \geq M$, $x_{n_k} \in U_{i_N}$. So U_{i_N} contains all but finitely many of the x_{n_k}. But by construction each U_{i_j} can only contain finitely many x_k, which is a contradiction. Thus, a finite subcover must exist. ◻

Theorem 4.10 (Heine-Borel) *A non-empty subset A of ℓ_n^p is compact if and only if A is closed and bounded.*

Proof We know compact implies closed and bounded. Suppose $A \subseteq \ell_n^p$ is closed and bounded. We show A is sequentially compact. Let $\mathbf{x}_k \in A$. By Bolzano-Weierstrass, since A is bounded, $\mathbf{x}_{n_k} \to \mathbf{x} \in \ell_n^p$ for some subsequence. Since A is closed, $\mathbf{x} \in A$. □

A reminder that Heine-Borel fails in other metric spaces, such as the example $f_n(x) = x^n$ in $C([0,1]) \cap \overline{B}_1(0)$, which has no (uniformly) convergent subsequence as discussed previously.

4.7 Product Spaces

We now define finite and countable products of metric spaces, and their associated properties.

Definition 4.34 If $(X_1, d_1), ..., (X_N, d_N)$ are metric spaces, we define the product metric space

$$X := X_1 \times \cdots \times X_N = \prod_{j=1}^{N} X_j$$

and define a metric d in X for $x = (x_1, ..., x_N), y = (y_1, ..., y_N)$ by

$$d(x, y) = \sum_{j=1}^{N} d_j(x_j, y_j)$$

Equivalently, we could define

$$\delta(x, y) = \sqrt{\sum_{j=1}^{N} d_j(x_j, y_j)^2}$$

where the equivalence is in the sense that they define the same topology on the product. In particular, we characterize two metrics being equivalent as follows:

Definition 4.35 If X is a set with metrics d_1 and d_2, then d_1 and d_2 are said to be ***equivalent*** if there exists $0 < C_0 \le C_1 < \infty$ such that

$$C_0 d_1(x, y) \le d_2(x, y) \le C_1 d_1(x, y)$$

for all $x, y \in X$.

Definition 4.36 For $p \in \mathbb{N}$, the ℓ_p norm on $\mathbb{R}^n$ is

$$||\mathbf{x}||_p = \left(\sum_{i=1}^{n} |x_i|^p\right)^{1/p}$$

which gives the metric

$$d_p(\mathbf{x}, \mathbf{y}) = ||\mathbf{x} - \mathbf{y}||_p$$

For $p = \infty$, define

$$||\mathbf{x}||_\infty = \max_{1 \le j \le n} |x_j|$$

and

$$d_\infty(\mathbf{x}, \mathbf{y}) = \max_{1 \le j \le n} |x_j - y_j|$$

It is an important, but non-trivial result, that d_p and d_q are equivalent for all $0 < p, q \le \infty$. Now we define countable products:

Definition 4.37 Let $(X_1, d_1), (X_2, d_2), ...$ be a countable collection of metric spaces. Define

$$X = \prod_{j=1}^{\infty} X_j$$

where $x \in X$ is a sequence $x = (x_j)$, $x_j \in X_j$. We define the metric by

$$d(x, y) = \sum_{j=1}^{\infty} 2^{-j} \frac{d_j(x_j, y_j)}{1 + d_j(x_j, y_j)}$$

We have that compactness of product spaces occurs if and only if we have compactness of the individual component spaces.

Proposition 4.29 *The product $X = \prod_{j=1}^{N} X_j$ is compact if and only if X_j is compact for all j.*

Proof First, suppose the product is compact and let $(x_{n,j})_{n=1}^{\infty} \subseteq X_j$. Let $x_i \in X_i$ for $i \neq j$. Then $(x_1, ..., x_{n,j}, ..., x_N) \subseteq X$ has a convergent subsequence $(x_1, ..., x_{n_k,j}, ..., x_N)$ converging to $(x_1, ..., x_j, ..., x_N)$ in X since X is compact. Then for all $\varepsilon > 0$, there exists $M \in \mathbb{N}$ such that for $k \ge M$, $d_j(x_{n_k,j}, x_j) = d((x_1, ..., x_{n_k,j}, ..., x_N), (x_1, ..., x_j, ..., x_N)) < \varepsilon$, so $(x_{n_k,j})$ is a convergent subsequence of (x_j). Thus X_j is compact for all j, as desired. Now suppose that X_j is compact for each j, and $((x_{n,1}, ..., x_{n,N})) \subseteq X$. As X_1 is compact there exists a subsequence $(x_{s_1(n),1})$ which converges to some $x_1 \in X_1$. Then $(x_{s_1(n),2}) \subseteq X_2$ has a convergent subsequence $(x_{s_2(n),2})$ since X_2 is compact, which converges to $x_2 \in X$. Proceeding in this way we arrive at $((x_{s_N(n),1}, ..., x_{s_N(n),N})) \subseteq ((x_{n,1}, ..., x_{n,N}))$, where $(x_{s_N(n),j}) \subseteq (x_{s_j(n),j})$ which converges to $x_j \in X_j$, and hence $(x_{s_N(n),1}, ..., x_{s_N(n),N})$ converges to $(x_1, ..., x_N)$. Thus X is compact, as desired. □

Proposition 4.30 *The product $X = \prod_{j=1}^{\infty} X_j$ is compact if and only if X_j is compact for all j.*

Proof The forward implication follows analogously to the proof of the previous proposition. Now, suppose each X_j is compact, and let $((x_{\nu,1}, x_{\nu,2}, ...)) \subseteq X$ be an arbitrary sequence. Then, as X_1 is compact, we have a subsequence $x_{s_1(\nu),1}$ which converges to some $x_1 \in X_1$. Then, we can take a subsequence $x_{s_2(\nu),2}$ of $x_{s_1(\nu),2}$ which converges to some $x_2 \in X_2$ since X_2 is compact. Let $(x_\nu^j) = ((x_{s_j(\nu),1}, x_{s_j(\nu),2}, ...))$. Then we have a decreasing sequence of subsequences $(x_\nu) \supseteq (x_\nu^1) \supseteq (x_\nu^2) \supseteq ...$, such that $x_{s_j(\nu),i}$ converges to $x_i \in X_i$ for all $1 \le j \le i$. Then, let (ξ_ν) be the subsequence defined by $\xi_\nu = x_\nu^\nu$, the diagonal. Then for all $j \in \mathbb{N}$, $(\xi_\nu)_{\nu=j}^{\infty} \subseteq (x_\nu^j)_{\nu=1}^{\infty}$, so for each j $x_{s_\nu(\nu),j}$ converges to $x_j \in X_j$. Thus, ξ_ν converges to $(x_1, x_2, ...) \in X$, and hence X is compact. □

4.8 Baire Category Theorem

We now construct a result on the size of complete metric spaces, known as the Baire's Category Theorem. First we define the notion of a category for a topological space:

Definition 4.38 For a topological space (X, τ), a subset A is said to be of the ***first category*** in X if and only if A can be written as a countable union of nowhere dense subsets of X.

Definition 4.39 A subset $S \subseteq X$ is ***nowhere dense*** if and only if $\overline{S}$ does not contain any non-empty open subsets, which occurs in a metric space if and only if $\overline{S}$ does not contain any open ball $B_r(x)$ for $r > 0$ and $x \in X$. Equivalently, S is nowhere dense if and only if $X\backslash\overline{A} = \overline{A}^c$ is dense in X.

Definition 4.40 For a topological space X, a subset A is said to be of the ***second category*** in X if and only if A is not of the first category.

Theorem 4.11 (Baire's Category Theorem) *If X is a complete metric space, then X is of second category.*

Proof Let $S_k \subseteq X$ be a sequence of nowhere dense sets. We claim $X\backslash \bigcup_{k\geq 1} S_k \neq \emptyset$. Let $T_k = \overline{\bigcup_{j=1}^k S_j} = \bigcup_{j=1}^k \overline{S}_j$, so T_k is closed and nowhere dense. Further, $T_1 \subseteq T_2 \subseteq ....$ Let $U_k = X\backslash T_k$, so U_k is open and dense. Indeed, if $x \in X$, and $N_x \subseteq N(x)$, an open neighborhood, $N_x \not\subseteq T_k$, so $N_x \cap U_k \neq \emptyset$. Thus $\overline{U_k} = X$, so U_k is dense. Then we have $U_1 \supseteq U_2 \supseteq ....$ We claim that the intersection is non-empty. Let $p_1 \in U_1$, so there exists $\varepsilon_1 > 0$ such that $\overline{B_\varepsilon(p_1)} \subseteq U_1$. By density of U_2, there exists $p_2 \in B_{\varepsilon/2} \cap U_2$. But U_2 is open so there exists $\varepsilon_2 < \varepsilon/2$ such that $\overline{B_{\varepsilon_2}(p_2)} \subseteq B_\varepsilon(p_1) \cap U_2$. Inductively, take $p_{k+1} \in B_{\varepsilon_k}(p_k) \cap U_{k+1}$, and $\varepsilon_{k+1} < \varepsilon_k/2$ such that $\overline{B_{\varepsilon_{k+1}}(p_{k+1})} \subseteq B_{\varepsilon_k}(p_k) \cap U_{k+1}$. Then $\varepsilon_k < \varepsilon/2^{k-1}$. Note (p_k) is a Cauchy sequence, because $p_l \in \overline{B_{\varepsilon_l}(p_l)} \subseteq B_{\varepsilon_k}(p_k)$ for all $l > k$, where $\varepsilon_k < \varepsilon/2^{k-1}$. By completeness of X, there exists $p \in X$ such that $p_l \to p$. Then $p \in \overline{B_{\varepsilon_k}(p_k)}$ for all k as they are closed, so in particular $p \in U_k$ for all k. Therefore, $p \in \bigcap_{i=1}^\infty U_i$, so $\bigcap_{i=1}^\infty U_i \neq \emptyset$, as claimed. □

We re-visit this investigation through a slightly different light:

Lemma 4.5 *A is nowhere dense in X if and only if $\overline{A}^\circ = \emptyset$.*

Proof First, if $B \subseteq X$, $X\backslash\overline{B} = (X\backslash B)^\circ$ and $X\backslash B^\circ = \overline{X\backslash B}$. Indeed, $X\backslash\overline{B} \subseteq X\backslash B$ and is open so $X\backslash\overline{B} \subseteq (X\backslash B)^{circ}$. Conversely, $(X\backslash B)^\circ \subseteq X\backslash B$, so $X\backslash(X\backslash B)^\circ \supseteq B$ is a closed set containing B, implying $X\backslash(X\backslash B)^\circ \supseteq \overline{B}$. It follows that $(X\backslash B)^\circ \subseteq X\backslash\overline{B}$ and we have equaltiy, $X\backslash\overline{B} = (X\backslash B)^\circ$. For the other equality, $X\backslash B^\circ \supseteq X\backslash B$ is a closed cover, so $X\backslash B^\circ \supseteq \overline{X\backslash B}$. Conversely, $X\backslash(\overline{X\backslash B}) \subseteq B$ is open, so $X\backslash(\overline{X\backslash B}) \subseteq B^\circ$, so $\overline{X\backslash B} \supseteq X\backslash B^{circ}$, and we have equality $X\backslash B^\circ = \overline{X\backslash B}$.

Then $\overline{A}^\circ = \emptyset$ if and only if $X\backslash\overline{A}^\circ = X$, if and only if $\overline{X\backslash\overline{A}} = X$, if and only if $X\backslash\overline{A}$ is dense in X, if and only if A is nowhere dense. □

Example 4.47 If $\{x_0\}$ in X is not isolated, then $\{x_0\}$ is nowhere dense, since $\overline{\{x_0\}} = \{x_0\}$ and it has no interior points. If X has no isolated points, then countable sets are of the fisrt category.

Example 4.48 $\mathbb{Q}$ is of the first category, being the countable union of singletons and having no isolated points. $\mathbb{R}\backslash\mathbb{Q}$ is of the second category, which follows from Baire's theorem.

Example 4.49 The middle thirds Cantor set C is nowhere dense. It is compact, and hence closed, and contains no intervals, and hence no interior.

Example 4.50 Collection of lines in $\mathbb{R}^2$: in ℓ_2^2, the line $\{(x, mx + b) | m, b \in \mathbb{R}\}$ is nowhere dense, since it is already closed and each point is a boundary point.

Theorem 4.12 (Baire) *Suppose X is a complete metric space and $\{U_n\}$ is a denumerable collection of open and dense subsets of X. Then*

$$\bigcap_{n\in\mathbb{N}} U_n$$

is dense in X.

Proof Note that $\overline{A} = X$ if and only if, for all $x \in X$, either $x \in A$ or $x \in A'$. If $x \in A$, then for every open neighborhood $U \subseteq X$ of x, $A \cap U \neq \emptyset$. If $x \in A'$, then for every $U \subseteq X$ open neighborhood of x, $A \cap (U\backslash\{x\}) \neq \emptyset$, so in particular $A \cap U \neq \emptyset$.

We must show that if $V \subseteq X$ is open and non-empty, then $\bigcap_{n=1}^{\infty} U_n \cap V \neq \emptyset$. Let $x_1 \in V \cap U_1$, which is possible since U_1 is dense. Then there exists $r_1 > 0$ with $\overline{B}_{r_1}(x_1) \subseteq V \cap U_1$, since $V \cap U_1$ is open. As U_2 is dense and $B_{r_1}(x_1)$ is open there exists $x_2 \in U_2 \cap B_{r_1}(x_1) \subseteq U_2 \cap U_1 \cap V$. Repeating this process, we can find $r_2 < r_1/2$ with

$$\overline{B}_{r_2}(x_2) \subseteq B_{r_1}(x_1) \cap U_2 \subseteq U_2 \cap U_1 \cap V$$

Inductively, find $x_n \in X$ and $r_n < r_{n-1}/2$ with

$$\overline{B}_{r_n}(x_n) \subseteq B_{r_{n-1}}(x_{n-1}) \cap U_n \subseteq \bigcap_{j=1}^{n} U_j \cap V$$

By construction, $(x_k)_{k\geq n} \subseteq \overline{B}_{r_n}(x_n)$. Since $r_n \to 0$, (x_k) is Cauchy and therefore has a limit x_0 since X is complete. Moreover, $\overline{B}_{r_n}(x_n)$ contains x_0 since it is closed. So by construction

$$x_0 \in \bigcap_{j=1}^{\infty} B_{r_j}(x_j) \subseteq \left(\bigcap_{j=1}^{\infty} U_j\right) \cap V$$

Corollary 4.6 (The Baire Category Theorem) *Complete metric spaces are of the second category.*

Proof Suppose the proposition does not hold. Let $X = \bigcup_{n=1}^{\infty} A_n$ with A_n nowhere dense (i.e. $X\backslash\overline{A_n}$ is dense in X, and open). By Baire's theorem $\bigcap_{n=1}^{\infty} X\backslash\overline{A_n}$ is dense in X, which holds if and only if $X\backslash\bigcup_{n=1}^{\infty}\overline{A_n}$ is dense in X. But, $\bigcup_{n=1}^{\infty}\overline{A_n} = X$, so this holds if and only if $\emptyset$ is dense in X. □

Example 4.51 The irrationals: $\mathbb{Q}$ is of the first category, so $\mathbb{R}\backslash\mathbb{Q}$ must be of the second, as otherwise $\mathbb{R} = \mathbb{Q} \cup (\mathbb{R}\backslash\mathbb{Q})$ would be of the first, which it is not being a complete metric space.

Example 4.52 Any basis of an infinite dimensional complete normed linear space is necessarily uncountable. However, there are infinite dimensional incomplete space having a countable basis: for example, the normed linear space of polynomials with norm $||a_0 + a_1 x + ... + a_n x^n|| = \sum_{i=1}^{n} |a_i|$ has countable basis $\{1, x, x^2, x^3, ...\}$ since each polynomial can be written as a unique finite linear combination of these elements.

Now, suppose to the contrary that $\{v_1, v_2, ...\}$ is a vector space basis for such a NLS V. For each n, let $V_n = \text{span}\{v_1, ..., v_n\}$. Each V_n is nowhere dense. Indeed, each V_n is closed, as to be shown below, and if U is an open ball in V_n, by subtracting the center of this ball we are still in V_n. So we may assume $B_r(0) \subseteq V_n$. So if $v \in V\backslash\{0\}$, we have $\frac{v}{2||v||} r \in V_n$. But V_n is a subspace, so $v \in V_n$. This implies $V = V_n$, contradicting the fact that V is finite dimensional. Thus V_n is nowhere dense.

Notice by assumption $V = \bigcup_{n=1}^{\infty} V_n$. By Baire's theorem, this contradicts the completeness of V.

Theorem 4.13 (Uniform Boundedness Principal) *Let X be a complete metric space and $S \subseteq C(X)$. Suppose that for each $x \in X$ there exists a positive constant $M_x > 0$ such that $|f(x)| \le M_x$ for all $f \in S$ (pointwise bound). Then there is a non-empty open subset U of X and $M > 0$ such that $|f(x)| \le M$ for all f in S and all x in U (uniform bound).*

If $\sup_X M_x < \infty$, we can just use this as M and $U = X$

Proof Fix $n \in \mathbb{N}$ and $f \in S$. Define

$$S_{n,f} = \{x \in X | |f(x)| \le n\} = f^{-1}([-n, n])$$

$S_{n,f}$ is closed being the pre-image of a closed set under a continuous map. Define $S_n := \bigcap_{f \in S} S_{n,f}$, which is again closed. Note $S_n \subseteq S_{n+1}$. Note $x \in S_n$ if and only if $|f(x)| \le n$ for all $f \in S$. By assumption, $X = \bigcup_{n=1}^{\infty} S_n$, which is a countable union of closed sets. By Baire's theorem, at least one of the S_N must contain an open set $U \subseteq S_N \subseteq X$. This means for all $x \in U$, we have $|f(x)| \le N$ for all $f \in S$, as desired. □

Example 4.53 Let $X = [0, \infty)$ with the usual metric, and let $S = \{f_n(x) = \sqrt[n]{(x)}\} \subseteq C(X)$. For fixed $x \ge 0$, $|f_n(x)| = \sqrt[n]{x} \le \max\{1, x\} =: M_x$. Note $\sup_x M_x = \infty$. Observe that $U = [0, a)$ is open in $[0, \infty)$ for any $a > 0$. In this case, $|f_n(x)| \le \max\{1, a\}$, for all $x \in U$ and all n.

Using this result we can explore certain examples of Banach spaces.

Definition 4.41 (Banach Space) A complete normed linear space is called a ***Banach space***.

Example 4.54 Note we have already shown that ℓ_n^p is a Banach space for $p \in [1, \infty]$ and $n \in \mathbb{N}$.

Example 4.55 We have also shown that $C_b(S)$ is a Banach space for $S \subseteq \mathbb{R}$.

Example 4.56 The classical Banach sequence spaces, ℓ^p, for $1 \leq p < \infty$: If $1 \leq p < \infty$,

$$\ell^p = \left\{ x = (x_k) \text{ sequence of real or complexes} | ||x||_p := \left(\sum_{k=1}^{\infty} |x_k|^p \right)^{1/p} < \infty \right\}$$

and for $p = \infty$:

$$\ell^\infty = \{x = (x_k) | | ||x||_\infty \sup_k |x_k| < \infty\}$$

We claim that $|| \cdot ||_p$ are all norms for $1 \leq p \leq \infty$. The first two axioms, positive definiteness and absolute homegeneity, are similar to the finite case. For the triangle inequality for $|| \cdot ||_\infty$, $|x_i + y_i| \leq |x_i| + |y_i| \leq ||x||_\infty + ||y||_\infty$ for all i, so $||x + y||_\infty \leq ||x||_\infty + ||y||_\infty$. For $p = 1$,

$$\sum_{i=1}^{n} |x_i + y_i| \leq \sum_{i=1}^{n} |x_i| + \sum_{i=1}^{n} |y_i| \leq ||x||_1 + ||y||_1$$

for all n, so $||x + y||_1 \leq ||x||_1 + ||y||_1$.

For $1 < p < \infty$, we have by Hölder's inequality that

$$\sum_{k=1}^{n} |x_k||y_k| \leq ||x||_p ||y||_q$$

for $x = [x_1 \ldots x_n]$, $y = [y_1 \ldots y_n]$ and $\frac{1}{p} + \frac{1}{q} = 1$. Now, let $x \in \ell^p$ and $y \in \ell^q$ such that $\frac{1}{p} + \frac{1}{q} = 1$. Then

$$\begin{aligned} \sum_{k=1}^{n} |x_k||y_k| &\leq \left(\sum_{k=1}^{n} |x_k|^p \right)^{1/p} \left(\sum_{k=1}^{n} |y_k|^q \right)^{1/q} \\ &\leq \left(\sum_{k=1}^{\infty} |x_k|^p \right)^{1/p} \left(\sum_{k=1}^{\infty} |y_k|^q \right)^{1/q} = ||x||_p ||y||_q \end{aligned}$$

This holds for all n, so the sum on the left converges as $n \longrightarrow \infty$ and

$$\sum_{k=1}^{\infty} |x_k||y_k| \leq ||x||_p ||y||_q$$

The triangle inequality now follows from Hölder's inequality, using the same argument as the finite dimensional case.

We claim ℓ^p is complete. For $1 \leq p < \infty$, let (x_m) be a Cauchy sequence in ℓ^p. Write $x_m = (x_{m,1}, x_{m,2}, \ldots)$. For each i,

$$|x_{m,i} - x_{l,i}|^p \leq \sum_{j=1}^{\infty} |x_{m,j} - x_{l,j}|^p = ||x_m - x_l||_p^p$$

which implies $(x_{m_i})_{m=1}^{\infty}$ is Cauchy for each i, and hence has a limit, say $\hat{x}_i$ since $\mathbb{R}$ is complete. Let $x = (\hat{x}_1, \hat{x}_2, \ldots)$. Fix $\varepsilon > 0$ and any $n \in \mathbb{N}$. Find $m, l \geq N$ such that

$$\sum_{j=1}^{m} |x_{m,k} - x_{l,j}|^p \leq ||x_m - x_l||_p^p < \varepsilon^p$$

The left hand side is a finite sum so we can let $l \to \infty$, giving

$$\sum_{j=1}^{n} |x_{m,j} - \hat{x}_j|^p \leq \varepsilon^p$$

But this is true for all n. So taking the limit on n, $||x_m - x||_p^p < \varepsilon^p$, so $x_m \to x$ in ℓ^p. Now, $x \in \ell^p$ since

$$\sum_{j=1}^{n} |\hat{x}_j|^p \leq ||x - x_m||_p + ||x_m||_p$$

for all n, and taking the sup over all n we have our result.

Recall that if V is a finite dimensional vector space over $\mathbb{R}$ with basis $\{v_1, ..., v_n\}$, there is a vector space isomorphism, the coordinate isomorphism for the basis, sending $e_i \in \mathbb{R}^n$ to v_i in V. Any norm on $\mathbb{R}^n$ imparts a norm onto V with $||r_1v_1 + ... + r_nv_n|| := ||[r_1 \; ... \; r_n]||$.

Theorem 4.14 *Any two norms on $\mathbb{R}^n$ are equivalent.*

Proof Let $\{e_1, ..., e_n\}$ be the standard orthonormal basis for $\mathbb{R}^n$. Each $v \in \mathbb{R}^n$ can be written uniquely as $v = r_1e_1 + ... + r_ne_n$. Let $|| \cdot ||$ be any norm on $\mathbb{R}^n$. Then

$$||v|| \leq \sum_{i=1}^{n} |r_i|||e_i|| \leq C \sum_{i=1}^{n} |r_o| = C||v||_1$$

where $C = \max_{1 \leq i \leq n}\{||e_i||\}$. Let $S = \{v \in \mathbb{R}^n | ||v||_1 = 1\}$. S is compact by Heine-Borel. Define $\alpha : S \to [0, \infty)$ by $\alpha(v) = ||v||$. α is continuous since $|||v|| - ||w||| \leq ||v - w||$. This implies $\alpha(S)$ is compact in $[0, \infty)$. Thus, α achieves its minimum at some $v_0 \in S$. That is, $\alpha(v_0) \leq \alpha(v)$ for all $v \in S$. Let $d = \alpha(v_0)$. If $v \neq 0$, we have $\frac{v}{||v||_1} \in S$, so

$$\left\|\frac{v}{||v||_1}\right\| = \alpha \frac{v}{||v||_1} \geq d$$

so $||v|| \geq d||v||_1$. Thus $|| \cdot ||$ and $|| \cdot ||_1$ are equivalent. □

Corollary 4.7 *Finite dimensional vector spaces over $\mathbb{R}$ (or $\mathbb{C}$) are Banach spaces.*

Proof From the theorem above, $\mathbb{R}^n$ is complete with any norm. Any finite dimensional vector space V will then be complete by the induced norm on it. □

4.9 Continuous Functions on Metric Spaces

We now explore the homomorphisms of the category of metric spaces: continuous functions.

Definition 4.42 Suppose (X, d_X) and (Y, d_Y) are metric spaces and $f : X \to Y$ is a function. Given $x_0 \in X' \cap X$ we say that the ***limit of f as x tends to x_0 is L in Y*** and write

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

if for every $\varepsilon > 0$, there is a $\delta > 0$ such that $d_Y(f(x), L) < \varepsilon$ whenever $0 < d(x, x_0) < \delta$.

Definition 4.43 A function $f : X \to Y$ for metric spaces $(X, d_X), (Y, d_Y)$, is ***continuous at*** $x \in X$ if whenever $x_j \to x$ in X, then $f(x_j) \to f(x)$ in Y. We say f is continuous on X if and only if f is continuous at $x \in X$ for all x.

Equivalently, ***f is continuous at*** $x_0 \in X$ if for every $\varepsilon > 0$, there is a $\delta > 0$ such that $d_Y(f(x), f(x_0)) < \varepsilon$ whenever $d_X(x, x_0) < \delta$. Note we do not require x_0 to be an accumulation point for continuity. We say that f is ***continuous on X*** if f is continuous at each $x_0 \in X$.

If x_0 is an accumulation point, continuity at x_0 is equivalent to $\lim_{x \to x_0} f(x) = f(x_0)$. If x_0 is an isolated point, continuity is automatic.

Example 4.57 If X is discrete and $f : X \to Y$ then f is continuous. Indeed, every singleton in X is open so every point is isolated. If $\varepsilon > 0$, choosing $\delta = 1$ we have $d_X(x, x_0) < 1$ implies $x = x_0$ and so $d_Y(f(x), f(x_0)) = 0 < \varepsilon$.

Proposition 4.31 *Suppose (X, d_X) and (Y, d_Y) are metric spaces and $f : X \to Y$. Then $\lim_{x \to x_0} f(x) = L$ if and only if whenever x_n is a sequence in $X \backslash \{x_0\}$ converging to x_0, then $f(x_n)$ converges to L in Y.*

Proof First assume $\lim_{x \to x_0} f(x) = L$. Fix $\varepsilon > 0$ and find $\delta > 0$ such that $0 < d_X(x, x_0) < \delta$ implies $d_Y(f(x), L) < \varepsilon$. If $(x_n) \subseteq X \backslash \{x_0\}$ with $x_n \to x_0$, find $N \in \mathbb{N}$ such that $n \geq N$ implies $0 < d_X(x_n, x_0) < \delta$. Then for $n \geq N$ we have $d_Y(f(x_n), L) < \varepsilon$ so $f(x_n) \to L$.

Conversely, suppose $\lim_{x \to x_0} f(x_0)$ is not L. Then there exists $\varepsilon > 0$ such that for all $\delta > 0$, there exists $x_\delta \in X \backslash \{x_0\}$ with $0 < d_X(x_\delta, x_0) < \delta$ but $d_Y(f(x_\delta), L) \geq \varepsilon$. For $\delta = 1/n$, $n \in \mathbb{N}$, write $x_n = x_{1/n}$. Then $0 < d(x_n, x_0) < 1/n$ so $x_n \to x_0$. But $d_Y(f(x_n), L) \geq \varepsilon$ for all n. This implies $f(x_n)$ does not converge to L, as desired. □

We now show that our two characterizations of continuity are equivalent:

Proposition 4.32 *Suppose (X, d_X) and (Y, d_Y) are metric spaces and $f : X \to Y$. Then f is continuous at x_0 if and only if $f(x_n)$ converges to $f(x_0)$ in Y whenever x_n converges to x_0 in X.*

Proof If $x_0 \in X'$, use $L = f(x_0)$ in the proposition above. If x_0 is isolated, then f is automatically continuous at x_0, and any sequence which converges to x_0 must be eventually constant and hence the images sequence becomes $f(x_0)$, which trivially converges to $f(x_0)$. □

Proposition 4.33 *A function $f : X \to Y$ is continuous if and only if $U \subseteq Y$ open implies $f^{-1}(U) \subseteq X$ is open, where $f^{-1}(U) = \{x \in X : f(x) \in U\}$.*

Proof Suppose $f : X \to Y$ is continuous and $U \subseteq Y$ is open. Let $x_0 \in f^{-1}(U)$, which is to say $f(x_0) \in U$. As U is open in Y there exists $\varepsilon > 0$ such that $B_\varepsilon(f(x_0)) \subseteq U$. By continuity, there exists $\delta > 0$ such that $x \in B_\delta(x_0)$ then $f(x) \in B_\varepsilon(f(x_0))$. This implies

$$f(B_\delta(x_0)) \subseteq B_\varepsilon(f(x_0)) \implies B_\delta(x_0) \subseteq f^{-1}(B_\varepsilon(f(x_0))) \subseteq f^{-1}(U)$$

so $f^{-1}(U)$ is open. The converse is the same argument reversed. □

Corollary 4.8 *Suppose X and Y are metric spaces and $f : X \to Y$ a function. f is continuous on X if and only if $f^{-1}(C)$ is closed in X whenever C is closed in Y.*

Follows from the previous result and $X \backslash f^{-1}(C) = f^{-1}(Y \backslash C)$.

Example 4.58 The function $f : \ell_2^2 \to \mathbb{R}$ given by

$$f(x, y) = \begin{cases} \frac{x^2 y}{x^4 + y^2} & (x, y) \neq (0, 0) \\ 0 & (x, y) = (0, 0) \end{cases}$$

is discontinuous at $(0, 0)$. Consider the sequence $(x_n, y_n) = (1/\sqrt{n}, 1/n) \to (0, 0)$. Then

$$f(1/\sqrt{n}, 1/n) = \frac{1/n^2}{1/n^2 + 1/n^2} = \frac{1}{2} \not\to 0 = f(0, 0)$$

so the function is not continuous at $(0, 0)$.

Example 4.59 The function $f : \ell_2^2 \to \mathbb{R}$ given by

$$f(x, y) = \begin{cases} \frac{x^2 y}{\sqrt{x^2 + y^2}} & (x, y) \neq (0, 0) \\ 0 & (x, y) = (0, 0) \end{cases}$$

is continuous at $(0, 0)$. Indeed,

$$\begin{aligned} |f(x, y) - f(0, 0)| &= \frac{x^2|y|}{\sqrt{x^2 + y^2}} \\ &\leq \frac{(x^2 + y^2)|y|}{\sqrt{x^2 + y^2}} \\ &= \sqrt{x^2 + y^2}|y| \\ &\leq \sqrt{x^2 + y^2}\sqrt{x^2 + y^2} \\ &\leq ||(x, y)||_2^2 \end{aligned}$$

Fix $\varepsilon > 0$ and choose $\delta = \sqrt{\varepsilon}$. If $||(x, y)||_2 < \delta$, then $|f(x, y) - f(0, 0)| < \delta^2 = \varepsilon$, as desired.

Example 4.60 The evaluation and integration maps on $C([0, 1])$ are continuous. For any $x \in [0, 1]$, define $E_x : C([0, 1]) \to \mathbb{R}$ by $E_x(f) = f(x)$. Then E_x is continuous. Suppose $(f_n) \subseteq C([0, 1])$ and

$f_n \to_u f$, noting that $||\cdot||_\infty$ is the standard norm on $C([0,1])$. Then $E_x(f_n) = f_n(x)$, which converges to $f(x) = E_x(f)$ since uniform convergence implies pointwise convergence. So E_x is continuous at f for any $f \in C([0,1])$. Define the integral map $I : C([0,1]) \to \mathbb{R}$ by $I(f) = \int_0^1 f(x)dx$. Suppose $f_n \to_u f$ in $C([0,1])$. Then, since the convergence is uniform we have that

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

4.9.1 Operators and NLS

Definition 4.44 (Bounded Operators) Suppose V and W are normed linear spaces and $T : V \to W$ is a linear transformation. We say that T is a ***bounded operator*** (or bounded linear operator, or bounded linear transformation) if T is continuous on V.

? Question

Why the word *bounded*?

We will see later that the bounded linear maps are precisely the ones that map bounded sets to bounded sets. They are not “bounded" in the usual sense of the word. For example, $\mathrm{Id} : \mathbb{R} \to \mathbb{R}, \mathrm{Id}(x) = x$, is continuous, linear, but “unbounded" in the sense that $\sup_{\mathbb{R}} |f| = \infty$.

Proposition 4.34 *Suppose V and W are normed linear spaces and $T : V \to W$ is linear. Then T is bounded if and only if T is continuous at* 0.

Proof Fix $\varepsilon > 0$. Suppose T is continuous at 0. Then there exists $\delta > 0$ such that $||v||_V < \delta$ implies $||Tv||_W < \varepsilon$. If $u \in V$ and $||u - u_0||_V < \delta$, then we get

$$||T(u - u_0)||_W < \varepsilon$$

by continuity at 0. But, by linearity $T(u - u_0) = Tu - Tu_0$, so $||Tu - Tu_0||_W < \varepsilon$, so T is continuous at u_0.

The implication follows by definition of a bounded operator. □

Example 4.61 The identity map between the same space with different norms may not be bounded. Let $V = \{C([0,1]), ||\cdot||_\infty\}$, and $W = \left\{C([0,1]), ||\cdot||_1 = \int_0^1 |\cdot|dx\right\}$, and consider $T : V \to W$ with $Tf = f$. T is bounded. Indeed, assume $f_n \to_u 0$. Then $Tf_n = f_n$, and we need to show $||f_n||_1 \to 0$. But, by uniform convergence of the f_n to 0,

$$\int_0^1 |f_n(x)|dx \to \int_0^1 0dx = 0$$

so indeed T is continuous at 0, and by the last result is consequently bounded.

However, $T^{-1} : W \to V$ is not bounded. We will find a sequence $f_n \to 0$ in W, but $T^{-1}f_n = f_n \not\to 0$ in V. Define $f_n(x)$ as $f_n(x) = -n(x - 1/n)$ for $0 \le x \le 1/n$ and $f_n(x) = 0$ for $1/n \le x \le 1$. Then $||f_n||_1 = \frac{1}{2n} \to 0$, but $||f_n||_\infty = 1 \not\to 0$.

Example 4.62 The anti-differentiation map: Define $A : C([0,1]) \to C([0,1])$, both with uniform norm, by

$$Af(x) = \int_0^x f(t)dt$$

By the FTOC we have that $Af(x) \in C([0,1])$, and $(Af)' = f$. A is linear because the Riemann integral is linear. To show boundedness, suppose $f_n \to_u 0$. Then

$$|Af_n(x)| = \left|\int_0^x f_n(t)dt\right| \le \int_0^x |f_n(t)|dt \le x||f_n||_\infty \le ||f_n||_\infty$$

Hence, it follows that $||Af_n||_\infty \le ||f_n||_\infty \to 0$, so $Af_n \to_u 0$, and A is bounded.

Example 4.63 The differentiation map on smooth functions: Let $C^1([0,1])$, which is a normed linear space with norm $|||f||| := ||f||_\infty + ||f'||_\infty$. Define $D : C^1([0,1]) \to C([0,1])$, both with the uniform norm, where $Df = f'$. Although D is linear, it is not continuous at 0. Consider

$$f_n(x) = \frac{\sin(nx)}{\sqrt{n}}, \ ||f_n||_\infty \le \frac{1}{\sqrt{n}} \to 0$$

But $f_n'(x) = \sqrt{n}\cos(nx)$ does not even converge pointwise. However, $D : (C^1([0,1]), ||| \cdot |||) \to (C([0,1]), || \cdot ||_\infty)$ is bounded. Indeed,

$$||Df||_\infty = ||f'||_\infty \le |||f|||$$

Then if $|||f_n||| \to 0$, then $||Df_n||_\infty \le |||f_n||| \to 0$ and we have continuity at 0.

Example 4.64 Maps on spaces of polynomials: Let $\mathcal{P} = \{a_0 + a_1x + ... + a_nx^n | n \in \mathbb{N} \cup \{0\}, a_0, ..., a_n \in \mathbb{R}\}$. Then $\mathcal{P}$ is an infinite dimensional space with basis $\{1, x, x^2, x^3, ...\}$. $\mathcal{P}$ has many norms. Let's use the 1-norm: for $p(x) = \sum_{k=0}^n a_kx^k$,

$$||p||_1 := \sum_{j=0}^n |a_j|$$

Define $A : \mathcal{P} \to \mathcal{P}$ by $Ap(x) := \int_0^x p(t)dt$, so

$$Ap(x) = \sum_{k=0}^n \frac{a_k}{k+1}x^{k+1}$$

Notice that

$$||Ap||_1 = \sum_{j=0}^{n} \left| \frac{a_j}{j+1} \right| \leq \sum_{j=0}^{n} |a_j| = ||p||_1$$

so if $||p_n||_1 \to 0$, then $||Ap_n||_1 \leq ||p_n||_1 \to 0$.

On the other hand, the derivative map $D : \mathcal{P} \to \mathcal{P}$, $Dp = p'$, is not bounded! Let $p_n(x) = \frac{x^n}{n}$. Then $||p_n||_1 = \frac{1}{n} \to 0$. But $||Dp_n||_1 = ||x^{n-1}||_1 = 1 \not\to 0$.

Example 4.65 Linear transformations on finite dimensional spaces are always continuous. Suppose $L : \mathbb{R}^n \to \mathbb{R}^m$ with any p-norm. Then L is bounded. The coordinate functions $p_i : \mathbb{R}^n \to \mathbb{R}$, $p_i(x_1, ..., x_n) = x_i$, for each $1 \leq i \leq m$ are all continuous as $|x_i| \leq ||[x_1 \, ... \, x_n]^T||_p$.

Next, we have that if S and T are bounded, so are $S + T$ and λS for all $\lambda \in \mathbb{R}$. Indeed,

$$||(S + T)(x_n)||_p \leq ||S(x_n)||_p + ||T(x_n)||_p \to 0$$

if $x_n \to 0$, and

$$||\lambda S(x_n)||_p = |\lambda|||S(x_n)||_p \to 0$$

So linear combinations of the f_i are continuous. That is function $f : \mathbb{R}^n \to \mathbb{R}$ of the form

$$f(x_1, ..., x_n) = \sum_{i=1}^{n} a_i f_i(x_1, ..., x_n)$$

If $L : \mathbb{R}^n \to \mathbb{R}^m$ is linear, we can write $L\mathbf{x} = A\mathbf{x}$ for some $m \times n$ matrix A. Write $A = [a_{ij}]$ then

$$A\mathbf{x} = \begin{bmatrix} a_{11}x_1 + a_{12}x_2 + ... + a_{1n}x_n \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + ... + a_{mn}x_n \end{bmatrix}$$

If $||\mathbf{x}_k||_p \to 0$, then by continuity of linear combinations of coordinate functions, each entry in $A\mathbf{x}$ must converge to 0. Consequently, $||A\mathbf{x}_k||_p \to 0$, so A is bounded.

Definition 4.45 (Homeomorphisms and Homeomorphic) Let X and Y be metric spaces and suppose that $\varphi : X \to Y$ is continuous. If φ is invertible and φ^{-1} is also continuous, we call φ a ***homeomorphism***. In this case, we say X and Y are ***homeomorphic*** and write $X \cong Y$.

Example 4.66 Continuous invertible maps need not be homeomorphisms. Indeed, from above we saw that $T : (C([0, 1]), || \cdot ||_\infty) \to (C([0, 1]), || \cdot ||_1)$, $Tf = f$, is continuous and invertible but has a discontinuous inverse.

Proposition 4.35 *The relation "≅" is an equivalence relation on metric spaces.*

Indeed $X \cong X$ using the identity, as it is a homeomorphism, symmetric since $X \cong Y$ implies there exists $\varphi : X \to Y$ a homeomorphism, which has $\varphi^{-1} : Y \to X$ also as a homeomorphism so $Y \cong X$. Finally, if $\varphi : X \to Y$ and $\psi : Y \to Z$ are homeomorphism, so is $\psi \circ \varphi : X \to Z$, so $X \cong Z$.

Example 4.67 With respect to the usual metric on $\mathbb{R}$, the sets $(0,1)$ and $(0,\infty)$ are homeomorphic. consider $f(t) = \frac{1}{1-t}$ on $(0,1)$. Then $f : (0,1) \to (1,\infty)$ is a homeomorphism with inverse $f^{-1}(t) = 1 - \frac{1}{t}$. Then $(0,\infty) \cong (1,\infty)$ via the map $t \mapsto t+1$.

Further, we have $(-\pi/2, \pi/2) \cong \mathbb{R}$ via $f(t) = \tan(t)$. But, $(a,b) \cup (c,d)$ for $a < b < c < d$ is not $\cong$ to $\mathbb{R}$. No homeomorphism is possible as $(a,b) \cup (c,d)$ can be expressed as the disjoint union of open balls, while $\mathbb{R}$ cannot be expressed as such.

Example 4.68 The unit circle and the unit cross are not homeomorphic, as we will see later by results on connected sets.

We remark that $f : X \to Y$ is a homeomorphism if and only if f is invertible and $f(x_n) \to f(x)$ in Y whenever $x_n \to x$ in X and $f^{-1}(y_n) \to f^{-1}(y)$ in X whenever $y_n \to y$ in Y. But f is bijective, so this is equivalent to $f(x_n) \to f(x)$ in Y if and only if $x_n \to x$ in X.

Definition 4.46 (Equivalent Norms) Suppose V is a normed linear space equipped with two norms $||\cdot||_a$ and $||\cdot||_b$. We say that these norms are ***equivalent*** if there exist positive constants c and C so that

$$c||x||_b \leq ||x||_a \leq C||x||_b$$

for all $x \in V$.

Proposition 4.36 *The relation of equivalence between norms on a normed linear space is an equivalence relation.*

Proof Indeed, $||\cdot||_a \sim ||\cdot||_a$ using $c = C = 1$, and if $||\cdot||_a \sim ||\cdot||_b$, so $c||\cdot||_b \leq ||\cdot||_a \leq C||\cdot||_b$, then $||\cdot||_b \sim ||\cdot||_a$ with $\frac{1}{C}||\cdot||_a \leq ||\cdot||_b \leq \frac{1}{c}||\cdot||_a$. Finally, if $||\cdot||_a \sim ||\cdot||_b \sim ||\cdot||_c$ with c, C and d, D, it follows that

$$dc||\cdot||_c \leq ||\cdot||_a \leq DC||\cdot||_c$$

so $||\cdot||_a \sim ||\cdot||_c$. □

Proposition 4.37 *Suppose V is a normed linear space with two norms $||\cdot||_a$ and $||\cdot||_b$. The identity map $\mathrm{Id}_V : (V, ||\cdot||_a) \to (V, ||\cdot||_b)$ is a homeomorphism if and only if $||\cdot||_a$ and $||\cdot||_b$ are equivalent.*

Proof First, suppose there exist $c < C$ positive with $c||\cdot||_b \leq ||\cdot||_a \leq C||\cdot||_b$ and note that Id_V is invertible. Then, if $v_n \to v$ in V with respect to $||\cdot||_a$, $||v_n - v||_b \leq \frac{1}{c}||v_n - v||_a \to 0$,so $\mathrm{Id}_V(v_n) \to \mathrm{Id}_V(v)$ with respect to $||\cdot||_b$. On the other hand, $||v_n - v||_a \leq C||v_n - v||_b$, so if $v_n \to v$ with respect to $||\cdot||_b$, then $v_n \to v$ with respect to $||\cdot||_b$.

Conversely, suppose Id_V is a homeomorphism. Let $B^a_r(0) = \{v \in V|||v||_a < r\}$ and $B^b_s(0) = \{v \in V|||v||_b < s\}$. Then $B^a_1(0)$ is open in $(V, ||\cdot||_a)$. Id_V is a homeomorphism, so Id_V^{-1} is continuous. So

$$(\mathrm{Id}_V^{-1})^{-1}(B^a_1(0)) = B^a_1(0)$$

is open in $(V, ||\cdot||_b)$. By definition of open, there exists $r > 0$ such that $B^b_r(0) \subseteq B^a_1(0)$. Hence, if $||v||_b < r$ then $||v||_a < 1$. If $v \neq 0$, then

$$\frac{vr}{2||v||_b} \in B_r^b(0)$$

so

$$\left\|\frac{vr}{2||v||_b}\right\|_a < 1$$

It follows that $||v||_a < \frac{2}{r}||v||_b$. On the other hand, Id_V is continuous so

$$\mathrm{Id}_v^{-1}(B_1^b(0) = B_1^b(0)$$

is open in $(V, ||\cdot||_a)$. By definition of open there exists $t > 0$ such that $B_t^a(0) \subseteq B_1^b(0)$. Hence if $||v||_a < t$, then $||v||_b < 1$. If $v \neq 0$, then

$$\frac{vt}{2||v||_a} \in B_t^1(0)$$

so

$$\left\|\frac{vt}{2||v||_a}\right\|_b < 1$$

Thus, $\frac{t}{2}||v||_b < ||v||_a < \frac{2}{r}||v||_b$, as desired. □

Proposition 4.38 *All of the p-norms on $\mathbb{R}^n$ are equivalent.*

Proof We will show $||\cdot||_\infty$ and $||\cdot||_p$ are equivalent for any $1 \le p < \infty$. Fix $\mathbf{x} = (x_1, ..., x_n) \in \mathbb{R}^n$. Then $||\mathbf{x}||_\infty = \max_{1\le i\le n}|x_i| = |x_j|$ for some $1 \le j \le n$. Then

$$||\mathbf{x}||_\infty = |x_j| = [|x_j|^p]^{1/p} \le \left[\sum_{i=1}^n |x_i|^p\right]^{1/p} = ||\mathbf{x}||_p$$

and

$$||\mathbf{x}||_p \le \left[\sum_{i=1}^n |x_j|^p\right]^{1/p} = n^{1/p}|x_j| = n^{1/p}||\mathbf{x}||_\infty$$

So $||\mathbf{x}||_\infty \le ||\mathbf{x}||_p \le n^{1/p}||\mathbf{x}||_\infty$, and the norms are equivalent. □

Definition 4.47 (Operator Norm) . Suppose V and W are normed linear spaces and $T : V \longrightarrow W$ is a linear transformation. The ***operator norm*** for T is

$$||T|| = \inf\{C > 0|||Tv||_W \le C||v||_V, \forall v \in V\}$$

We will adopt the convention that if the set on the right is empty, then $||T|| = \infty$.

Note that by definition $||Tv||_W \le ||T||||v||_V$ for all $v \in V$.

Proposition 4.39 *Suppose V and W are normed linear spaces and $T : V \longrightarrow W$ is a linear transformation. The following are equivalent:*

(1)T is bounded (i.e. continuous on V)

(2)T is continuous at 0

(3)$||T|| < \infty$

Proof We have already shown the equivalence between (1) and (2). We will show (2) $\iff$ (3).

First, suppose (2) holds. Fix $\varepsilon = 1$. Then there exists $\delta > 0$ such that $||v||_V < \delta$ then $||Tv||_W < 1$. If $v \neq 0$, then $\left\|\frac{\delta v}{2||v||_V}\right\|_V = \frac{\delta}{2}$, so

$$\left\|T\frac{\delta v}{2||v||_V}\right\|_W < 1$$

As T is linear it follows that

$$||Tv||_W < \frac{2}{\delta}||v||_V$$

Thus, T is bounded and $||T|| \leq \frac{2}{\delta}$.

Conversely, suppose (3) holds. We have $||Tv||_W \leq ||T||||v||_V$ for all $v \in V$, where $||T|| < \infty$. So if $v_n \longrightarrow 0$ in V, then

$$||Tv_n||_W \leq ||T|| \cdot ||v_n||_V \longrightarrow 0$$

so $Tv_n \longrightarrow 0$ in W and T is continuous at 0. □

In light of the above result, we can now say that an operator is bounded if and only if its operator norm is finite.

Definition 4.48 The set of all bounded operators from V to W is denoted $B(V, W)$. We use the notation $B(V)$ for $B(V, V)$.

Proposition 4.40 *Suppose $T \in B(V, W)$. Then*

$$||T|| = \sup_{||v||=1} ||Tv||_W = \sup_{||v||\leq 1} ||Tv||_W = \sup_{v\neq 0} \frac{||Tv||_W}{||v||_V}$$

Proof We shall show the second equals the third, and that $(1) \leq (2) \leq (3) \leq (1)$. For $(2) = (4)$, if $v \neq 0$ then $\frac{v}{||v||_V}$ is a unit vector and

$$T(v/||v||_V) = \frac{1}{||v||_V}Tv$$

applying supremums to both sides we obtain the desired inequality.

Next, for $(1) \leq (2)$, if $v \neq 0$ then

$$\left\|T\left(\frac{v}{||v||_V}\right)\right\|_W \leq (2)$$

by definition, so $||Tv||_W \leq ||v||_V(2)$m and so by definition of $||T||$, $||T|| \leq (2)$. $(2) \leq (3)$ is immediate as (3) is a supremum over a larger set. Finally, for $(3) \leq (1)$ suppose $C \geq 0$ satisfying $||Tv||_W \leq C||v||_V$ for all $v \in V$. We must show $(3) \leq C$ and ($(3) \leq ||T||$ by definition). If $||v|| \leq 1$, then $||Tv||_W \leq C$, so

$$(3) = \sup_{||v||\leq 1} ||Tv||_W \leq C$$

and the result follows. □

Proposition 4.41 *Let V and W be normed linear spaces.*

(1)$B(V, W)$ is a normed linear space using the operator norm

(2)$B(V, W)$ is complete when W is complete

(3)if Z is a normed vector space, $T \in B(V, W)$ and $S \in B(W, Z)$, then $S \circ T \in B(V, Z)$ and

$$||S \circ T|| \leq ||S|| \cdot ||T||$$

which is to say the operator norm is submultiplicative

Proof (1): First, let $\lambda \in \mathbb{R}$. Note that $||(\lambda T)v||_W = |\lambda|||Tv||_W \leq |\lambda|||T||||v||_V$ for all $v \in V$, so $||\lambda T|| \leq |\lambda|||T||$. Further, if $\lambda = 0$ the result is immediate, so suppose $\lambda \neq 0$, and it follows that

$$||Tv||_W = \frac{1}{|\lambda|}||(\lambda T)v||_W \leq \frac{||\lambda T||}{|\lambda|}||v||_V$$

for all $v \in V$, so $||T|| \cdot |\lambda| \leq ||\lambda T||$, and we have equality $||\lambda T|| = |\lambda| \cdot ||T||$. If $T = 0$, then $||Tv|| = 0$ for all $v \in V$, so $||T||0$. Conversely, if $||T|| = 0$, then $||Tv||_W = 0$ for all $||v|| \leq 1$. Then for all $w \neq 0$, we have

$$||T\frac{w}{||w||_V}||_W = 0$$

so $||Tw||_W = 0$, and as $|| \cdot ||_W$ is a norm, $Tw = 0$ for all $w \in V$. Thus, $T \equiv 0$. Finally, suppose $T, S \in B(V, W)$. Then if $||v||_V = 1$,

$$||(T + S)v||_W = ||Tv + Sv||_W \leq ||Tv||_W + ||Sv||_W \leq ||T|| + ||S|| < \infty$$

so using our previous equality chain result, $||T + S|| \leq ||T|| + ||S||$.

(2) Suppose $(T_n) \subseteq B(V, W)$ is a Cauchy sequence. Then for all $\varepsilon > 0$ there exists $N \in \mathbb{N}$ such that for $m, n \geq N$, $||T_n - T_m|| < \varepsilon$. Note for any $v \neq 0$ in V, we have

$$T_n(v/||v||_V) - T_m(v/||v||_V)||_W < \varepsilon$$

so

$$||T_n v - t_m v||_W < \varepsilon||v||_V$$

This implies $(T_n v)$ is a Cauchy sequence in W. W being complete, we can define $T : V \longrightarrow W$ by

$$Tv := \lim_{n\to\infty} T_n v$$

This defines a linear map $T : V \longrightarrow W$ since

$$T(v + w) = \lim_{n\to\infty} T_n(v + w) = \lim_{n\to\infty} T_n v + \lim_{n\to\infty} T_n w = Tv + Tw$$

and

$$T(\lambda v) = \lim_{n\to\infty} T_n(\lambda v) = \lim_{n\to\infty} \lambda T_n v = \lambda T v$$

First we show that $||T - T_n|| \to 0$. If $v \in V$ with $||v|| = 1$, then

$$||Tv - T_n v||_W \leq ||Tv - T_m v||_W + ||T_m v - T_n v||_W < ||Tv - T_m v||_W + \varepsilon \to 0$$

since $||Tv - T_m v||_W \to 0$ as $m \to \infty$. Then $||Tv - T_n v|| \leq \varepsilon$. Taking the supremum over all $||v|| = 1$,

$$||T - T_n|| \leq \varepsilon$$

Finally, $T \in B(V, W)$ since $||T|| \leq ||T - T_n|| + ||T_n|| < \infty$.

(3) If $T \in B(V, W), S \in B(W, Z)$, then $S \circ T : V \to Z$ is linear. If $||v||_V = 1$, then

$$||STv||_Z \leq ||S||||Tv||_W \leq ||S||||T||||v||_V = ||S||||T||$$

Now taking the supremum over all $||v|| = 1$ we get $||ST|| \leq ||S||||T||$. □

Remark 4.2 Recall that we can represent linear transformations on finite dimensional vector spaces by matrices. If $T : \ell^2_n \to \ell^2_m$ is linear, than its standard matrix representation is

$$A = [Te_1\ Te_2\ \dots\ Te_n]$$

for $\{e_1, \dots, e_n\}$ the standard orthonormal basis for $\mathbb{R}^n$. In this way, $T(\mathbf{x}) = A\mathbf{x}$.

Example 4.69 We consider the ***operator spectral norm*** for a diagonal matrix. The ***spectral norm*** means the operator norm on $A \in M_{m\times n}$. Consider $T_A : \ell^2_n \to \ell^2_n$ with $A = \begin{bmatrix} a_1 & & & 0 \\ & a_2 & & \\ & & \ddots & \\ 0 & & & a_n \end{bmatrix}$, so $T\mathbf{x} = [a_1 x_2\ \dots\ a_n x_n]^T$. Let $a = \max\{|a_1|, \dots, |a_n|\}$. Then

$$||T\mathbf{x}||_2^2 = \sum_{i=1}^{n} a_i^2 x_i^2 \leq \sum_{i=1}^{n} a^2 x_i^2 = a^2 ||\mathbf{x}||_2^2$$

so $||T|| \leq a$. There is some i with $a = |a_i|$, so $||Te_i||_2 = |a_i| = a||e_i||_2$, as $||e_i||_2 = 1$, so $||T|| \geq a$, since $||T|| = \sup_{||v||=1} ||Tv||$. Thus, $||T|| = a$.

Example 4.70 For any $T : \ell^2_n \to \ell^2_m$, we have seen T is bounded. Let $A = [T]$ be the standard matrix for T. The $n \times n$ matrix $A^T A$ is positive semidefinite. That is $A^T A$ is symmetric, and has all eigenvalues ≥ 0. Indeed, observe that for all $v \in \ell^2_n$, $v^T A^T A v = (Av) \cdot (Av) = ||Av||_2^2 \geq 0$. Then, if $\lambda \in \mathbb{R}$ is an eigenvalue with eigenvector $v \in \ell^2_n$, then

$$0 \leq ||Av||_2^2 = v^T A^T A v = v^T(\lambda v) = \lambda ||v||_2^2$$

As $v \neq 0$, $||v||_2^2 \neq 0$, so $\lambda \geq 0$, as claimed.

Now, there is an orthogonal matrix U, $U^T = U^{-1}$, such that $U^T(A^T A)U = D$ is diagonal. The diagonal entries of D, $\lambda_1, \dots, \lambda_n$, are called the singular values for A with $\lambda_i \geq 0$. Define $\sqrt{D}$ as the diagonal matrix with $\sqrt{\lambda_i}$ on the diagonal, and note that $\sqrt{D}^2 = D$. We have $UU^T = U^TU = I_n$, so for any $\mathbf{x} \in \ell_n^2$,

$$||\mathbf{x}||_2^2 = \mathbf{x} \cdot \mathbf{x} = U^TU\mathbf{x} \cdot \mathbf{x} = U\mathbf{x} \cdot U\mathbf{x} = ||U\mathbf{x}||_2^2$$

and $||U^T\mathbf{x}||_2 = ||\mathbf{x}||_2$. Let $\mathbf{x}$ be a unit vector. Then

$$\begin{aligned} ||A\mathbf{x}||_2^2 = A\mathbf{x} \cdot A\mathbf{x} &= A^TA\mathbf{x} \cdot \mathbf{x} \\ &= UDU^T\mathbf{x} \cdot \mathbf{X} \\ &= DU^T\mathbf{x} \cdot U^T\mathbf{x} \\ &= \sqrt{D}U^T\mathbf{x} \cdot \sqrt{D}U^T\mathbf{x} \\ &= ||\sqrt{D}U^T\mathbf{x}||_2^2 \\ &\leq ||\sqrt{D}||^2||U^T\mathbf{x}||_2^2 = ||\sqrt{D}||^2 \end{aligned}$$

So $||A|| \leq ||\sqrt{D}||$. Similarly, $||\sqrt{D}\mathbf{x}||_2^2 = ||AU\mathbf{x}||_2^2 \leq ||A||^2$, so $||\sqrt{D}|| \leq ||A||$. Together these inequalities imply $||A|| = ||\sqrt{D}|| = \max_{1 \leq i \leq n} \sqrt{\lambda_i}$.

Example 4.71 Consider $A = \begin{bmatrix} 2 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix} : \ell_3^2 \to \ell_2^2$. Then

$$A^TA = \begin{bmatrix} 5 & 2 & 0 \\ 2 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

has characteristic polynomial $t[(t-5)(t-1)-4] = t(t^2-6t+1)$, so its eigenvalues are $\lambda = 0, 3 \pm 2\sqrt{2}$. Thus,

$$||A|| = \sqrt{3 + 2\sqrt{2}}$$

Even if A is $n \times n$ and diagonalizable, it may not be the case that $||A||$ is equal to the largest eigenvalue of A (in modulus). It turns out that $||A||$ is always at least this number, though.

By convention, we will also insist that if $L : \mathbb{R}^n \to \mathbb{R}^m$ is linear, we are using the 2-norm for both spaces.

Definition 4.49 Suppose (X, d_X) and (Y, d_Y) are metric spaces and $\varphi : X \to Y$. We say that φ is an ***isometry*** if $d_Y(\varphi(x), \varphi(y)) = d(x, y)$ for all $x, y \in X$

That is to say, isometries preserve distances.

Proposition 4.42 *Isometries are continuous.*

Proof Fix $\varepsilon > 0$ and $\delta = \varepsilon$. Then if $d_X(x, y) < \delta$, $d_Y(\varphi(x), \varphi(y)) = d_X(x, y) < \delta = \varepsilon$. □

Example 4.72 We saw above if $U : \ell_n^2 \to \ell_n^2$ is orthogonal, then $||U\mathbf{x}||_2^2 = ||\mathbf{x}||_2^2$. So$U$ is an isometry since

$$||U\mathbf{x} - U\mathbf{y}||_2^2 = ||U(\mathbf{x} - \mathbf{y})||_2^2 = ||\mathbf{x} - \mathbf{y}||_2^2$$

Definition 4.50 An ***affine function*** $\varphi : \ell_n^2 \to \ell_n^2$ is given by

$$\varphi(\mathbf{x}) = U\mathbf{x} + \mathbf{b}$$

for some orthogonal matrix U and some vector $\mathbf{b} \in \ell_n^2$.

From the previous example, affine functions are common examples of isometries. In fact there are no other isometries.

Example 4.73 Suppose $\varphi : \mathbb{R} \to \mathbb{R}$ is an isometry. First assume $\varphi(0) = 0$. Then $|\varphi(x) - \varphi(0) = |x - 0| = |x|$, so $\varphi(x) = \pm x$. We want $\varphi(x) = x$ for all $x \in \mathbb{R}$, or $\varphi(x) = -x$ for all x. Suppose this was not the case. Then there exist $x \neq y \in \mathbb{R}\backslash\{0\}$ with $\varphi(x) = x$ and $\varphi(y) = -y$. But then $0 \neq |x - y| = |\varphi(x) - \varphi(y)| = |x + y|$, which implies $(x - y)^2 = (x + y)^2$, which occurs if and only if $-2xy = 2xy$, or either $x = 0$ or $y = 0$, which is a contradiction. Thus, $\varphi(x) = x$ or $\varphi(x) = -x$. If $p : \mathbb{R} \to \mathbb{R}$ is any isometry, then $\varphi : \mathbb{R} \to \mathbb{R}$ defined by $\varphi(x) = p(x) - p(0)$ is an isometry with the same property as above, so $\varphi(x) = x$ or $-x$. Thus, $p(x) = x + b$ or $p(x) = -x + b$ for some $b \in \mathbb{R}$.

Example 4.74 The space of smooth functions is complete: the space $(C^1([0, 1]), ||| \cdot |||)$. Let $(f_n) \subseteq C^1([0, 1])$ be Cauchy with respect to $||| \cdot |||$. Then the f_n are Cauchy with respect to $|| \cdot ||_\infty$, so as $(C([0, 1]), || \cdot ||_\infty)$ is complete, there exists $f \in C([0, 1])$ such that $f_n \to_u f$. Similarly, f_n' are Cauchy with respect to $|| \cdot ||_\infty$, so there exists $g \in C([0, 1])$ such that $f_n' \to_u g$. Then, for $x \in [0, 1]$,

$$\begin{aligned} \int_0^x g(t)dt &= \lim_{n\to\infty} \int_0^x f_n'(t)dt && \text{(by uniform convergence)} \\ &= \lim_{n\to\infty} (f_n(x) - f_n(0)) && \text{(by FTOC2)} \\ &= f(x) - f(0) && \text{(by pointwise convergence)} \end{aligned}$$

Then, by the FTOC1, $f(x) - f(0)$ is differentiable at $x \in [0, 1]$ as $g(x)$ is continuous on $[0, 1]$, and $(f(x) - f(0))' = f'(x) = g(x)$. Thus, $f' \in C([0, 1])$. Hence, $f \in C^1([0, 1])$, and

$$|||f - f_n||| = ||f - f_n||_\infty + ||f' - f_n'||_\infty = ||f - f_n||_\infty + ||g - f_n'||_\infty \to 0$$

so $(C^1([0, 1]), ||| \cdot |||)$ is complete.

4.9.2 Maps on Compact Spaces

Proposition 4.43 *If $f : X \to Y$ is continuous and $K \subseteq X$ is compact in X, then $f(K)$ is compact in Y.*

Proof Let $(y_j) \subseteq f(K)$. Then there exists $(x_j) \subseteq K$ such that $f(x_j) = y_j$ for all j. But, K is compact so there exists a convergent subsequence (x_{j_ν}) which converges to some $x \in K$. As f is continuous, $f(x_{j_\nu}) \to f(x)$. But then $y_{j_\nu} = f(x_{j_\nu}) \to f(x) \in f(K)$, so (y_j) has a convergent subsequence and $f(K)$ is compact. □

Proposition 4.44 *If X is a compact metric space and $f : X \to \mathbb{R}$ is continous, then f assumes a max and min value in X.*

This follows from the characterization of compact sets in $\mathbb{R}$, which are precisely the closed and bounded sets.

Definition 4.51 For $f : X \to \mathbb{R}$, we define

$$\sup_X f = \begin{cases} \sup_{x\in X} f(x) & \text{if } f(X) \text{ is bounded from above} \\ \infty & \text{if not bounded above} \end{cases}$$

and

$$\inf_X f = \begin{cases} \inf_{x\in X} f(x) & \text{if } f(X) \text{ is bounded from below} \\ -\infty & \text{if not bounded below} \end{cases}$$

Using this convention, we define a notion of a limit which always exists in the extended reals, even if the actual limit does not exist.

Definition 4.52 For any sequence $(x_n) \subseteq \mathbb{R}$, we define

$$\limsup_{n\to\infty} x_n := \lim_{n\to\infty} \left(\sup_{k\geq n} x_k\right)$$

and

$$\liminf_{n\to\infty} x_n := \lim_{n\to\infty} \left(\inf_{k\geq n} x_k\right)$$

Note $\sup_{k\geq n} x_k$ is a ***decreasing sequence*** and $\inf_{k\geq n} x_k$ is an ***increasing sequence***.

So lim sup is the limit of a monotone decreasing sequence and lim inf is the limit of a monotone increasing sequence. Further we have that $(x_n) \subseteq \mathbb{R}$ is convergent if and only if $\limsup x_n = \liminf x_n$.

4.9.3 Uniform Continuity

We now define a more powerful notion of continuity of functions on metric spaces.

Definition 4.53 Let (X, d_X) and (Y, d_Y) be metric spaces. A function $f : X \to Y$ is said to be ***uniformly continuous*** on X if for all $\varepsilon > 0$ there exists $\delta > 0$ such that

$$d_X(x, y) < \delta \implies d_Y(f(x), f(y)) < \varepsilon$$

for all $x, y \in X$.

This is the same definition as for uniformly continuous functions on the real line.

Example 4.75 An example of a continuous function which is not uniformly continuous is the ***topoligists sine curve***, $\sin(1/x)$ (Image here)

Example 4.76 $f(x) = x^2$ is continuous. For any $\varepsilon > 0$, observe $|x^2 - y^2| < \varepsilon \implies |x-y||x+y| < \varepsilon$. If $|x - y| < \delta$, $|x-y||x+y| \leq (2|x|+1)\delta$ if $\delta = \min\left\{\frac{\varepsilon}{2|x|+1}, 1\right\}$. But his depends on x. This does work for x' such that $|x'| \leq |x|$, but for bigger x' we would need a smaller δ, so f is not uniformly continuous.

Proposition 4.45 *Suppose $\varphi : X \to Y$ is uniformly continuous. If (x_n) is Cauchy in X, then $(\varphi(x_n))$ is Cauchy in Y.*

Proof Fix $\varepsilon > 0$ such that $d_X(x, y) < \delta$ implies $d_Y(\varphi(x), \varphi(y)) < \varepsilon$. If (x_n) is Cauchy there exists $N \in \mathbb{N}$ such that for $m, n \geq N$, $d_X(x_n, x_m) < \delta$. Then, for all $m, n \geq N$ we have

$$d_Y(\varphi(x_n), \varphi(x_m)) < \varepsilon$$

so $(\varphi(x_n))$ is Cauchy in Y. □

Example 4.77 Consider $f(x) = 1/x$ and $g(x) = \sin(1/x)$, which are not uniformly continuous on $(0, \infty)$. For example, $x_n = 1/n$ is Cauchy in $(0, \infty)$ but $f(1/n) = n$ is not, and $g(1/n) = \sin(n)$ is also not.

Example 4.78 Consider $f : \mathbb{R}^2 \to \mathbb{R}$, $f(x, y) = x^2 + y^2$. Fix $\varepsilon = 1$ and $\delta > 0$. Let $x = 1/\delta, y = 1/\delta$. Then

$$|f(x + \delta/2, y + \delta/2) - f(x, y)| = 2(1/\delta + \delta/2)^2 - 2/\delta^2| = 2|1 + \delta^2/4| > 2 > \varepsilon$$

Then $||(x + \delta/2, y + \delta/2) - (x, y)||_2 = ||(\delta/2, \delta/2)||_2 = \sqrt{2}\delta/2 < \delta$. Thus, f cannot be uniformly continuous

Example 4.79 Consider $f : \mathbb{R}^2\backslash\{0\} \to \mathbb{R}$ defined by $f(x, y) = \frac{1}{x^2+y^2}$. Then $f(1/n, 0) = n^2$ is not Cauchy in $\mathbb{R}$, but $(1/n, 0)$ is Cauchy in $\mathbb{R}^2$.

Proposition 4.46 *Bounded operators are uniformly continuous.*

Proof Let $T \in B(V, W)$ and let $\varepsilon > 0$. Then

$$||Tv - Tw|| \leq ||T||||v - w||$$

for $||T|| < \infty$. Let $\delta = \frac{\varepsilon}{||T||+1}$. Then if $||v - w|| < \delta$, $||Tv - Tw|| < \varepsilon$. □

Further, we also have that all isometries are uniformly continuous as they are in fact Lipschitz continuous.

Proposition 4.47 *If X is compact and $f : X \to Y$ is continuous, then f is uniformly continuous.*

Proof Let $\varepsilon > 0$. Let $f(X) = \bigcup_{x\in X} B_{\varepsilon/2}(f(x))$ cover the image. For each x there exist $\delta_x > 0$ such that $f(B_{\delta_x}(x)) \subseteq B_{\varepsilon/2}(f(x))$. Then $X \subseteq \bigcup_{x\in X} B_{\delta_x/2}(x)$ is an open cover, so there exists $x_1, ..., x_N \in X$ such that $X \subseteq \bigcup_{i=1}^N B_{\delta_{x_i}/2}(x_i)$ since X is compact. Let $\delta = \min_{1\le i\le N}(\delta_{x_i}/2)$. Then let $x, y \in X$ such that $d_X(x,y) < \delta$. Since the $B_{\delta_{x_i}/2}(x_i)$ cover X, there exists $1 \le i \le N$ such that $x \in B_{\delta_{x_i}/2}(x_i)$. Then

$$d_X(x_i, y) \le d_X(x_i, x) + d_X(x, y) < \delta_{x_i}/2 + \delta \le \delta_{x_i}$$

so $x, y \in B_{\delta_{x_i}}(x_i)$. It follows that $f(x), f(y) \in B_{\varepsilon/2}(f(x_i))$ so

$$d_Y(f(x), f(y)) \le d_Y(f(x), f(x_i)) + d_Y(f(x_i), f(y) < \varepsilon/2 + \varepsilon/2 = \varepsilon$$

completing the proof. □

Now we come to the notion of an isomorphism in the category of metric spaces:

Definition 4.54 A function $f : X \to Y$ is said to be a ***homeomorphism*** if it is continuous, bijective, and f^{-1} is continuous.

Proposition 4.48 *If X is a compact metric space, then $f : X \to Y$ being continuous and bijective implies f^{-1} is continuous.*

Proof Let $g = f^{-1} : Y \to X$. Note g is continuous in Y if and only if $g^{-1}(V) = f(V)$ is closed in Y for all V closed in X. Note V is compact in X if it is closed, so $f(V)$ is compact. Then as Y is a metric space V being compact implies it is closed. Thus g is continuous, as desired. □

4.9.4 Sequences and Series of Functions

Next we consider convergence of sequences of functions:

Definition 4.55 Let $f_j : X \to Y$, $j \in \mathbb{N}$, be a sequence of functions. If $f : X \to Y$, and $f_j(x) \to f(x)$ for all $x \in X$, we say that $f_j \to f$ ***pointwise*** on X.

Definition 4.56 A sequence $f_j : X \to Y$ converges ***uniformly*** to $f : X \to Y$ if and only if for all $\varepsilon > 0$, there exists $N \in \mathbb{N}$ such that if $j \ge N$,

$$d_Y(f_j(x), f(x) < \varepsilon, \forall x \in X$$

or equivalently

$$\sup_{x\in X} d_Y(f_j(x), f(x)) \le \varepsilon$$

Proposition 4.49 *If $f_j : X \to Y$ are continuous and converge uniformly to $f : X \to Y$, then f is continuous.*

Proof Let $\varepsilon > 0$ and $x \in X$. Then there exists $N \in \mathbb{N}$ such that for $j \geq N$, $d_Y(f_j(y), f(y)) < \varepsilon/3$ for all $y \in X$. As f_N is continuous, there exists $\delta > 0$ such that $f_N(B_\delta(x)) \subseteq B_{\varepsilon/3}(f_N(x))$. Then for $d_X(x, y) < \delta$,

$$d_Y(f(x), f(y)) \leq d_Y(f(x), f_N(x)) + d_Y(f_N(x), f_N(y)) + d_Y(f_N(y), f_N(x) < \varepsilon$$

so f is continuous. □

Example 4.80 If $f_j : [0, 1] \to [0, 1]$ is defined by $f_j(x) = x^j$, all of which are continuous, then $f_j(x)$ converge pointwise to $f(x) = \begin{cases} 0 & x < 1 \\ 1 & x = 1 \end{cases}$, which is discontinuous, and hence the convergence can't be uniform.

Definition 4.57 A sequence of functions $f_j : X \to Y$ is said to be ***uniformly Cauchy*** if for all $\varepsilon > 0$, there exists $N \in \mathbb{N}$ such that for $j, k \geq N$,

$$\sup_{x \in X} d_Y(f_j(x), f_k(x)) \leq \varepsilon$$

or equivalently $\lim\limits_{j,k \to \infty} \sup_{x \in X} d_Y(f_j(x), f_k(x)) = 0$.

Proposition 4.50 *If Y is a complete metric space and $f_j : X \to Y$ is uniformly Cauchy, then there exists $f : X \to Y$ such that f_j converge uniformly to f.*

Proof As f_j is uniformly Cauchy, for each $x \in X$, $(f_j(x)) \subseteq Y$ is Cauchy. As Y is complete there exists $y_x \in Y$ such that $f_j(x) \to y_x$. Then define $f : X \to Y$ by $f(x) = \lim\limits_{j \to \infty} f_j(x)$. Fix $\varepsilon > 0$. As f_j is uniformly Cauchy there exists $N \in \mathbb{N}$ such that $j \geq N$ and $k \geq 0$ implies $d_Y(f_j(x), f_{j+k}(x)) < \varepsilon$. Taking the limit as k goes to infinity we have $d_Y(f_j(x), f(x)) \leq \varepsilon$ for all $x \in X$. Thus, f_j converges uniformly to f. □

Now we consider basic properties of series:

Definition 4.58 We say a sequence $f_j : X \to \mathbb{R}^n$ is ***pointwise summable*** if and only if the sequence

$$s_n(x) = \sum_{j=0}^{n} f_j(x)$$

is ***pointwise convergent***.

Definition 4.59 $f_j : X \to \mathbb{R}^n$ is ***uniformly summable*** if and only if $s_n(x) = \sum_{j=0}^{n} f_j(x)$ is ***uniformly convergent***.

Theorem 4.15 (Weierstrass M-Test (General)) *Let $f_j : X \to \mathbb{R}^n$ be a sequence of functions. Assume there exist $M_k \in \mathbb{R}$ such that* $\sup_{x \in X} ||f_k(x)|| \leq M_k$ *and*

$$\sum_{k=0}^{\infty} M_k < \infty$$

Then the series $\sum_{k=0}^{n} f_k(x)$ *converges uniformly on* X *to a limit* $s(x)$.

Proof Suppose the hypotheses of the theorem. As $\sum_{k=0}^{n} M_k$ is Cauchy, for $\varepsilon > 0$ there exists $N \in \mathbb{N}$ such that for $j, k \geq N$,

$$\left| \sum_{n=j+1}^{k} f_n(x) \right| \leq \sum_{n=j+1}^{k} |f_n(x)| \leq \sum_{n=j+1}^{k} M_n < \varepsilon$$

for all $x \in X$, so $\sum_{n=0}^{k} f_n(x)$ is uniformly Cauchy, and hence uniformly convergent since $\mathbb{R}^n$ is complete. □

4.10 Connected Sets

We now define a notion for what it means for a space to be connected, or not disconnected.

Definition 4.60 A topological space (X, τ) is ***not connected*** or ***disconnected*** if there exists $U, V \in \tau$ such that $U, V \neq \emptyset$, $U \cap V = \emptyset$, and

$$X = U \sqcup V$$

If no such U and V exist, then X is said to be ***connected***.

Equivalently a space X is connected if and only if the only clopen sets are X and $\emptyset$. For metric spaces we say:

Definition 4.61 A subset S of a metric space X is ***disconnected*** if there are open subsets U and V of X such that

- The intersection $U \cap S$ and $V \cap S$ are non-empty
- $(U \cap S) \cap (V \cap S) = \emptyset$
- $S = (U \cap S) \cup (V \cap S)$

The pair (U, V) is called a ***separation of** S*. If S has no separations, then we say that S is ***connected***

If $S = X$, this just means $X = U \cap V$ with $U \cap V = \emptyset$.

Example 4.81 A disconnected set in ℓ_2^2 is $B_1((-1, 0)) \cup B_1((1, 0))$.

Example 4.82 Connected subsets of a discrete space: $Y \subseteq X$ is connected if and only if $Y = \{y\}$. $\{y\}$ is connected. If $|Y| \geq 2$, let $y_0 \in Y$, $U = \{y_0\}$, and $V = Y \backslash \{y_0\}$, which is a separation.

Proposition 4.51 *All intervals in $\mathbb{R}$ are connected.*

Proof Suppose I is an interval in $\mathbb{R}$. That is for all $a, b \in I$, with $a < b$, $[a, b] \subseteq I$. Suppose we have a separation $A, B \subseteq \mathbb{R}$ open such that $I = (I \cap A) \cup (I \cap B)$, with $I \cap A, I \cap B \neq \emptyset$, and $(I \cap A) \cap (I \cap B) = \emptyset$. Let $A' = I \cap A$ and $B' = I \cap B$. Let $a \in A' \subseteq I$ and $b \in B' \subseteq I$ so $a \neq b$. Without loss of generality suppose $a < b$. Then $[a, b] \subseteq I$ and covered by A', B'. Then let $s = \sup A' \cap [a, b]$. Then $s \leq b$. We proceed by cases:

- Suppose $s \in A'$. Then as A is open, there exists $\varepsilon > 0$ such that $(s - \varepsilon, s + \varepsilon) \subseteq A$. Let $\varepsilon' = \min\{\varepsilon, |s - b|\} > 0$. Then $[s, s + \varepsilon') \subseteq A' \cap [a, b]$. In particular $s + \varepsilon'/2 \in A' \cap [a, b]$, contradicting the fact that $s = \sup A' \cap [a, b]$.
- Suppose $s \in B' \cap [a, b]$. As B is open, there exists $\varepsilon > 0$ such that $B_\varepsilon(s) \subseteq B$. Then let $\varepsilon' = \min\{\varepsilon, |s - a|\} > 0$, so $(s - \varepsilon', b] \subseteq B' \cap [a, b]$ and $s - \varepsilon'/2 \in B' \cap [a, b]$, so $t \leq s - \varepsilon/2$ for all $t \in A' \cap [a, b]$, contradicting the fact $s = \sup A' \cap [a, b]$. □

Definition 4.62 If (X, τ) is a topological space and $A \subseteq X$, the subspace topology on A is defined by

$$\tau_A := \{U \cap A \subseteq A : U \in \tau\}$$

Note if $\iota : A \hookrightarrow X$ is the inclusion, τ_A is the coarsest topology/weakest topology making ι continuous

$$U \subseteq A \text{ open} \iff \exists V \in \tau; \iota^{-1}(V) = V \cap A = U$$

Definition 4.63 A topological space (X, τ) is ***path connected*** if and only if for all $p, q \in X$, there exists a continuous map

$$\gamma : [0, 1] \to X,\ \gamma(0) = p,\ \gamma(1) = q$$

Proposition 4.52 *Path connected implies connected.*

Proof Suppose X is path connected. Towards a contradiction suppose X has a separation A, B. Let $a \in A$, $b \in B$, and $\gamma : [0, 1] \to X$ with $\gamma(0) = a, \gamma(1) = b$. Then $\gamma^{-1}(A) \subseteq [0, 1]$, $\gamma^{-1}(B) \subseteq [0, 1]$ are open, disjoint, and cover $[0, 1]$, contradicting the fact that $[0, 1]$ is connected. Thus X must be connected. □

Example 4.83 The metric space

$$X = \{(0, y) \in \mathbb{R}^2 : y \in [-1, 1]\} \cup \{(x, \sin 1/x) \in \mathbb{R}^2 : x \in (0, 1]\}$$

with metric $d_X = d_{\mathbb{R}^2}$ is compact and connected, but not path connected.

Theorem 4.16 (Intermediate Value Theorem) *Suppose (X, τ) is a connected space and $f : X \to \mathbb{R}$ is continuous. Suppose $p, q \in X$ such that $f(p) = a < b = f(q)$. Then for all $c \in (a, b)$ there exists $z \in X$ such that $f(z) = c$.*

Proof Let $A = f^{-1}((-\infty, c))$ and $B = f^{-1}((c, \infty))$, so A and B are open, non-empty, and disjoint. Thus, as X is connected, $X \neq A \cup B$ so there must exist $t \in f^{-1}(\{c\})$, so in particular $f(t) = c$. □

Proposition 4.53 *Suppose X and Y are metric spaces and $f : X \to Y$ is continuous. If S is connected in X, then $f(S)$ is connected in Y.*

Proof If U, V is a separation of $f(S)$, so $f(S) \subseteq U \cup V$, then

$$S \subseteq f^{-1}(U \cup V) = f^{-1}(U) \cup f^{-1}(V)$$

Since f is continuous, these are both open, and are in fact a separation of S, which is a contradiction.□

Chapter 5

Approximations and Continuous Functions

5.1 Algebras

Definition 5.1 (Algebras and Subalgebras) Suppose $\mathcal{A}$ is a (real or complex) vector space. We say that $\mathcal{A}$ is an ***algebra*** if there is a multiplication $\mathcal{A} \times \mathcal{A} \to \mathcal{A}$ satisfying the following properties for all $A, B, C \in \mathcal{A}$ and constants r:

$(1)(AB)C = A(BC)$ (associative)

$(2)r(AB) = (rA)B = A(rB)$ (scalar multiplication)

$(3)A(B + C) = AB + AC$ and $(A + B)C = AC + BC$ (distributive)

If in addition there is an element $1 \in \mathcal{A}$ satisfying $1A = A1 = A$ for all A, then we say that $\mathcal{A}$ is ***unital***. A subspace $\mathcal{B}$ of $\mathcal{A}$ is called a ***subalgebra*** if $\mathcal{B}$ is itself an algebra with the multiplication inherited from $\mathcal{A}$.

Definition 5.2 If $\mathcal{A}$ is an algebra equipped with a vector space norm that also satisfies $||AB|| \leq ||A||\ ||B||$ for all $A, B \in \mathcal{A}$, it is called a ***normed algebra***. If $\mathcal{A}$ is complete with respect to this norm, it is called a ***Banach algebra***.

Example 5.1 Algebras of matrices: $M_n(\mathbb{R})$ or $M_n(\mathbb{C})$ is a unital Banach algebra under the spectral norm and matrix multiplication. Recall the spectral norm is an operator norm for maps from $\ell_n^2 \to \ell_n^2$. Such norms always satisfy $||AB|| \leq ||A||\ ||B||$, as shown in a separate section. The upper triangular matrices $T_n(\mathbb{R})\{[a_{ij}]|a_{i,j} = 0\ if\ i > j\}$ is a unital subalgebra of $M_n(\mathbb{R})$. The strictly upper triangular $J_n(\mathbb{R}) = \{[a_{ij}|a_{i,j} = 0\ if\ i \geq j\}$ are a non-unital subalgebra of $T_n(\mathbb{R})$ and $M_n(\mathbb{R})$.

Definition 5.3 Let X be a compact metric space. Define the space of continuous functions on X by

$$C(X) = \{f : X \to \mathbb{R}|f \text{ is continuous}\}$$

Note, we can replace $\mathbb{R}$ with $\mathbb{C}$ above, but for the remainder of this section we will restrict to the real case.

Proposition 5.1 *$C(X)$ is a Banach algebra with the uniform norm and pointwise product.*

Proof As X is compact and $f \in C(X)$ is continuous, $||f||_\infty := \sup_{x \in X} |f(x)| < \infty$. The product is pointwise, so $fg(x) = f(x)g(x)$. It is evident that $C(X)$ is a NLS under $|| \cdot ||_\infty$. It is also a normed algebra since for $f, g \in C(X)$,

$$|f(x)g(x)| = |f(x)||g(x)| \leq ||f||_{|infty}||g||_\infty$$

Taking the supremum, $||fg||_\infty \leq ||f||_\infty ||g||_\infty$. For sompleteness, suppose (f_n) is Cauchy. Then

$$|f_n(x) - f_m(x)| \leq ||f_n - f_m||_\infty$$

which implies $(f_n(x))$ is Cauchy in $\mathbb{R}$ for each $x \in X$, and hence there is a pointwise limit $\lim\limits_{n \to \infty} f_n(x) =: f(x)$. For every $\varepsilon > 0$, there exists $N \in \mathbb{N}$ such that for $n \geq N$ and $k \in \mathbb{N}$, $||f_n - f_{n+k}||_\infty < \varepsilon$. Taking the limit as $k \to \infty$, $||f_n - f||_\infty \leq \varepsilon$. Thus, $||f_n - f||_\infty \to 0$. Further, for all $x, y \in X$,

$$|f(x) - f(y)| \leq |f(x) - f_n(x)| + |f_n(x) - f_n(y)| + |f_n(y) - f(y)| \to 0$$

as $d_X(x, y) \to 0$, so $f \in C(X)$. □

Definition 5.4 Let X be a compact metric space. A subset $S \subseteq C(X)$ is said to ***separate*** X if for all $x \neq y$ in X, there is an $f \in S$ such that $f(x) \neq f(y)$.

Example 5.2 The algebra of even polynomials on $[-1, 1]$: Let $\mathcal{A} = \operatorname{span}\{1, x^2, x^4, ...\} \subseteq C([-1, 1])$. $\mathcal{A}$ is a unital subalgebra of $C([-1, 1])$. $\mathcal{A}$ does not separate $[-1, 1]$. If f is any function in $\mathcal{A}$ we have $f(-x) = f(x)$ for all $x \in [-1, 1]$. $\mathcal{A}$ does seperate $[01,]$. For example, if $0 \leq x < y \leq 1$, then $f(x) = x^2 \neq y^2 = f(y)$.

Example 5.3 Let $S = \operatorname{span}\{1, x(x-1)f(x) | f \in C([0, 1])\} \subseteq C([0, 1])$. A generic element in S is of the form $g(x) = c + x(x-1)f(x)$ where $f \in C([0, 1])$. S is a subspace since it is a span. S is also a subalgebra since

$$[c_1 + x(x-1)f_1(x)][c_2 + x(x-1)f_2(x)] = c_1c_2 + x(x-1)[c_2 + c_2 + x(x-1)f_1(x)f_2(x)] \in S$$

S does not seperate $[0, 1]$ since $g(0) = g(1)$ for all $g \in S$.

Example 5.4 The algebra of polynomials: Let $K \subseteq \mathbb{R}$ be compact. The algebra of polynomials $\mathcal{P} \subseteq C(K)$ separates K. If $x_1 \neq y_1$ are in K and $a, b \in \mathbb{R}$, then define

$$p_1(x) = \frac{a}{x_1 - y_1}(x - y_1) \text{ and } p_2(x) = \frac{b}{y_1 - x_1}(x - x_1)$$

and notice $p_1(x_1) = a, p_1(y_1) = 0$, and $p_2(x_1) = 0, p_2(y_1) = b$.

5.2 Stone-Weierstrass

Theorem 5.1 (Weierstrass 1885) *Suppose* $K \subseteq \mathbb{R}^n$ *is compact. Then for each* f *in* $C(K)$*, there is a sequence of polynomials* p_n *so that* p_n *converges uniformly to* f*.*

We will not prove this result exactly, but instead a more general formulation of it due to Stone.

Example 5.5 If $S = [0, 1]$, the polynomials in Weierstrass' theorem can be explicitly written down. Define

$$b_{1,n}(t) = \binom{n}{i} t^i (1-t)^{n-1}, \quad p_n(t) = \sum_{k=0}^{n} f(k/n) b_{k,n}(t)$$

Then $||p_n - f||_\infty \to 0$.

One of our main results is Stone's theorem:

Theorem 5.2 (Stone 1941) *Suppose* X *is a compact metric space and* $\mathcal{A}$ *is a* **unital subalgebra** *of* $C(X)$*. Then* $\overline{A} = C(X)$ *if and only if* $\mathcal{A}$ *separates* X*.*

We first must prove several preliminary results:

Theorem 5.3 (Dini's Theorem) *Suppose* X *is a compact metric space and* f_n *is a sequence in* $C(X)$ *converging pointwise to* $f \in C(X)$*. Suppose further that* $f_n(x)$ *is a monotonically increasing sequence for each* $x \in X$*, so* $f_{n+1}(x) \geq f_n(x)$ *for all* $n \in \mathbb{N}$*. Then* f_n *converges uniformly to* f *on* X*.*

Proof Let $g_n(x) = f(x) - f_n(x) \geq 0$ for all $x \in X$. $g_n(x)$ is decreasing in n to 0. Fix $\varepsilon > 0$ and let $E_n = \{x \in X | g_n(x) < \varepsilon\} = g_n^{-1}((-\infty, \varepsilon))$, so E_n is continuous as g_n is continuous and $(-\infty, \varepsilon)$ is open. Note $E_n \subseteq E_{n+1}$ since $g_{n+1}(x) \leq g_n(x)$. Since $g_n(x) \to 0$, this implies that $X = \bigcup_{n=1}^{\infty} E_n$.

By compactness there exists a finite subcover $X = E_{n_1} \cup ... \cup E_{n_k} = E_N$ where $N = \max\{n_1, ..., n_k\}$. Then $X = E_N$ implies $g_N(x) < \varepsilon$ for all $x \in X$. But then for $n \geq N$, $0 \leq g_n(x) \leq g_N(x) < \varepsilon$. So $g_n \to_u 0$, so

$$||f - f_n||_\infty = ||g_n||_\infty \to 0$$

and $f_n \to_u f$ as desired. □

Lemma 5.1 *Suppose* X *is compact and* $\mathcal{A}$ *is a subalgebra of* $C(X)$*. If* $f \in \mathcal{A}$ *and* $0 \leq f(x) \leq 1$ *for all* $x \in X$*, then* $\sqrt{f} \in \overline{\mathcal{A}}$*.*

Proof Let $f_1 = 0 \in \mathcal{A}$ and $f_{n+1} = f_n + \frac{f - f_n^2}{2}$. By induction, $f_n \in \mathcal{A}$. We claim $0 \leq f_n(x) \leq \sqrt{f(x)}$ for all $x \in X$. If $n = 1$, the result is immediate. Suppose inductively that the result holds for some n. Then $f(x) - f_n(x)^2 \geq 0$ for all x, so

$$f_{n+1}(x) = f_n(x) + \frac{f(x) - f_n(x)^2}{2} \geq 0$$

We also have

$$\begin{aligned} f_{n+1}(x) &= f_n(x) + \frac{1}{2}(\sqrt{f(x)} - f_n(x))(\sqrt{f(x)} + f_n(x)) \\ &\leq f_n(x) + \frac{1}{2}(\sqrt{f(x)} - f_n(x))(\sqrt{f(x)} + \sqrt{f(x)}) \\ &\leq f_n(x) + \frac{1}{2}2(\sqrt{f(x)} - f_n(x)) = \sqrt{f(x)} \end{aligned}$$

We also get $f_{n+1}(x) \geq f_n(x)$ for all $x \in X$ since $f_{n+1}(x) - f_n(x) = \frac{f(x) - f_N(x)^2}{2} \geq 0$. By Dini's Theorem, $f_n \to_u g \in C(X)$ and $g(x) \leq \sqrt{f(x)}$. But then

$$f_{n+1} = f_n + \frac{f - f_n^2}{2} \implies g = g + \frac{f - g^2}{2} \implies g = \sqrt{f}$$

so $f_n \to_u \sqrt{f}$ and $\sqrt{f} \in \overline{\mathcal{A}}$. □

Definition 5.5 (Lattice) Suppose X is a compact metric space. A subspace L of $C(X)$ is called a ***lattice*** if $|f| \in L$ whenever $f \in L$.

Example 5.6 The lemma above shows that if $\mathcal{A}$ is a subalgebra of $C(X)$, then $\overline{\mathcal{A}}$ is a lattice. Let $f \neq 0$ in $\mathcal{A}$, and $g = \frac{f^2}{||f||_\infty^2}$ so that $0 \leq g \leq 1$. By the aboce result, $\sqrt{g} \in \overline{\mathcal{A}}$, but $\sqrt{g} = \frac{|f|}{||f||_\infty}$, so $|f|$ is in $\overline{\mathcal{A}}$.

If $f, g \in L$, a lattice, then

$$\max\{f(x), g(x)\} = \frac{f(x) + g(x) + |f(x) - g(x)|}{2} \in L$$

and

$$\min\{f(x), g(x)\} = \frac{f(x) + g(x) - |f(x) - g(x)|}{2} \in L$$

Lemma 5.2 (Key part of Stone's) *Suppose X is compact and $L \subseteq C(X)$ is a lattice. Suppose further that for each $x, y \in X$ and $a, b \in \mathbb{R}$, there is a function $f_{xy} \in L$ with $f_{xy}(x) = a$ and $f_{xy}(y) = b$. Then $\overline{L} = C(X)$.*

Proof Let $f \in C(X)$ and fix $\varepsilon > 0$. Let $x, y \in X$ and write $a := f(x)$ and $b := f(y)$. Find $f_{xy} \in L$ by assumption with $f_{xy}(x) = a$, $f_{xy}(y) = b$. Let

$$U_{xy} = \{z \in X | f_{xy}(z) < f(z) + \varepsilon\}$$

which is non-empty since $x, y \in U_{xy}$, and

$$V_{xy} = \{z \in X | f(z) - \varepsilon < f_{xy}(z)\}$$

which is also non-empty. Both are pre-images of open sets for continuous functions and are therefore open.

Fix $y \in X$, then $X = \bigcup_{x \in X} U_{xy}$ since $x \in U_{xy}$. By compactness, there exist $x_1, ..., x_n \in X$ with $X = \bigcup_{i=1}^{n} U_{x_i y}$. Let $h_y = \min\{f_{x_1 y}, f_{x_2 y}, ..., f_{x_n y}\} \in L$ as each $f_{xy} \in L$ by assumption and L is a lattice. Then, by definition, $h_y(z) < f(z) + \varepsilon$ for all $z \in X$. Also note that $f(z) - \varepsilon < h_y(z)$ for all $z \in \bigcap_{i=1}^{n} V_{x_i y}$, which is open. Let $V_y = \bigcap_{i=1}^{n} V_{x_i y}$, and note $y \in V_y$. Then $X = \bigcup_{y \in X} V_y$, so by compactness there exist $y_1, ..., y_m \in X$ with $X = \bigcup_{j=1}^{m} V_{y_j}$. Let $h = \max\{h_{y_1}, h_{y_2}, ..., h_{y_m}\}$. We already know $h(z) < f(z) + \varepsilon$ for all $z \in X$. Further, $h(z) > f(z) - \varepsilon$ for all $z \in X$, so

$$|h(z) - f(z)| < \varepsilon$$

for all z so $||h - f||_\infty \leq \varepsilon$. As $h \in L$ and $f \in C(X)$, and this holds for all $\varepsilon > 0$, $f \in \overline{L}$, and $\overline{L} = C(X)$. □

Proof (Stone's Theorem Proof) Let $\mathcal{A}$ be a unital subalgebra of $C(X)$ that separates points in X. We already know $\overline{\mathcal{A}}$ is a lattice. By the lemma above, we need to show for all $x \neq y \in X$ and $a, b \in \mathbb{R}$, there exists $f_{xy} \in \overline{\mathcal{A}}$ with $f_{xy}(x) = a$ and $f_{xy}(y) = b$. This will prove by the lemma that $\overline{\mathcal{A}} = \overline{\overline{\mathcal{A}}} = C(X)$. Assume $x \neq y$ and find $g \in A$ with $g(x) \neq g(y)$. Let

$$f_{xy}(t) = a\left(\frac{g(t) - g(y)}{g(x) - g(y)}\right) + b\left(\frac{g(t) - g(x)}{g(y) - g(x)}\right) \in \mathcal{A}$$

Then $f_{xy}(x) = a + 0 = a$ and $f_{xy}(y) = 0 + b = b$, as desired.

The converse is also true. If $\mathcal{A}$ is unital and $\overline{\mathcal{A}} = C(X)$, then $\mathcal{A}$ separates X. Suppose $\mathcal{A}$ does not seperate X. Then there exists $x \neq y$ in X with $f(x) = f(y)$ for all $f \in \mathcal{A}$. But then $g(x) = g(y)$ for all $g \in C(X)$ since univorm convergence implies pointwise convergence, and $\mathcal{A}$ is dense in $C(X)$. This is a contradiction, as all $C(X)$ separates X. □

Example 5.7 The polynomials $\mathcal{P}$ are a separating unital subalgebra for $C(K)$, $K \subseteq \mathbb{R}$ compact. So $\overline{\mathcal{P}} = C(K)$, i.e. Stone's theorem implies Weierstrass' polynomial approximation theorem.

Further, Ston'es theorem implies Weierstrass' theorem in the general case since the polynomials are an algebra separating any subset of points in $\mathbb{R}^n$.

Example 5.8 Take $\mathcal{A} = \text{span}\{f(x)g(y) | f, g \in C([0,1])\}$ is dense in $C([0,1] \times [0,1])$. Note something like $x^2 y + 1 + \frac{x}{y^2+1}$ is in $\mathcal{A}$, but $\sin^{-1}(x + y)$ is not, so it is a proper subspace. Not $\mathcal{A}$ is unital since we can just let $f(x) = 1 = g(y)$. $\mathcal{A}$ is a subalgebra as

$$\left(\sum_{i=1}^{n} f_i(x) g_i(y)\right)\left(\sum_{j=1}^{m} h_j(x) k_j(y)\right) = \sum_{i,j} f_i(x) h_j(x) g_i(y) k_j(y)$$

To confirm $\overline{\mathcal{A}} = C([0,1] \times [0,1])$, we show $\mathcal{A}$ separates points and then use Stone's theorem. Suppose $(x_0, y_0) \neq (x_1, y_1) \in [0,1]^2$. If $x_0 \neq x_1$, then $f(x, y) = x \in \mathcal{A}$ and $f(x_0, y_0) = x_0 \neq x_1 = f(x_1, y_1)$. If $y_0 \neq y_1$, we just use $f(x, y) = y$ instead.

Example 5.9 (Real Trig Polynomials) Let $\mathcal{A} = \text{span}_{m,n \in \mathbb{N} \cup \{0\}}\{\cos(nx), \sin(mx)\} \subseteq C([0,1])$. Note $1 \in \mathcal{A}$ setting $m = n = 0$. $\mathcal{A}$ is a span, and hence a subspace. $\mathcal{A}$ is an algebra using angle reduction formulas, such as $\cos^2(x) = \frac{1+\cos(2x)}{2}$. $\mathcal{A}$ separate $[0,\pi]$ since, for example, $f(x) = \cos(x)$ is injective on $[0,\pi]$. Thus, $\overline{\mathcal{A}} = C([0,\pi])$.

Stone's theorem also works for complex-valued functions, but we need the following additional assumption on $\mathcal{A}$; for every $f \in \mathcal{A}$ we also have $\overline{f} \in \mathcal{A}$.

Example 5.10 (Complex trig polynomials on the torus) Let $\prod = \{z \in \mathbb{C} | |z| = 1\}$ (the unit circle). Let $\mathcal{A} = \text{span}_{n\in\mathbb{Z}}\{e^{in\theta} | \theta \in [0,2\pi]\} \subseteq C(\prod, \mathbb{C})$. $\mathcal{A}$ is unital since $e^{i0\theta} = e^0 = 1 \in \mathcal{A}$. If $f \in \mathcal{A}$ then $\overline{f} \in \mathcal{A}$ since $\overline{e^{in\theta}} = e^{-in\theta} \in \mathcal{A}$. Just as in the previous example, $\mathcal{A}$ separates $\prod$, so $\overline{\mathcal{A}} = C(\prod, \mathbb{C})$.

Example 5.11 (Density of nowhere differentiable functions) Recall there exist $W \in C([0,1])$ that is nowhere differentiable. We know if $f \in C([0,1])$, there exist polynomials $p_n \to_u f$ by Stone-Weierstrass. Let $f_n := p_n + \frac{1}{n}W$, which are nowhere differentiable. Then

$$||f - f_n||_\infty \leq ||f - p_n||_\infty + \frac{1}{n}||W||_\infty \to 0$$

5.3 Equicontinuity and Arzéla-Ascoli

Definition 5.6 (Equicontinuity) Suppose X is a compact metric space and $\mathcal{F} \subseteq C(X)$. We say that $\mathcal{F}$ is ***equicontinuous on*** $\underline{X}$ if for every $\varepsilon > 0$ there is a $\delta > 0$ such that for any $x, y \in X$ satisfying $d_X(x,y) < \delta$, we have

$$|f(x) - f(y)| < \varepsilon$$

for every $f \in \mathcal{F}$.

Note δ only depends on ε. This can be thought of as simultaneous uniform continuity.

Remark 5.1 Think of:

- Separating X corresponds to “many function"
- Equicontinuous corresponds to “few functions"

Definition 5.7 In general, for metric spaces X and Y, $\mathcal{F} \subseteq C(X,Y)$ is equicontinuous if for every $\varepsilon > 0$ there exists $\delta > 0$ such that if $f \in \mathcal{F}$ then $d_Y(f(x), f(z)) < \varepsilon$ for all $x, z \in X$ such that $d_X(x,z) < \delta$.

Proposition 5.2 *Let X and Y be metric spaces with X compact, and equip $C(X,Y)$ with the L_∞-norm. That is, the distance d in $C(X,Y)$ is*

$$d(f,g) = \sup_{x\in X} d_Y(f(x), g(x))$$

Then if $\mathcal{F} \subseteq C(X,Y)$ is compact, it is equicontinuous.

Proof Suppose $\mathcal{F} \subseteq C(X,Y)$ is compact. Fix $\varepsilon > 0$. As X is compact, each $f \in C(X,Y)$ is uniformly continuous, so there exists $\delta_f > 0$ such that $d_Y(f(x), f(y)) < \varepsilon/3$ for all $x, y \in X$ such that $d_X(x,y) < \delta_f$. Then $\mathcal{F} \subseteq \bigcup_{f\in\mathcal{F}} B_{\varepsilon/3}(f)$ is an open cover, so as $\mathcal{F}$ is compact there exist $f_1, ..., f_N$ such that $\mathcal{F} \subseteq \bigcup_{j=1}^{N} B_{\varepsilon/3}(f_j)$. Then, let $\delta = \min_{1\leq j\leq N}\{\delta_{f_j}\}$. Now let $f \in \mathcal{F}$ and let $x, y \in X$ with $d_X(x,y) < \delta$. Then there exists j such that $f \in B_{\varepsilon/3}(f_j)$. It follows that

$$d_Y(f(x), f(y) \leq d_Y(f(x), f_j(x)) + d_Y(f_j(x), f_j(y)) + d_Y(f_j(y), d(y) < \varepsilon/3 + \varepsilon/3 + \varepsilon/3 = \varepsilon$$

so indeed $\mathcal{F}$ is equicontinuous. □

Example 5.12 (Uniformly converging sequences are equicontinuous) If $\mathcal{F} = \{f\}$ is a singleton, then $\{f\}$ is equicontinuous since f is uniformly continuous. If $\mathcal{F} = \{f_1, ..., f_n\} \subseteq C(X)$ then it is equicontinuous. Fix $\varepsilon > 0$. For each f_i, there exists $\delta_i > 0$ from uniform continuity of f_i. Then $\delta = \min\{\delta_1, ..., \delta_n\}$ works for the definition of uniform continuity. Equivalently, we could recognize that every finite subset of a topological space is compact, and use the previous result.

Suppose $f_n \to_u f$ in $C(X)$. Then $\mathcal{F} = \{f_n\}_{n\in\mathbb{N}}$ is equicontinuous:

$$|f_n(x) - f_n(y)| \leq |f_n(x) - f(x)| + |f(x) - f(y)| + |f(y) - f_n(y)|$$

f is uniformly continous so there exists $\delta > 0$ such that $d_X(x,y) < \delta$ implies $|f(x) - f(y)| < \varepsilon/3$. By uniform convergence there exist $N \in \mathbb{N}$ such that $n \geq N$ implies $|f(x) - f_n(x)| < \varepsilon/3$ for all $x \in X$. For $d_X(x,y) < \delta$, we obtain $|f_n(x) - f_n(y)| < \varepsilon$ (then $k < N$ is a finite set and we can use the previous part for that).

Example 5.13 Totally bounded subsets of $C(X)$ are equicontinuous. Recall that compact is equivalent to complete and totally bounded in metric spaces. Suppose $\mathcal{F} \subseteq C(X)$ is totally bounded. Then if $\varepsilon > 0$, there exist $f_1, ..., f_n$ such that $\mathcal{F} \subseteq \bigcup_{i=1}^{n} B_{\varepsilon/3}(f_i)$. If $f \in \mathcal{F}$, find i such that $||f - f_i||_\infty < \varepsilon/3$. Then

$$|f(x) - f(y)| < 2\varepsilon/3 + |f_i(x) - f_i(y)|$$

As $\{f_1, ..., f_n\}$ is equicontinuous, there exists $\delta > 0$ such that $d_X(x,y) < \delta$ implies $|f_i(x) - f_i(y)| < \varepsilon/3$, and the result follows.

Note this result holds in general for totally bounded subsets of $C(X,Y)$ for X compact.

Example 5.14 Let α and K be fixed positive constants. The set

$$\mathcal{F} = \{f \in C([0,1])|\ |f(x) - f(y)| \leq L(x-y|^\alpha \text{ for all } x, y \in [0,1]\}$$

is equicontinuous.

Fix $\varepsilon > 0$. Let $\delta = (\varepsilon/K)^{1/\alpha}$. If $|x - y| < \delta$,

$$|f(x) - f(y)| < K((\varepsilon/K)^{1/\alpha})^\alpha = \varepsilon$$

for all $f \in \mathcal{F}$.

Let $G = \{f \in C([0,1]) | f'$ is continuous on $[0,1]$ and $||f'||_\infty \le 1\}$. G is equicontinuous since if $x < y$ there exists $c \in (x, y)$ with

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

implies

$$|f(y) - f(x)| = |f'(c)| \, |y - x| \le |y - x|$$

So $G \subseteq \mathcal{F}$ with $\alpha = K = 1$.

Example 5.15 Let

$$\mathcal{F} = \left\{ F(x) = \int_0^x f(t)dt | f \in C([0,1]) \text{ and } ||f||_\infty \le 1 \right\} \subseteq C([0,1])$$

is equicontinuous and not closed. $F(0) = 0$ for all $F \in \mathcal{F}$. If $F = \int_0^x f$, then $F'(x) = f(x)$ by the FTOC1. Then

$$|F(x) - F(y)| = \left| \int_y^x f(t)dt \right| \le \left| \int_y^x |f(t)|dt \right| \le \left| \int_y^x 1dt \right| = |y - x|$$

So $\mathcal{F}$ is a subset of the $\mathcal{F}$ in the previous example with $\alpha = K = 1$.

$\mathcal{F}$ is not closed in $C([0,1])$, as

$$f_n(x) = \sqrt{(x - 1/2)^2 + 1/n} - \sqrt{1/4 + 1/n}$$

has $f_n(0) = 0$, and $|f_n'(x)| = \frac{|x-1/2|}{\sqrt{(x-1/2)^2+1/n}} \le 1$ and $f_n' \in C([0,1])$, so

$$f_n(x) = \int_0^x f_n'(t)dt \in \mathcal{F}$$

by FTOC2. But $f_n \to |x - 1/2| - 1/2$ which is not differentiable, and hence not in $\mathcal{F}$.

Theorem 5.4 (Arzéla 1884; Ascoli 1895) *Let X be a compact metric space. A set $\mathcal{F} \subseteq C(X)$ is compact in $C(X)$ if and only if $\mathcal{F}$ is closed, bounded, and equicontinuous.*

Proof We saw $\mathcal{F}$ compact implies $\mathcal{F}$ equicontinuous. We also already know compact implies closed and bounded in metric spaces. Conversely, let $(f_n) \subseteq \mathcal{F}$ with $\mathcal{F}$ equicontinuous. We will show (f_n) admits a uniformly convergent subsequence.

Note since X is compact, it is totally bounded. Then for each $n \in \mathbb{N}$, there exists $x_{n(1)}, ..., x_{n(k)}$ with $X \subseteq \bigcup_{j=1}^k B_{1/n}(x_{n(j)})$. Then the countable set $\bigcup_{n \in \mathbb{N}} \{x_{n(j)}\}$ is dense in X.

Let $\{x_1, x_2, ...\}$ be a countable dense subset in X. The sequence $\{f_n(x_1)\}_{n=1}^\infty$ is bounded in $\mathbb{R}$ since $\mathcal{F}$ is assumed bounded, so there exists $C > 0$ such that $||f||_\infty \le C$ for all $f \in \mathcal{F}$. By Bolzano-Weierstrass, $\{f_n(x_1)\}$ has a convergent subsequence $\{f_{s_1(n)}(x_1)\}$. Then $\{f_{s_2(n)}(x_2)\}$ is a bounded sequence in $\mathbb{R}$, and also has a convergent subsequence $\{f_{s_2(n)}(x_2)\}$. Inductively define $\{f_{s_k(n)}\} \subseteq \{f_{s_{k-1}(n)}\} \subseteq ... \subseteq \{f_n\} \subseteq \mathcal{F}$, with $\{f_{s_k(n)}(x_k)\}$ convergent. Define the subsequence

$\xi_\nu = f_{s_\nu(\nu)}$. By construction, $\xi_\nu(x_i)$ converges for all x_i. Fix $\varepsilon > 0$ and find $\delta > 0$ by equicontinuity such that $d_X(x, y) < \delta$ implies $|\xi_n(x) - \xi_n(y)| < \varepsilon$ for all $n \in \mathbb{N}$. We want to establish the Cauchy criterion for ξ_ν. Then

$$|\xi_n(x) - \xi_m(x)| \le |\xi_n(x) - \xi_n(x_i)| + |\xi_n(x_i) - \xi_m(x_i)| + |\xi_m(x_i) - \xi_m(x)|$$

We can find x_i such that $x \in B_j(x_i)$ by density, so using equicontinuity and the fact ξ_n is cauchy on the countable dense subset, we obtain an $\varepsilon/3$ argument yielding our desired result. □

We can extend this result, without any difficulty, to subsets of $C(X, \mathbb{R}^n)$ for $n \in \mathbb{N}$.

Corollary 5.1 $\mathcal{F} \subseteq C(X)$ *is totally bounded if and only if* $\mathcal{F}$ *is bounded and equicontinuous.*

Proof $\mathcal{F}$ is totally bounded if and only if $\overline{\mathcal{F}}$ is totally bounded. The if direction is immediate, and for the only if direction, fix $\varepsilon > 0$. Then there exist $f_1, ..., f_n$ such that $\mathcal{F} \subseteq \bigcup_{i=1}^n B_\varepsilon(f_i)$, then

$$\overline{\mathcal{F}} \subseteq \overline{\bigcup_{i=1}^n B_\varepsilon(f_i)} \subseteq \bigcup_{i=1}^n \overline{B_\varepsilon(f_i)} \subseteq \bigcup_{i=1}^n B_{2\varepsilon}(f_i)$$

Then, $\overline{\mathcal{F}}$ is totally bounded if and only if it $\overline{\mathcal{F}}$ is complete and totally bounded since $C(X)$ is complete and $\overline{\mathcal{F}}$ is closed, which occurs if and only if $\mathcal{F}$ is complete, if and only if $\overline{\mathcal{F}}$ is closed, bounded, and equicontinuous by Arzéla-Ascoli, which holds if and only if $\mathcal{F}$ is bounded and equicontinuous. □

Example 5.16 Closed balls in $C([0, 1])$ are not equicontinuous. Consider $\overline{B}_r(f_0)$, which contains the sequence $f_0 + \frac{1}{2r}x^n$, which we know has no uniformly convergent subsequence, and is not compact and hence is not equicontinuous by Arzéla-Ascoli.

Chapter 6

Higher-Dimensional Differentiation

6.1 Basic Notions of Higher-Dimensional Derivatives

In this section, we will always use the 2-norm on $\mathbb{R}^n$ and denote it $||\cdot||$. For a linear operator or matrix, $||\cdot||$ will also denote the operator or spectral norm. U will denote an open subset of $\mathbb{R}^n$.

Definition 6.1 (The Derivative) Let $f : U \to \mathbb{R}^m$. We say that f is ***differentiable*** at $a \in U$ if there exists a linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ such that

$$\lim_{h\to 0} \frac{||f(a+h) - f(a) - Th||_{\mathbb{R}^m}}{||h||_{\mathbb{R}^n}} = 0$$

for $a + h \in U$, where as U is open $a + h \in U$ for h small enough. This can be interpreted to mean $f(a+h) \approx f(a)+Th$ for h small enough. We call T the ***derivative of f at a*** and write it as $T = Df(a)$. We say that f is ***differentiable on U*** if f is differentiable at each $a \in U$.

Remark 6.1 The linear transformation T in the above definition is unique. Suppose S and T both satisfy the definition of th ederivative at a. Then

$$\frac{||Th - Sh||}{||h||} \leq \frac{||f(a+h) - f(a) - Th||}{||h||} + \frac{||f(a+h) - f(a) - Sh||}{||h||} \to 0$$

Indeed, for any $\alpha > 0$ constant, and $1 \leq j \leq n$,

$$0 = \lim_{\alpha\to 0} \frac{||T\alpha e_j - S\alpha e_j||}{||\alpha e_j||} = \lim_{\alpha\to 0} ||Te_j - Se_j||$$

so $Te_j = Se_j$, and hence $T = S$.

If the case of $m = n = 1$, if $f : U \to \mathbb{R}$ is a function with $U \subseteq \mathbb{R}$ open, the linear transformation T is just the constant $f'(a)$, where as $f'(a) = \lim\limits_{h\to 0} \frac{f(a+h)-f(a)}{h}$,

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

and in this case $f(a+h) \approx f(a) + f'(a)h$.

Note 6.1 Given a function $f : U \to \mathbb{R}^m$, we will write f as $f(x) = (f_1(x), ..., f_m(x))^T$, where each $f_i : U \to \mathbb{R}$ and $f_i = p_i \circ f$, for $p_i : \mathbb{R}^m \to \mathbb{R}$ the i-th projection function.

Proposition 6.1 *If $f : U \to \mathbb{R}^m$ is differentiable at $a \in U$, then f is continuous at a.*

Proof We must show that $\lim_{h\to 0} f(a+h) = f(a)$. Find $\delta > 0$ such that $||h|| < \delta$ implies

$$\frac{||f(a+h) - f(a) - Df(a)h||}{||h||} < 1$$

Then $||f(a+h) - f(a) - Df(a)h|| \leq ||h||$, so by the triangle inequality, and using the spectral norm,

$$||f(a+h) - f(a)|| \leq ||Df(a)h|| + ||h|| \leq ||Df(a)||\ ||h|| + ||h|| \to 0$$

as $h \to 0$, so f is continuous at a. □

Definition 6.2 (Directional Derivative) Let $f : U \to \mathbb{R}^m$ and let $v \in \mathbb{R}^n$ be a unit vector. The ***directional derivative for f at $a \in U$ in the direction of v*** is given by

$$D_v f(a) := \lim_{t\to 0} \frac{f(a+tv) - f(a)}{t} \in \mathbb{R}^m$$

for all $a \in U$ and t small enough such that $a + tv \in U$, provided the limit above exists. If $v = e_i$, the canonical orthonormal basis element for $\mathbb{R}^n$, we call $D_{e_i} f(a)$ the ***partial derivative of f with respect to x_i at a*** and denote it $D_i f(a)$ or $\frac{\partial f}{\partial x_i}(a)$.

Note that the limit for $D_v f(a)$ can be written as

$$D_v f(a) = \lim_{t\to 0} \begin{pmatrix} \frac{f_1(a+tv) - f_1(a)}{t} \\ \vdots \\ \frac{f_m(a+tv) - f_m(a)}{t} \end{pmatrix} = \begin{pmatrix} D_v f_1(a) \\ \vdots \\ D_v f_n(a) \end{pmatrix}$$

if the $D_v f_i(a)$ existm as we know that sequences in $\mathbb{R}^m$ converge if and only if each component sequence converges. Thus, $D_v f(a)$ exists if and only if $D_v f_i(a)$ exists for each $i = 1, ..., m$.

Example 6.1 (Affine maps are differentiable) Let $f : \mathbb{R}^n \to \mathbb{R}^m$ be defined by $f(x) = Ax + b$ for A $m \times n$ and $b \in \mathbb{R}^m$. Then $Df(x) = A$ for all $x \in \mathbb{R}^n$, since for any $h \in \mathbb{R}^n$

$$||f(a+h) - f(a) - Ah|| = ||Aa + Ah - Aa - Ah|| = 0$$

Example 6.2 The function $f : \mathbb{R}^2 \to \mathbb{R}^2$ given by $f(x,y) = (x^2, y^2)^T$ is differentiable. We claim $Df(x,y) = \begin{bmatrix} 2x & 0 \\ 0 & 2y \end{bmatrix}$. To see this, let $h = [h_1\ h_2]^T$. Since all norms on $\mathbb{R}^n$ and $\mathbb{R}^m$ are equivalent, we can use whichever norm makes the easiest computation:

$$\begin{aligned}
\lim_{h\to 0}\frac{||\langle (x+h_1)^2-x^2,(y+h_2)^2-y^2\rangle-\langle 2xh_1,2yh_2\rangle||_1}{||h||_1} &= \lim_{h\to 0}\frac{|(x+h_1)^2-x^2-2xh_1|+|(y+h_2)^2-y^2-2yh_2|}{|h_1|+|h_2|}\\
&= \lim_{h\to 0}\frac{h_1^2+h_2^2}{|h_1|+|h_2|}\\
&\le \lim_{h\to 0}\frac{h_1^2}{|h_1|}+\frac{h_2^2}{|h_2|}=0
\end{aligned}$$

Proposition 6.2 *If $f : U \to \mathbb{R}^m$ and $g : U \to \mathbb{R}$ are differentiable at $a \in U$, then so is fg and*

$$D(fg)(a) = f(a)Dg(a) + g(a)Df(a)$$

Proof First, note that $fg : U \to \mathbb{R}^m$. Then observe that for all $||h|| > 0$, with $a + h \in U$,

$$\begin{aligned}
\frac{||fg(a+h)-fg(a)-f(a)Dg(a)h-g(a)Df(a)h||}{||h||} &= \frac{\begin{matrix}|||f(a+h)g(a+h)-f(a)g(a+h)+f(a)g(a+h)\\ -f(a)g(a)-f(a)Dg(a)h-g(a)Df(a)h||\end{matrix}}{||h||}\\
&\le \frac{||f(a+h)g(a+h)-f(a)g(a+h)-g(a)Df(a)h||}{||h||}\\
&+\frac{||f(a)g(a+h)-f(a)g(a)-f(a)Dg(a)h||}{||h||}\\
&= \frac{\begin{matrix}||f(a+h)g(a+h)-f(a)g(a+h)\\ -g(a+h)Df(a)h+(g(a+h)-g(a))Df(a)h||\end{matrix}}{||h||}\\
&+\frac{||f(a)g(a+h)-f(a)g(a)-f(a)Dg(a)h||}{||h||}\\
&\le \frac{|g(a+h)|\ ||f(a+h)-f(a)-Df(a)h||}{||h||}\\
&+\frac{||Df(a)h||\ |g(a+h)-g(a)|}{||h||}\\
&+\frac{||f(a)(g(a+h)-g(a)-Dg(a)h)||}{||h||}\\
&\le \frac{|g(a+h)|\ ||f(a+h)-f(a)-Df(a)h||}{||h||}\\
&+\frac{||Df(a)||\ ||h||\ |g(a+h)-g(a)|}{||h||}\\
&+\frac{||f(a)||\ |g(a+h)-f(a)g(a)-f(a)Dg(a)h|}{||h||}\\
&= \frac{|g(a+h)|\ ||f(a+h)-f(a)-Df(a)h||}{||h||}\\
&+||Df(a)||\ |g(a+h)-g(a)|\\
&+\frac{||f(a)||\ |g(a+h)-f(a)g(a)-f(a)Dg(a)h|}{||h||}
\end{aligned}$$

Then, note that each term in the above expansion goes to zero since g being differentiable implies it is continuous. □

Proposition 6.3 *Suppose $f : U \to \mathbb{R}^m$ is differentiable at $a \in U$. Then $D_v f(a)$ exists for all unit vectors v. Moreover,*

$$D_v f(a) = Df(a)v$$

(note $D_v f(a) \in \mathbb{R}^m$, $Df(a) \in \mathbb{R}^{m\times n}$, and $v \in \mathbb{R}^n$)

Proof We need to show $\lim\limits_{t\to 0} \frac{f(a+tv)-f(a)}{t}$ exists. In the definition of $Df(a)$, let $h = tv$. Then $||h|| = |t|$, so $h \to 0$ if and only if $t \to 0$, since v is assumed to be a unit vector. Then

$$\begin{aligned} 0 &= \lim_{t\to 0} \frac{||f(a+tv) - f(a) - Df(a)tv||}{|t|} \\ &= \lim_{t\to 0} \left\| \frac{f(a+tv) - f(a)}{t} - \frac{Df(a)tv}{t} \right\| \\ &= \lim_{t\to 0} \left\| \frac{f(a+tv) - f(a)}{t} - Df(a)v \right\| \end{aligned}$$

Thus, the limit exists and $\lim\limits_{t\to 0} \frac{f(a+tv)-f(a)}{t} = Df(a)v$. □

Example 6.3 If $U \subseteq \mathbb{R}^n$ and $f : U \to \mathbb{R}$, then Df is a $1 \times n$ row (called the gradient), so $Df(a)v = Df(a) \cdot v$ for $v \in \mathbb{R}^n$.

If $\{e_1, ..., e_n\}$ is the canonical orthonormal basis for $\mathbb{R}^n$, then Ae_i is the ith column of A. The ith partial derivative for f at a is $Df(a)e_i = D_i f(a) =$ the ith column of $Df(a)$, so $Df(a) = [D_1 f(a)\ D_2 f(a)\ \dots\ D_n f(a)]$. Now, if $f = (f_1, ..., f_m)^T$, then by the chain rule, which we shall prove shortly, $Df(a)$ exists if and only if $Df_i(a)$ exists for all $i = 1, ..., m$ and

$$Df(a) = \begin{bmatrix} Df_1(a) \\ \vdots \\ Df_m(a) \end{bmatrix}$$

Together we get $Df(a) = [D_i f_j(a)]$ for $i = 1, ..., n$, $j = 1, ..., m$, provided $Df(a)$ exists.

Example 6.4 Consider

$$f(x,y) = \begin{cases} 0 & (x,y) = (0,0) \\ \frac{xy}{x^2+y^2} & (x,y) \neq (0,0) \end{cases}$$

which is not continuous, and hence not differentiable, at $(0,0)$. Indeed,

$$f(1/n, 1/n) = \frac{1/n^2}{2/n^2} = \frac{1}{2} \not\to 0 = f(0,0)$$

Then, let $v = (v_1, v_2)^T$ be a unit vector, so $v_1^2 + v_2^2 = 1$. Then

$$D_v f(0,0) = \lim_{t\to 0} \frac{f(tv_1, tv_2) - f(0,0)}{t}$$

$$
\begin{aligned}
&= \lim_{t\to 0} \frac{tv_1v_2}{t^2v_1^2 + t^2v_2^2} \\
&= \lim_{t\to 0} \frac{v_1v_2}{t}
\end{aligned}
$$

which does not exist for any unit vector v unless $v_1 = \pm 1$ and $v_2 = 0$, or $v_1 = 0$ and $v_2 = \pm 1$. So $D_1 f(0,0) = D_2 f(0,0) = 0$, but no other directionals exist!

Example 6.5 Consider

$$
f(x,y) = \begin{cases} 0 & (x,y) = (0,0) \\ \frac{x^3}{x^2+y^2} & (x,y) \neq (0,0) \end{cases}
$$

which is continuous at $(0,0)$, all directional derivatives exist at $(0,0)$, but f is not differentiable at $(0,0)$. For continuity

$$
\left|\frac{x^3}{x^2+y^2}\right| = \frac{|x|x^2}{x^2+y^2} \leq \frac{\sqrt{x^2+y^2}}{x^2+y^2}x^2 \leq \frac{\sqrt{x^2+y^2}}{x^2+y^2}(x^2+y^2) = \sqrt{x^2+y^2} \to 0
$$

as $(x,y) \to 0$. For directionals, let $v = (v_1, v_2)$ be a unit vector. Then $\lim_{t\to 0} \frac{t^2v_1^3}{t^2v_1^2+t^2v_2^2} = \lim_{t\to 0} v_1^3 = v_1^3$ So $D_v f(0,0)$ exists for all v. In particular, $D_1 f(0,0) = 1^3 = 1$ and $D_2 f(0,0) = 0$. We know that if $Df(0,0)$ exists, it must be $[1\ 0]$. Then

$$
\begin{aligned}
\lim_{h\to 0} \frac{|f(h_1,h_2) - f(0,0) - [1\ 0](h_1,h_2)^T|}{||h||} &= \lim_{h\to 0} \frac{\left|\frac{h_1^3}{h_1^2+h_2^2} - h_1\right|}{\sqrt{h_1^2+h_2^2}} \\
&= \lim_{h\to 0} \frac{|h_1h_2^2|}{(h_1^2+h_2^2)^{3/2}}
\end{aligned}
$$

which does not tend to 0 since the sequence $(1/n, 1/n)^T$ yields

$$
\lim_{n\to\infty} \frac{1/n^3}{(2/n^2)^{3/2}} = \frac{1}{2^{3/2}} \neq 0
$$

Example 6.6 (Projection maps are Differentiable) Define $\pi_i : \mathbb{R}^n \to \mathbb{R}$ by $\pi_i(x_1, ..., x_n) = x_i$. π_i is linear, given by the matrix $\pi_i(x) = [0 \dots 0\ 1\ 0 \dots 0]x$, for 1 in the ith position. Thus, as it is linear it is its own derivative at any point $a \in \mathbb{R}^n$.

Proposition 6.4 (Chain Rule) *Suppose $U \subseteq \mathbb{R}^n$ is open, that $f : U \to \mathbb{R}^m$ is differentiable on U, and that $g : V \to \mathbb{R}^p$ is differentiable on an open set $V \subseteq \mathbb{R}^m$ containing $f(U)$. Then $g \circ f$ is differentiable on U, and*

$$
D(g \circ f)(a) = Dg(f(a))Df(a)
$$

where $Dg(f(a))$ is a $p \times m$ matrix and $Df(a)$ is an $m \times n$ matrix, and $D(g \circ f)(a)$ is $p \times n$.

Proof Let $b = f(a), L = Df(a), M = Dg(b)$, and define

$$
\varphi(x) = f(x) - f(a) - L(x-a),\ \varphi : U \to \mathbb{R}^m
$$

$$\psi(y) = g(y) - g(b) - M(y-b),\ \psi : V \to \mathbb{R}^p$$
$$p(x) = g(f(x)) - g(f(a)) - ML(x-a),\ p : U \to \mathbb{R}^p$$

By differentiability of f and g,

$$\lim_{x\to a} \frac{||\varphi(x)||}{||x-a||} = 0, \text{ and } \lim_{y\to b} \frac{||\psi(y)||}{||y-b||} = 0$$

We have $p(x) = \psi(f(x)) + M\varphi(x)$ and we want to show

$$\lim_{x\to a} \frac{||p(x)||}{||x-a||} = 0$$

Compute

$$\frac{||p(x)||}{||x-a||} \le \frac{||\psi(f(x))||}{||x-a||} + \frac{||M\varphi(x)||}{||x-a||} \le \frac{||\psi(f(x))||}{||x-a||} + \frac{||M||\ ||\varphi(x)||}{||x-a||}$$

where the second term goes to zero since $||\varphi(x)||/||x-a|| \to 0$. Fix $\varepsilon > 0$ and find $\delta > 0$ such that $||x-a|| < \delta$ implies $||\psi(f(x))|| < \varepsilon||f(x)-b||$. Rearrange φ to get $f(x) - b = \varphi(x) + L(x-a)$, so $||f(x)-b|| \le ||\varphi(x)|| + ||L||\ ||x-a||$. Then for $||x-a|| < \delta$,

$$||\psi(f(x))|| < \varepsilon(||\varphi(x)|| + ||L||\ ||x-a||)$$

so

$$\frac{||\psi(f(x))||}{||x-a||} \le \varepsilon\left(\frac{||\varphi(x)||}{||x-a||} + ||L||\right)$$

Since $\frac{||\varphi(x)||}{||x-a||} \to 0$, and $||L|| < \infty$ is a number, we can make the right hand side as small as we wish, so $\frac{||\psi(f(x))||}{||x-a||} \to 0$. Thus, $\lim\limits_{x\to a} \frac{||p(x)||}{||x-a||} = 0$ as desired, giving $D(g\circ f)(a) = Dg(f(a))Df(a)$. □

Example 6.7 Let $g : \mathbb{R} \to \mathbb{R}^n$ be differentiable with $g(t) = \begin{pmatrix} g_1(t) \\ \vdots \\ g_n(t) \end{pmatrix}$, a parametric curve, and $f : \mathbb{R}^n \to \mathbb{R}$ also differentiable. Then $Dg(t) = \begin{pmatrix} g_1'(t) \\ \vdots \\ g_n'(t) \end{pmatrix}$, and $n \times 1$ column, and $Df(x_1, ..., x_n) = [D_1 f(x_1, ..., x_n)\ \cdots\ D_n f(x_1, ..., x_n)]$, a $1\times n$ row. Then $f\circ g : \mathbb{R} \to \mathbb{R}$, with $f(g_1(t), ..., g_n(t)$, and

$$D(f\circ g)(t) = Df(g(t))Dg(t) = [D_1 f(g(t))\ \cdots\ D_n f(g(t))] \begin{bmatrix} g_1'(t) \\ \vdots \\ g_n'(t) \end{bmatrix} = \sum_{i=1}^{n} D_i f(g(t)) g_i'(t)$$

Proposition 6.5 $f = (f-1, ..., f_m)^T : U \to \mathbb{R}^m$ *is differentiable at* $a \in U$ *if and only if each* f_i *is differentiable at* a.

Proof Recall $\pi_i : \mathbb{R}^m \to \mathbb{R}$ with $\pi_i(x_1, ..., x_m) = x_i$, which is a linear transformation, and so differentiable. If f is differentiable then so is $\pi_i \circ f = f_i$ by the chain rule, and

$$Df_i(x) = D\pi_i(f(x))Df(x) = \pi_i Df(x)$$

which is the ith row of $Df(x)$. That is,

$$Df(x) = \begin{bmatrix} Df_1(x) \\ \vdots \\ Df_m(x) \end{bmatrix}$$

Conversely, if each f_i is differentiable, write T to be the linear transformation just found, but at a. Then

$$\begin{aligned} \lim_{h\to 0} \frac{||f(a+h) - f(a) - Th||_1}{||h||_1} &= \lim_{h\to 0} \frac{||f(a+h) - f(a) - [Df_1(a)h \ \dots \ Df_m(a)h]^T\,||_1}{||h||_1} \\ &= \lim_{h\to 0} \frac{\sum_{i=1}^m |f_i(a+h) - f_i(a) - Df_i(a)h|}{||h||_1} \\ &= \lim_{h\to 0} \sum_{i=1}^m \frac{|f_i(a+h) - f_i(a) - Df_i(a)h|}{||h||_1} \end{aligned}$$

which goes to zero by the differentiability of the f_i. □

Example 6.8 Consider

$$f(x,y) = \begin{cases} 0 & (x,y) = (0,0) \\ (x^2+y^2)\sin\left(\frac{1}{\sqrt{x^2+y^2}}\right) & (x,y) \neq (0,0) \end{cases}$$

which is differentiable at $(0,0)$, but both partials are not continuous at $(0,0)$. For differentiability at $(0,0)$, we claim $Df(0,0) = [0\ 0]$ since

$$\begin{aligned} \lim_{h\to 0} \frac{|f(h_1,h_2) - f(0,0) - [0\ 0](h_1,h_2)^T\,|}{||h||} &= \lim_{h\to 0} \frac{(h_1^2+h_2^2)\sin\left(\frac{1}{\sqrt{h_1^2+h_2^2}}\right)}{\sqrt{h_1^2+h_2^2}} \\ &= \lim_{h\to 0} \sqrt{h_1^2+h_2^2}\sin\left(\frac{1}{\sqrt{h_1^2+h_2^2}}\right) = 0 \end{aligned}$$

since sin is bounded. Then $\frac{\partial f}{\partial x}(0,0) = 0$ and $\frac{\partial f}{\partial y}(0,0) = 0$. But

$$\frac{\partial f}{\partial x}(x,y) = \begin{cases} (x^2+y^2)\cos\left(\frac{1}{\sqrt{x^2+y^2}}\right)(x^2+y^2)^{-3/2}\cdot\left(\frac{-1}{2}\right)\cdot 2x + 2x\sin\left(\frac{1}{\sqrt{x^2+y^2}}\right) & (x,y)\neq(0,0) \\ 0 & (x,y) = (0,0) \end{cases}$$

which has an infinite discontinuity at $(0,0)$, and similarly for $\frac{\partial f}{\partial x}(x,y)$.

Example 6.9 Consider

$$f(x,y)=\begin{cases} 0 & (x,y)=(0,0) \\ \frac{x^2y}{x^4+y^2} & (x,y)\neq(0,0) \end{cases}$$

which is not continuous at $(0,0)$, but every directional derivative exists at $(0,0)$. For continuity

$$f(1/n,1/n^2)=\frac{1/n^4}{1/n^4+1/n^4}=\frac{1}{2}\not\to 0=f(0,0)$$

so f is not continuous at $(0,0)$. Let $u=[u_1\ u_2]^T$ be a unit vector. Then

$$\begin{aligned} D_uf(0,0) &= \lim_{t\to 0}\frac{f(tu_1,tu_2)-f(0,0)}{t} \\ &= \lim_{t\to 0}\frac{t^2u_1^2u_2}{t^4u_1^4+t^2u_2^2} \\ &= \lim_{t\to 0}\frac{u_1^2u_2^2}{t^2u_1^4+u_2^2}=\frac{u_1^2u_2}{u_2^2}=\frac{u_1^2}{u_2} \end{aligned}$$

provided $u_2\neq 0$. But it is 0 if $u_2=0$, $u_1=\pm 1$.

Example 6.10 Consider

$$f(x,y)=\begin{cases} 0 & (x,y)=(0,0) \\ \frac{x^2y\sqrt{x^2+y^2}}{x^4+y^2} & (x,y)\neq(0,0) \end{cases}$$

which is continuous at $(0,0)$, all directional derivatives exist at $(0,0)$, but $Df(0,0)$ does not exist. For continuity

$$\frac{|x^2y|\sqrt{x^2+y^2}}{x^4+y^2}\leq\frac{\frac{1}{2}(x^4+y^2)\sqrt{x^2+y^2}}{x^4+y^2}\to 0$$

since $(x^2\pm y)^2\geq 0$ implies $x^4\pm 2x^2y+y^2\geq 0$, and so $\frac{x^4+y^2}{2}\geq|x^2y|$. For the directionals,

$$\begin{aligned} D_uf(0,0) &= \lim_{t\to 0}\frac{f(tu_1,tu_2)-f(0,0)}{t} \\ &= \lim_{t\to 0}\frac{t^2u_1^2u_2\sqrt{t^2u_1^2+t^2u_2^2}}{t^4u_1^4+t^2u_2^2} \\ &= \lim_{t\to 0}\frac{tu_1^2u_2^2}{t^2u_1^4+u_2^2}=0 \end{aligned}$$

so $D_1f(0,0)=0=D_2f(0,0)$. If $Df(0,0)$ exists, it must be $[0\ 0]$. But then

$$0=\lim_{h\to 0}\frac{|f(h_1,h_2)-f(0,0)-[0\ 0](h_1,h_2)^T|}{\sqrt{h_1^2+h_2^2}}=\lim_{h\to 0}\frac{h_1^2h_2}{h_1^4+h_2^2}$$

which does not hold since for $h_1=1/n$, $h_2=1/n^2$, the limit is

$$\lim_{h\to 0}\frac{1/n^4}{1/n^4+1/n^4}=\frac{1}{2}$$

Example 6.11 Space curves on the unit sphere are orthogonal to their derivative. Consider $f : \mathbb{R} \to \mathbb{R}^3$ differentiable. Write $f(t) = (f_1(t), f_2(t), f_3(t))^T$, so that $f'(t) = (f_1'(t), f_2'(t), f_3'(t))^T$. Further suppose $||f(t)|| = 1$ for all t (that is the curve is on the unit sphere). Then $f_1(t)^2 + f_2(t)^2 + f_3(t)^2 = 1$, and so we have

$$2f_1(t)f_1'(t) + 2f_2(t)f_2'(t) + 2f_3(t)f_3'(t) = 0$$

which implies $f(t) \cdot f'(t) = 0$.

Example 6.12 Define $f : [0, \infty) \times [0, 2\pi] \times [0, \pi]$ by

$$f(p, \theta, \varphi) = \begin{pmatrix} p\cos\theta\sin\varphi \\ p\sin\theta\sin\varphi \\ p\cos\varphi \end{pmatrix} = \begin{pmatrix} x \\ y \\ z \end{pmatrix}$$

Then

$$Df(p, \theta, \varphi) = \begin{bmatrix} \cos\theta\sin\varphi & -p\sin\theta\sin\varphi & p\cos\theta\cos\varphi \\ \sin\theta\sin\varphi & p\cos\theta\sin\varphi & p\sin\theta\cos\varphi \\ \cos\varphi & 0 & -p\sin\varphi \end{bmatrix}$$

Recall the change of volume element:

$$\det(Df(p, \theta, \varphi)) = p^2\sin\varphi$$

Proposition 6.6 *Suppose $f : U \to \mathbb{R}^m$ and write $f = (f_1, ..., f_m)^T$. Suppose that all partial derivatives for the f_i exist and that $D_j f_i$ are continuous on U for all i, j. Then f is differentiable on U (and $Df(a) = [D_j f_i(a)]_{ij}$ for all $a \in U$)*

Proof Assume $m = 1$. (If we can prove this case we already have Df exists if and only if Df_i exist). Suppose $x = (x_1, ..., x_n) \in U$ and $h = (h_1, ..., h_n) \in \mathbb{R}^n$ such that $x + h \in \mathbb{R}^n$. Let $\tilde{h}_i = (0, ..., 0, h_i, ..., h_n)$. Then observe that

$$\frac{|f(x+h) - f(x) - \sum_{i=1}^n D_i f(x)h_i|}{||h||} \leq \sum_{i=1}^n \frac{|f(x+\tilde{h}_i) - f(x+\tilde{h}_{i+1}) - D_i f(x)h_i|}{||h||}$$

By the Mean Value Theorem, there exist c_i in between x_i and $x_i + h_i$ with

$$f(x_1, ..., x_i+h_i, x_{i+1}+h_{i+1}, ..., x_n+h_n) - f(x_1, ..., x_i, x_{i+1}+h_{i+1}, ..., x_n+h_n) = D_i f(x_1, ..., c_i, x_{i+1}+h_{i+1}, ..., x_n+h_n)h_i$$

Then

$$\begin{aligned} \frac{|f(x+h) - f(x) - \sum_{i=1}^n D_i f(x)h_i|}{||h||} &\leq \sum_{i=1}^n \frac{|D_i f(x_1, ..., c_i, x_{i+1} + h_{i+1}, ..., x_n + h_n)h_i - D_i f(x)h_i|}{||h||} \\ &\leq \sum_{i=1}^n \frac{|h_i|\, |D_i f(x_1, ..., c_i, x_{i+1} + h_{i+1}, ..., x_n + h_n) - D_i f(x)|}{|h_i|} \\ &= \sum_{i=1}^n |D_i f(x_1, ..., c_i, x_{i+1} + h_{i+1}, ..., x_n + h_n) - D_i f(x)| \end{aligned}$$

Since the $D_i f$ are continuous by assumption, and as $h \to 0$ so $h_i \to 0$, we have $c_i \to x_i$, so the above limit tends to 0. □

Example 6.13 Suppose $f : \mathbb{R}^n \to \mathbb{R}$ satisfies $|f(x)| \le ||x||^\alpha$ for all $x \in \mathbb{R}^n$ where $\alpha > 1$. Show that f is differentiable at 0 and show that $Df(0) = [0 \dots 0]$.

Observe that
$$\frac{|f(0+h) - f(0) - [0 \dots 0]h|}{||h||} = \frac{|f(h)|}{||h||} \le ||h||^{\alpha-1} \to 0$$
as $\alpha - 1 > 0$.

Theorem 6.1 (Mean value Theorem in Several Variables) *Suppose U is an open and convex subset of $\mathbb{R}^n$. (Note: Convex means that if $x, y \in U$, then the entire line segment connecting x and y is also in U). If $f : U \to \mathbb{R}$ is differentiable, then for all $x \ne y \in U$ there is a $c \in (0,1)$ so that*

$$f(y) - f(x) = Df(z)(y - x)$$

where $z = (1-c)x + cy$.

Proof Convex means if $x \ne y$ are in U, then $(1-t)x + ty \in U$ for all $0 \le t \le 1$. Defin $F : [0,1] \to \mathbb{R}$ by $F(t) = f((1-t)x + ty)$ where $x \ne y$ are fixed in U. By the chain rule,

$$F'(t) = Df((1-t)x + ty)\frac{d}{dt}((1-t)x + ty) = Df((1-t)x + ty)(y - x)$$

By the Mean Value Theorem, there exists $c \in (0,1)$ such that

$$f(y) - f(x) = F(1) = F(0) = F'(c) = Df((1-c)x + cy)(y - x)$$

as desired. □

Corollary 6.1 *Suppose $f : U \to \mathbb{R}$ is differentiable where U is an open and convex subset of $\mathbb{R}^n$. If $Df(a) = 0$ for all $a \in U$, then f is constant in U.*

By the generalized mean value theorem $f(y) - f(x) = 0$ for all $x \ne y$, so f must be constant.

Corollary 6.2 *Suppose $f : U \to \mathbb{R}^m$ is differentiable with U open and convex in $\mathbb{R}^n$. If $x \ne y \in U$ and $v \in \mathbb{R}^m$, then there exists a $c \in (0,1)$ so that*

$$v \cdot (f(y) - f(x)) = v \cdot (Df(z)(y - x))$$

for $z = (1-c)x + cy$.

We apply the theorem to the function $v \cdot f(x) = \sum_{j=1}^m v_i f_i(x)$, so $D(v \cdot f)(z) = \sum_{j=1}^m v_i Df_i(z)$, by linearity of D.

Definition 6.3 (Higher Order Derivatives) Given a differentiable function $f : U \subseteq \mathbb{R}^n \to \mathbb{R}^m$, we may regard $Df : U \to \mathbb{R}^{mn}$ and discuss its differentiability. If Df is differentiable on U, we say that f is ***twice differentiable***. In this case, we may identify $D(Df)(a)$ with an $mn \times n$ matrix.

Proposition 6.7 *If $f : U \longrightarrow \mathbb{R}^m$ is differentiable on U and $Df : U \longrightarrow \mathbb{R}^{mn}$ is differentiable at $a \in U$, then $D(Df)(a) = [D_i(D_j f_k)]$ where $1 \leq i, j \leq n$ and $1 \leq k \leq m$.*

Proof First, we note that $Df(a) = [D_1 f_1(a) \ \dots \ D_n f_1(a) \ \dots \ D_1 f_m(a) \ \dots \ D_n f_m(a)]^T \in \mathbb{R}^{mn}$. Then, by the chain rule row $1 \leq (k-1)n + i \leq mn$, with $1 \leq k \leq m$, $1 \leq i \leq n$, is $D(D_i f_k)(a)$. Next, for $1 \leq j \leq n$ column j of $D(D_i f_k)(a)$ is $D(D_i f_k)(a)e_j = D_j(D_i f_k)(a)$. Thus, the entries of $D(Df)(a)$ are $D_j(D_i f_k)$ for $1 \leq i, j \leq n$ and $1 \leq k \leq m$, as desired. □

Note that $\mathbb{R}^{mn} \cong M_{m\times n}(\mathbb{R})$ as linear spaces. Further, the 2-norm on $\mathbb{R}^{mn}$ is equivalent to the spectral norm on $M_{m\times n}(\mathbb{R})$ since all norms on finite dimensional vector spaces are equivalent. Thus $\mathbb{R}^{mn} \cong M_{m\times n}(\mathbb{R})$ are isomorphic as normed linear spaces, and hence metric spaces.

Example 6.14 Consider

$$f(x,y) = \begin{cases} 0 & (x,y) = (0,0) \\ \frac{x^3y - xy^3}{x^2+y^2} & (x,y) \neq (0,0) \end{cases}$$

is twice differentiable at $(0,0)$, but $D_{12}f(0,0) \neq D_{21}f(0,0)$. Note $D_1 f(0,0) = f_x(0,0) = 0 = D_2 f(0,0)$. For $x, y \neq 0$

$$f_x(x,y) = \frac{4x^2y^3 + x^4y - y^5}{(x^2+y^2)^2}$$

Now

$$f_{xy}(0,0) = \lim_{k\to 0} \frac{f_x(0,k) - f_x(0,0)}{k} = \lim_{k\to 0}(-1) = -1$$

Similarly we find $f_{yx}(0,0) \neq -1$.

Proposition 6.8 *Suppose $f : U \longrightarrow \mathbb{R}^m$ is twice differentiable. If $D_i(D_j f)$ and $D_j(D_i f)$ are continuous on U, then they are equal.*

Proof Without loss of generality we may suppose $m = 1$, and glue together rows in the general case using the Chain rule. So $Df = [D_1 f \ \dots \ D_n f]$, is $1 \times n$. Since we are considering all others variables besides i and j to be fixed when taking partial derivatives, they can be regarded as constants and we can regard f as a function of two variables. That is $n = 2$. Let $(a,b) \in U$ and find $h > 0$ such that $B_{2h}(a,b) \subseteq U$. Define

$$A(h) = \frac{1}{h^2}\left[f(a+h, b+h) - f(a, b+h) + f(a+h, b) - f(a,b)\right]$$

where all of the inputs to f are in U. By the mean value theorem there exist $c_1, c_2 \in (a, a+h)$ with

$$f(a+h, b+h) - f(a, b+h) = f_x(c_1, b+h)h$$
$$f(a+h, b) - f(a,b) = f_x(c_2, b)h$$

Apply the mean value theorem again to f_x. THere exists c_3 between c_1 and c_2, and d_1 between b and $b+h$ such that

$$f_x(c_1, b+h) - f_x(c_2, b) = hf_{xy}(c_3, d_1)$$

where the left hand side is $hA(h)$. Similarly, there exist c_4, d_2 with $hA(h) = hf_{yx}(c_4, d_2)$, that is, $f_{xy}(c_3, d_1) = f_{yx}(c_4, d_2)$. As $h \longrightarrow 0$, $c_3, c_4 \longrightarrow a, d_1, d_2 \longrightarrow b$. Now, since both f_{xy} and f_{yx} are continuous by assumption, we have

$$f_{xy}(a,b) = f_{yx}(a,b)$$

completing the proof. $\square$

6.2 The Inverse and Implicit Function Theorems

Simple cases of what we are trying to do:

- $f : I \longrightarrow \mathbb{R}$, where I is an interval, and $f'(x) \neq 0$ on I. From analysis one we have f is invertible, and $f^{-1} : f(I) \longrightarrow I$. f^{-1} is differentiable on $f(I)$, and $(f^{-1})'(f(x)) = \frac{1}{f'(x)}$
- Suppose $f : \mathbb{R}^m \longrightarrow \mathbb{R}^n$ is affine, so $f(x) = Ax + b$, where A is $n \times n$ and $b \in \mathbb{R}^n$. Then $Df = A$. If A is invertible, then f is invertible and $f^{-1}(x) = A^{-1}(x - b)$, so $Df^{-1} = A^{-1}$

Definition 6.4 If $f : U \longrightarrow \mathbb{R}^m$ is differentiable and $Df : U \longrightarrow \mathbb{R}^{mn}$ is continuous, we say that f is C^1.

Lemma 6.1 *$f : U \longrightarrow \mathbb{R}^{,}$ is C^1 if and only if all partials $D_i f_j$ exist and are continuous on U.*

Proof (Sketch) If $D_i f_j$ are all continuous, we know by a previous result that Df exists. Moreover,

$$Df(a+h) - Df(a) = [D_i f_j(a+h) - D_i f_j(a)] \longrightarrow 0$$

since each entry goes to 0 by continuity of $D_i f_j$. On the other hand, if Df is continuous, so is $\pi_i Df \pi_j$ = the ij-th entry of Df, which is $D_j f_i$. $\square$

Lemma 6.2 (Key to Inverse Function Theorem) *Suppose $f : U \longrightarrow \mathbb{R}^n$ is C^1 and $x_0 \in U$ with $Df(x_0)$ an invertible $n \times n$ matrix. There is an open set $W \subseteq U$ containing x_0 and a $c > 0$ such that*

$$||f(y) - f(x)|| \geq c||y - x||$$

for all $x, y \in W$.

Proof If $A \in M_n$ then $||Ax - Ay|| \leq ||A||\ ||x - y||$, where $||A||$ is the spectral norm. If A is invertible, then

$$||x - y|| = ||A^{-1}Ax - A^{-1}Ay|| \leq ||A^{-1}||\ ||Ax - Ay||$$

Since $A^{-1} \neq 0$, $||A^{-1}|| > 0$ and we can divide by it to obtain

$$\frac{1}{||A^{-1}||}||x - y|| \leq ||Ax - Ay||$$

Now suppose $Df(x_0)$ is invertible and let $c = \frac{1}{2||Df(x)^{-1}||}$. Let $f = (f_1, ..., f_n)$. All f_i are C^1 on U, so there exists an open ball W in U containing x_0 such that

$$||Df_i(y) - Df_i(x_0)|| \leq \frac{c}{n}$$

when $y \in W$, where the left hand side is the spectral or operator norm of a row.

Suppose A is $n \times n$ with rows A_i, $1 \times n$. Suppose $||A_i|| \leq \frac{c}{n}$, so $||A_i v|| \leq \frac{c}{n}||v||$ for all $v \in \mathbb{R}^n$. Then, we have that

$$||Av|| = \left\| \begin{pmatrix} A_1 v \\ \vdots \\ A_n v \end{pmatrix} \right\| \leq \sum_{i=1}^{n} ||A_i v|| \leq c$$

This implies $||Df(y) - Df(x_0)|| \leq c$ for all $y \in W$. (Note we could have gotten this immediately from continuity of Df) By the mean value theorem in multiple variables, there exists $c_i \in (0,1)$ and $z_i = (1-c_i)x + c_i y \in W$ since W is an open ball and open balls are convex, with $f_i(y) - f_i(x) = Df_i(z_i)(y-x)$. We have

$$\begin{aligned} ||f_i(y) - f_i(x) - Df_i(x_0)(y-x)|| &= ||Df_i(z_i)(y-x) - Df_i(x_0)(y-x)|| \\ &\leq ||Df_i(z_i) - Df_i(x_0)||\; ||y-x|| \\ &< \frac{c}{n}||y-x|| \end{aligned}$$

Then we have that

$$||Df(x_0)(y-x)|| - ||f(y) - f(x)|| \leq ||f(y) - f(x) - Df(x_0)(y-x)|| \leq c||y-x||$$

This implies

$$||Df(x_0)(y-x)|| - c||y-x|| \leq ||f(y) - f(x)||$$

By the first part of the proof,

$$||Df(x_0)(y-x)|| \geq \frac{1}{||Df(x_0)^{-1}||}||y-x||$$

By our previous inequality this implies

$$||f(y) - f(x)|| \geq \frac{1}{||Df(x_0)^{-1}||}||y-x|| - c||y-x|| = c||y-x||$$

since $c = \frac{1}{2||Df(x_0)^{-1}||}$. □

Corollary 6.3 *Suppose $f : U \to \mathbb{R}^n$ is C^1 and $x_0 \in U$ with $Df(x_0)$ an invertible $n \times n$ matrix. There is an open set $W \subseteq V$ containing x_0 so that $f : W \to f(W)$ is invertible.*

Using W from the proof above, $f(x) = f(y)$ for $x, y \in W$ only if $x = y$ by the inequality $||f(x) - f(y)|| \geq c||x-y||$, so $f|_W$ is injective.

Example 6.15 The assumption that f is C^1 in the above results is required. Consider

$$f(x) = \begin{cases} x + 2x^2 \sin(1/x) & x \neq 0 \\ 0 & x = 0 \end{cases}$$

is differentiable on $\mathbb{R}$ but f' is not continuous. $f'(0) = 1$ (so invertible), but f is not monotone in any neighborhood of 0. Hence it cannot be invertible in any neighborhood of 0.

Lemma 6.3 *Suppose $f : U \to \mathbb{R}^n$ is C^1 and $x_0 \in U$ with $Df(x_0)$ an invertible $n \times n$ matrix. There is an open ball $V \subseteq U$ containing x_0 so that $f|_V$ is a homeomorphism.*

Proof There exists an open ball W containing x_0 so that

(i) there exists $c > 0$ with $||f(y) - f(x)|| \geq 0||y - x||$ for all $x, y \in W$

(ii)$\overline{W} \subseteq U$

(iii)$Df(x)$ is invertible for all $x \in W$.

Note, Df is continuous so $Df(x)$ is invertible if x is close to x_0. To see (iii) is invertible in M_n, we note that the space of invertible linear transformations is an open set, since $\det : M_n \to \mathbb{R}$ is continuous and the invertibles are $\det^{-1}(\mathbb{R}\backslash\{0\})$.

This means if $Df(x)$ is close enough to $Df(x_0)$ then it is invertible. But f is C^1, which means $Df : U \to M_n$ is continuous. So there exists $V \subseteq W$ containing x_0 such that $x \in V$ implies $Df(x)$ is close enough to $Df(x_0)$ so that $Df(x)$ is invertible. Take this V and consider $f|_V : V \to f(V)$. All we must show is that $f^{-1}|_{f(V)}$ is continuous since it is already the restriction of a bijective continuous map. We do this by showing $f(O)$ is open whenever O is open in V. It is enough, by considering unions, to assume O is an open ball with $\overline{O} \subseteq V$. Let $S = \partial O$ be the boundary of O, so S is a sphere, which is compact. Note $\overline{O} = S \cup O \subseteq V$. If $x \in V\backslash S$, then

$$d = dist(f(x), f(S)) := \int_{s \in S} ||f(x) - f(s)|| > 0$$

since S is compact, f is continuous, and so $f(S)$ is a compact set not containing $f(x)$ as f is injective on V. To show $f(O)$ is open, we will show

$$B_{d/2}(f(x)) \subseteq f(O)$$

Let $z \in B_{d/2}(f(x))$, so $||z - f(x)|| < d/2$. Then

$$dist(z, f(\overline{O})) = \inf_{t \in \overline{O}} ||z - f(t)|| \leq \int_{t \in S} ||z - f(t)||$$

Let $s \in S$. Then

$$\begin{aligned} ||z - f(s)|| &= ||z - f(x) + f(x) - f(s)|| \\ &\geq ||f(x) - f(s)|| - ||z - f(x)|| \\ &> d - d/2 = d/2 \end{aligned}$$

It follows that $dist(z, f(S)) \geq d/2$. For $x \in \overline{O} = O \cup S$, define

$$g(x) = ||z - f(x)||^2 = \sum_{i=1}^{n} |z_i - f_i(x)|^2$$

$\overline{O}$ is compact and g is continuous, hence it attains its minimum on some $x_1 \in \overline{O}$, that is

$$||z - f(x_1)|| = dist(z, f(\overline{O})$$

We claim $x_1 \in O$ and not in S. Otherwise,

$$d/2 \leq dist(z, f(S)) = dist(z, f(\overline{O})) < d/2$$

a contradiction. So $x_1 \in O$. So x_1 is interior which implies $Dg(x_1) = 0$. For $1 \leq j \leq n$,

$$0 = D_j g(x_1) = -2 \sum_{i=1}^{n} (z_i - f_i(x_1)) D_j f_i(x_1)$$

by the chain rule, so

$$Df(x_1)(z - f(x_1)) = 0$$

which implies $z = f(x_1) \in f(O)$ since $Df(x_1)$ is invertible. □

Theorem 6.2 (Inverse Function Theorem) *Suppose* $f : U \to \mathbb{R}^n$ *is* C^1 *and* $x_0 \in U$ *with* $Df(x_0)$ *an invertible* $n \times n$ *matrix. There is an open ball* $V \subseteq U$ *containing* x_0 *with*

(1) $f : V \to f(V)$ *is a homeomorphism*

(2) $f^{-1} : f(V) \to V$ *is* C^1 *and*

$$Df^{-1}(f(x)) = Df(x)^{-1}$$

for all $x \in V$ *(that is* $Df^{-1}(y) = Df(f^{-1}(y))^{-1}$*)*

Proof Let V be as in the previous proof. (1) is already established by the previous result. We also know by the proof that $Df(x)$ is invertible for $x \in V$. Let $x \in V$ and $y = f(x)$. Write for $z \in f(V)$,

$$\begin{aligned} \frac{||f^{-1}(z) - f(y) - Df(x)^{-1}(z-y)||}{||z-y||} &= \frac{||Df(x)^{-1}(Df(x)f^{-1}(z) - Df(x)f^{-1}(y)) - Df(x)^{-1}(z-y)||}{||z-y||} \\ &\leq ||Df(x)^{-1}|| \frac{||Df(x)f^{-1}(z) - Df(x)f^{-1}(y) - (z-y)||}{||z-y||} \\ &\leq ||Df(x)^{-1}|| \frac{||Df(x)(f^{-1}(z) - x) - (z-y)||}{c||f^{-1} - x||} \end{aligned}$$

where there exists $c > 0$ with $||z-y|| \geq c||f^{-1}(z) - f^{-1}(y)|| = c||f^{-1}(z) - x||$ using our key lemma. The last line goes to 0 as $f^{-1}(z) \to f^{-1}(y) = x$ since f is differentiable at x.

Finally, Df^{-1} is continuous if and only if its entries are continuous. The map $^{-1} : \mathbf{GL}_n \to \mathbf{GL}_n$, $A \mapsto A^{-1}$, is continuous. Since the entries of $Df(x)$ are continuous as f is C^1, this implies the entries of $Df(x)^{-1} = Df^{-1}(f(x))$ are continuous. □

Example 6.16 Consider the function $f(x, y) = (e^x \cos y, e^x \sin y)^T = (u, v)^T$, $f : \mathbb{R}^2 \to \mathbb{R}^2$. Note if $z = x + iy$, this f is the complex exponential. f is not invertible, since $f(x, y) = f(x, y + 2\pi)$ for all $x, y \in \mathbb{R}^2$. Observe

$$Df(x,y) = \begin{bmatrix} u_x & u_y \\ v_x & v_y \end{bmatrix} = \begin{bmatrix} e^x \cos y & -e^x \sin y \\ e^x \sin y & e^x \cos y \end{bmatrix}$$

and $\det(Df(x,y)) = e^{2x} > 0$, so $Df(x,y)$ is invertible and the inverse function theorem applies. THis means there is a neighborhood U of (x,y) such that $f|_U$ is invertible, and

$$\begin{bmatrix} x_u & x_v \\ y_u & y_v \end{bmatrix} = Df^{-1}(u(x,y), v(x,y)) = \begin{bmatrix} u_x & u_y \\ v_x & v_y \end{bmatrix}^{-1}$$

Example 6.17 More generally, if $f(x,y) = (u(x,y), v(x,y))^T$ is C^1 with

$$Df(x,y) = \begin{bmatrix} u_x & u_y \\ v_x & v_y \end{bmatrix}$$

invertible, then

$$\begin{bmatrix} x_u & x_v \\ y_u & y_v \end{bmatrix} = Df^{-1}(u,v) = \begin{bmatrix} u_x & u_y \\ v_x & v_y \end{bmatrix}^{-1}$$

by the IFT.

Example 6.18 Consider $x^2 + y^2 = 1$, the unit circle. Implicitly, y as a function of x for $x \neq \pm 1$ and $\frac{dy}{dx} = \frac{-x}{y}$. Of course $y = \pm\sqrt{1-x^2}$ is an explicit representation of this, but depends on the sign of y. We can also view x as a function of y for $x \neq 0$, and $\frac{dx}{dy} = -\frac{y}{x}$. For all (x,y) on the unit circle, one variable can be implicitly written as a C^1 function of the other.

Example 6.19 (Systems of linear equations) Let $\mathbf{x} = (x_1, ..., x_n)^T \in \mathbb{R}^n$ and $\mathbf{y} = (y_1, ..., y_m)^T \in \mathbb{R}^m$, so $(\mathbf{x}, \mathbf{y}) \in \mathbb{R}^{n+m}$. Suppose A is an $n \times (n+m)$ matrix. Write

$$A = \left[\begin{array}{c|c} A_x & A_y \end{array}\right]$$

where A_x is an $n \times n$ matrix block and A_y is an $n \times m$ matrix block. The linear system $A(\mathbf{x}, \mathbf{y}) = 0$ is a system of n equations in $m + n$ variables, has m parameters in its solution if A_x is invertible (i.e. A has rank n). In this case $0 = A(\mathbf{x}, \mathbf{y}) = A_x\mathbf{x} + A_y\mathbf{y}$ if and only if $\mathbf{x} = -A_x^{-1}(A_y\mathbf{y})$ ($\mathbf{x}$ is our n variables and $\mathbf{y}$ is our m parameters).

For the next few results, a point $(x_1, ..., x_n, y_1, ..., y_m)$ in $\mathbb{R}^{n+m}$ will be abbreviated as $(\mathbf{x}, \mathbf{y})$ with $\mathbf{x} = (x_1, ..., x_n) \in \mathbb{R}^n$ and $\mathbf{y} = (y_1, ..., y_m) \in \mathbb{R}^m$.

Theorem 6.3 (Implicit Function Theorem) *Suppose U is an open subset of $\mathbb{R}^{m+n}$ and $f : U \to \mathbb{R}^n$ is C^1. Suppose $f(\mathbf{a}, \mathbf{b}) = 0$ where $a \in \mathbb{R}^n$ and $b \in \mathbb{R}^m$ and*

$$A = Df(\mathbf{a}, \mathbf{b}) = \left[\begin{array}{c|c} A_x & A_y \end{array}\right]$$

where A_x is $n \times n$ and A_y is $n \times m$. Write $f = (f_1, ..., f_n)^T$, each f_i depends on $m + n$ variables. Note $f(\mathbf{a}, \mathbf{b}) = 0$ is n equations in $m + n$ variables.

If A_x is invertible, then there exist open sets $V \subseteq U$ and $W \subseteq \mathbb{R}^m$ with $(a, b) \in V$ and $b \in W$ so that for all $y \in W$ there is a unique x with $(x, y) \in V$ and $f(x, y) = 0$. Write $x = g(y)$ in this case, so that $f(g(y), y) = 0$ on W. Then $g : W \longrightarrow \mathbb{R}^n$ is C^1, $g(b) = a$, and

$$Dg(b) = -A_x^{-1}A_y$$

In other words, we can implicitly solve for the variables $x_1, ..., x_n$ in terms of the parameters $y_1, ..., y_m$ with a C^1 function on W.

Proof Define $F : \mathbb{R}^{n+m} \longrightarrow \mathbb{R}^{n+m}$ by $F(\mathbf{x}, \mathbf{y}) = (f(\mathbf{x}, \mathbf{y}, \mathbf{y})$. Then

$$DF(\mathbf{x}, \mathbf{y}) = \left[\begin{array}{c} Df(\mathbf{x}, \mathbf{y}) \\ \hline 0 \,|\, I_m \end{array}\right]$$

We know $Df(\mathbf{a}, \mathbf{b}) = [A_x \mid A_y]$ so

$$DF(\mathbf{a}, \mathbf{b}) = \begin{bmatrix} A_x & A_y \\ 0 & I_m \end{bmatrix}$$

So F is C^1 and $DF(\mathbf{a}, \mathbf{b})$ is invertible since A_x is invertible, so $\det(DF(\mathbf{a}, \mathbf{b})) = \det(A_x)\det(I_m) \neq 0$.

Apply the Inverse Function Theorem to F at $(\mathbf{a}, \mathbf{b})$. There exists W' open in $\mathbb{R}^{m+n}$ containing $(\mathbf{a}, \mathbf{b})$ such that $F : W' \longrightarrow F(W')$ is a homeomorphism. $F(W')$ is open and $F(\mathbf{a}, \mathbf{b}) = (f(\mathbf{a}, \mathbf{b}), \mathbf{b}) = (0, \mathbf{b}) \in F(W')$. Let $W = \{\mathbf{y} \in \mathbb{R}^m | (0, \mathbf{y}) \in F(W')\}$ so $\mathbf{b} \in W$. Then $F(W') \cap \{\mathbf{0}_n\} \times \mathbb{R}^m = \{\mathbf{0}_n\} \times W$, $\{\mathbf{0}_n\} \times W$ is open in $\{\mathbf{0}_n\} \times \mathbb{R}^m \cong \mathbb{R}^m$, so W is open in $\mathbb{R}^m$. If $\mathbf{y} \in W$, then $(0, \mathbf{y}) \in F(W')$, which implies $(0, \mathbf{y}) = F(\mathbf{x}, \mathbf{y}) = (f(\mathbf{x}, \mathbf{y}), \mathbf{y})$ for some $\mathbf{x} \in \mathbb{R}^n$. This implies $f(\mathbf{x}, \mathbf{y}) = 0$. That is there is a neighborhood of $(\mathbf{a}, \mathbf{b})$ consisting of solutions to $f(\mathbf{x}, \mathbf{y}) = 0$. Fix $\mathbf{y}$ and suppose $(\mathbf{x}, \mathbf{y}) \in F(W')$ and $(\mathbf{x}_1, \mathbf{y}) \in F(W')$ such that

$$F(\mathbf{x}, \mathbf{y}) = (0, \mathbf{y}) = F(\mathbf{x}_1, \mathbf{y})$$

F is also injective on W', so $\mathbf{x}_1 = \mathbf{x}$. This shows for all $\mathbf{y} \in W$, there exists a unique $\mathbf{x} \in \mathbb{R}^m$ with $f(\mathbf{x}, \mathbf{y}) = 0$. Write $\mathbf{x} = g(\mathbf{y})$ in this case. So $g : W \longrightarrow \mathbb{R}^m$. By construction $f(g(\mathbf{y}), \mathbf{y}) = 0$ for all $\mathbf{y} \in W$. In particular, $g(\mathbf{b}) = \mathbf{a}$. Now $(g(\mathbf{y}), \mathbf{y}) = F^{-1}(0, \mathbf{y})$ and we know F^{-1} is C^1 on $F(W')$ by the inverse function theorem. This gives by the chain rule that g is C^1 on W, as $g(\mathbf{y} = \pi_{1-m} \circ F^{-1}(0, \mathbf{y})$. Write $\Phi(\mathbf{y}) = (g(\mathbf{y}), \mathbf{y})$, so

$$D\Phi(\mathbf{y}) = \begin{bmatrix} Dg(\mathbf{y}) \\ I_m \end{bmatrix}$$

which is $(m + n) \times m$. We know $f \circ \Phi = 0 \in \mathbb{R}^n$. Applying the chain rule:

$$0 = Df(\Phi(\mathbf{y}))D\Phi(\mathbf{y})$$

Let $\mathbf{y} = \mathbf{b}$, so $\Phi(\mathbf{b}) = (g(\mathbf{b}), \mathbf{b}) = (\mathbf{a}, \mathbf{b})$. Then

$$0 = Df(\mathbf{a}, \mathbf{b}) \begin{bmatrix} Dg(\mathbf{b}) \\ I_m \end{bmatrix} = [A_x \; A_y] \begin{bmatrix} Dg(\mathbf{b}) \\ I_m \end{bmatrix} = A_x D_g(\mathbf{b}) + A_y$$

so it follows that

$$Dg(\mathbf{b}) = -A_x^{-1}A_y$$

Example 6.20 Consider the unit sphere $x^2+y^2+z^2=1$. Here $f(x,y,z)=x^2+y^2+z^2-1=0$. Note

$$Df(x,y,z)=[2x \mid 2y\ 2z]$$

If $x \neq 0$, then A_x is invertible, so there exists g C^1 such that $x=g(y,z)$. In this case

$$Dg(y,z)=-A_x^{-1}A_y=\frac{-1}{2x}[2y\ 2z]$$

Note we could have used any square block in Df as our A_x. For example, taking $A_x=2y$, $A_y=[2x\ 2z]$, if $y \neq 0$ then there exists C^1 h with $y=h(x,z)$ and $Dh(x,z)=-\frac{1}{2y}[2x\ 2z]$.

Example 6.21 Consider

$$f(x_1,x_2,y_1,y_2,y_3)=\begin{bmatrix} 2e^{x_1}+x_2y_1-4y_2+3 \\ x_2\cos(x_1)-6x_1+2y_1-y_3 \end{bmatrix} : \mathbb{R}^5 \to \mathbb{R}^2$$

We have the solution $f(0,1,3,2,7)=[0\ 0]^T$, which corresponds with 2 equations in 5 unknowns. Applying the implicit function theorem to $\mathbf{a}=(0,1)^T$ and $\mathbf{b}=(3,2,7)^T$ (i.e. solve for x_1,x_2 in terms of y_1,y_2,y_3),

$$\begin{aligned} Df(\mathbf{a},\mathbf{b}) &= \left.\begin{bmatrix} 2e^{x_1} & y_1 & x_2 & -4 & 0 \\ (-x_2\sin(x_1)-6) & \cos(x_1) & 2 & 0 & -1 \end{bmatrix}\right|_{(\mathbf{a},\mathbf{b})} \\ &= \begin{bmatrix} 2 & 3 & 1 & -4 & 0 \\ -6 & 1 & 2 & 0 & -1 \end{bmatrix} \end{aligned}$$

where $A_x=\begin{bmatrix} 2 & 3 \\ -6 & 1 \end{bmatrix}$ and $A_y=\begin{bmatrix} 1 & -4 & 0 \\ 2 & 0 & -1 \end{bmatrix}$. A_x is invertible since $\det(A_x)=2+18=20 \neq 0$. Therefore, by the implicit function theorem, there is a $g=(g_1,g_2)^T$ which is C^1 with $x_1=g_1(y_1,y_2,y_3)$ and $x_2=g_2(y_1,y_2,y_3)$ and

$$Dg(\mathbf{b})=-A_x^{-1}A_y=\begin{bmatrix} 1/4 & 1/5 & -3/20 \\ -1/2 & 6/5 & 1/10 \end{bmatrix}$$

so $-1/2=\frac{\partial}{\partial y_1}g_2(3,2,7)$.

Can we do this with y_2,y_3 as the dependent variables? Yes! In this case $\mathbf{b}=(0,1,3)^T$ and $\mathbf{a}=(2,7)^T$, where $A_y=\begin{bmatrix} 2 & 3 & 1 \\ -6 & 1 & 2 \end{bmatrix}$ and $A_x=\begin{bmatrix} -4 & 0 \\ 0 & -1 \end{bmatrix}$. $\det(A_x)=4 \neq 0$, so A_x is invertible and $A_x^{-1}=\begin{bmatrix} -1/4 & 0 \\ 0 & -1 \end{bmatrix}$. Then there exists $h=(h_1,h_2)^T$, C^1, with $y_2=h_1(x_1,x_2,y_1)$ and $y_3=h_2(x_1,x_2,y_1)$, and

$$\begin{aligned} Dh(0,1,3) &= -A_x^{-1}A_y \\ &= \begin{bmatrix} 1/4 & 0 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} 2 & 3 & 1 \\ -6 & 1 & 2 \end{bmatrix} \\ &= \begin{bmatrix} 1/2 & 3/4 & 1/4 \\ -6 & 1 & 2 \end{bmatrix} \end{aligned}$$

Chapter 7

Higher-Dimensional Integration

Chapter 8

Manifolds

Chapter 9

Differential Forms

Chapter 10

Integration On Chains

Chapter 11

Integration on Manifolds

Part III

Function Spaces

Chapter 12

Normed Spaces

Chapter 13

Hilbert Spaces

Chapter 14

Banach Spaces

Chapter 15

Differentiation and Integration on Functions

Chapter 16

Banach Algebras

Part IV

Measure Theory

Chapter 17

Riemann Integration

In this chapter we define the Riemann integral for bounded functions on a closed interval $I = [a, b]$ in the real line. To do this we partition I into smaller intervals.

Definition 17.1 A ***partition*** $\mathcal{P}$ of I is a finite collection of subintervals $\{J_k : 0 \leq k \leq N\}$, disjoint except for their endpoints, whose union is I. We can order the J_k so that $J_k = [x_k, x_{k+1}]$, where

$$x_0 < x_1 < \cdots < x_N < x_{N+1},\ x_0 = a, x_{N+1} = b$$

We call the points x_k the ***endpoints*** of $\mathcal{P}$. We set

$$\ell(J_k) = x_{k+1} - x_k,\ \text{maxsize}(\mathcal{P}) = \max_{0 \leq k \leq N} \ell(J_k)$$

$$\text{minsize}(\mathcal{P}) = \min_{0 \leq k \leq N} \ell(J_k)$$

Definition 17.2 Let $f : [a, b] \to \mathbb{R}$ be a bounded function, and let $\mathcal{P} \in \prod([a, b])$ be a partition. We then define the ***upper sum*** and ***lower sum*** of f on $[a, b]$ with respect to $\mathcal{P}$ to be

$$\overline{I}_{\mathcal{P}}(f) = \sum_k \sup_{J_k} f(x)\ell(J_k)$$

$$\underline{I}_{\mathcal{P}}(f) = \sum_k \inf_{J_k} f(x)\ell(J_k)$$

Note that $\underline{I}_{\mathcal{P}}(f) \leq \overline{I}_{\mathcal{P}}(f)$. These quantities should approximate the Riemann integral of f if the partition $\mathcal{P}$ is sufficiently “fine."

Definition 17.3 Let $I = [a, b] \subseteq \mathbb{R}$. For $\mathcal{P}, \mathcal{Q} \in \prod([a, b])$, we say that $\mathcal{P}$ ***refines*** $\mathcal{Q}$, and we write $\mathcal{P} \succ \mathcal{Q}$, if $\mathcal{P}$ is formed by partitioning each interval in $\mathcal{Q}$. Equivalently, $\mathcal{P} \succ \mathcal{Q}$ if and only if all the endpoints of $\mathcal{Q}$ are also endpoints of $\mathcal{P}$.

Note that if $\mathcal{P}, \mathcal{Q} \in \prod(I)$, then $\mathcal{P} \succ \mathcal{Q}$ implies that

$$\overline{I}_{\mathcal{P}}(f) \leq \overline{I}_{Q}(f) \text{ and } \underline{I}_{\mathcal{P}}(f) \geq \underline{I}_{Q}(f)$$

since we are taking infimums and supremums over smaller sets. Consequently, if $\mathcal{P}_1, \mathcal{P}_2$ are any partitions of I, with common refinement Q,

$$\underline{I}_{\mathcal{P}_1}(f) \leq \underline{I}_{Q}(f) \leq \overline{I}_{Q}(f) \leq \overline{I}_{\mathcal{P}_2}(f)$$

We remind ourselves that if $f : I \to \mathbb{R}$ is bounded, all of these quantities are well defined.

Definition 17.4 Let $f : [a, b] \to \mathbb{R}$ be a bounded function. Then we define the upper and lower Riemann sums to be

$$\overline{I}(f) = \inf_{\mathcal{P} \in \prod([a,b])} \overline{I}_{\mathcal{P}}(f), \quad \underline{I}(f) = \sup_{\mathcal{P} \in \prod([a,b])} \underline{I}_{\mathcal{P}}(f)$$

Evidently $\underline{I}(f) \leq \overline{I}(f)$. We then say that f is ***Riemann integrable*** provided $\overline{I}(f) = \underline{I}(f)$, and in such a case we write

$$\int_I f(x)dx = \overline{I}(f) = \underline{I}(f)$$

We denote the set of Riemann integrable functions on an interval I by $\mathcal{R}(I)$.

Proposition 17.1 *If $f, g \in \mathcal{R}(I)$, then $f + g \in \mathcal{R}(I)$ and*

$$\int_I (f + g)dx = \int_I f dx + \int_I g dx$$

Proof If J_k is any subinterval of I, then $\sup_{J_k}(f + g) \leq \sup_{J_k} f + \sup_{J_k} g$ and $\inf_{J_k}(f + g) \geq \inf_{J_k} f + \inf_{J_k} g$. So for any partition $\mathcal{P}$,

$$\underline{I}_{\mathcal{P}}(f) + \underline{I}_{\mathcal{P}}(g) \leq \underline{I}_{\mathcal{P}}(f + g) \leq \overline{I}_{\mathcal{P}}(f + g) \leq \overline{I}_{\mathcal{P}}(f) + \overline{I}_{\mathcal{P}}(g)$$

We can simultaneously approximate the upper and lower Riemann sums to obtain

$$\underline{I}(f) + \underline{I}(g) \leq \underline{I}(f + g) \leq \overline{I}(f + g) \leq \overline{I}(f) + \overline{I}(g)$$

But the leftmost and rightmost terms are equal, so the whole inequality chain must be equalities and we obtain the desired result. □

Proposition 17.2 *If f is continuous on I, then f is Riemann integrable.*

Proof Since I is a compact set, f is uniformly continuous on I. Let $\omega(\delta)$ be a modulus of continuity for f, so $|x-y| \leq \delta$ implies $|f(x)-f(y)| \leq \omega(\delta)$, and $\omega(\delta) \to 0$ as $\delta \to 0$. Then for maxsize$(\mathcal{P}) \leq \delta$, $\overline{I}_{\mathcal{P}}(f) - \underline{I}_{\mathcal{P}}(f) \leq \omega(\delta) \cdot \ell(I)$, which yields the proposition. □

Proposition 17.3 *Let $f \in \mathcal{R}(I)$, and take $\varepsilon > 0$. Let $\mathcal{P}_0$ be a partition such that*

$$\overline{\mathcal{P}_0}(f) - \varepsilon \leq \int_I f dx \leq \underline{I}_{\mathcal{P}_0}(f) + \varepsilon$$

Let $M = \sup_I |f(x)|$ *and* $\delta = minsize(\mathcal{P}_0)$. *If* $\mathcal{P} \in \prod(I)$, *satisfying* $maxsize(\mathcal{P}) \leq \frac{\delta}{k}$, *for* $k \in \mathbb{N}$, *then*

$$\overline{I}_{\mathcal{P}}(f) - \varepsilon_1 \leq \int_I f dx \leq \underline{I}_{\mathcal{P}}(f) + \varepsilon_1, \quad with \;\; \varepsilon_1 = \varepsilon + \frac{2M}{k}\ell(I)$$

Proof Consider those intervals in $\mathcal{P}$ that are contained in intervals in $\mathcal{P}_0$ and those not contained in intervals in $\mathcal{P}_0$, whose lengths sum to $\leq \ell(I)/k$. Let $\mathcal{P}_1$ be the minimal common refinement of $\mathcal{P}$ and $\mathcal{P}_0$. We obtain

$$\overline{I}_{\mathcal{P}}(f) \leq \overline{I}_{\mathcal{P}_1}(f) + \frac{2M}{k}\ell(I), \quad \underline{I}_{\mathcal{P}}(f) \geq \underline{I}_{\mathcal{P}_1}(f) - \frac{2M}{k}\ell(I)$$

Since also $\overline{I}_{\mathcal{P}_1}(f) \leq \overline{I}_{\mathcal{P}_0}(f)$ and $\underline{I}_{\mathcal{P}_1}(f) \geq \underline{I}_{\mathcal{P}_0}(f)$, this implies the result. □

Corollary 17.1 (Darboux's Theorem) *Let* $\mathcal{P}_\nu$ *be any sequence of partitions of* I *into* ν *intervals* $J_{\nu k}$, $1 \leq k \leq \nu$, *such that*

$$maxsize(\mathcal{P}_\nu) = \delta_\nu \to 0$$

and let $\xi_{\nu k}$ *be any choice of one point in each interval* $J_{\nu k}$ *of the partition* $\mathcal{P}_\nu$. *Then whenever* $f \in \mathcal{R}(I)$,

$$\int_I f(x)dx = \lim_{\nu\to\infty} \sum_{k=1}^{\nu} f(\xi_{\nu k})\ell(J_{\nu k})$$

Example 17.1 For $x \in I$, set $\vartheta(x) = 1$ if $x \in \mathbb{Q}$, and $\vartheta(x) = 0$ if $x \notin \mathbb{Q}$. Now, as $\mathbb{Q}$ is dense in $\mathbb{R}$, for any partition $\mathcal{P}$ of I, we have $\overline{I}_{\mathcal{P}}(\vartheta) = \ell(I)$ and $\underline{I}_{\mathcal{P}}(\vartheta) = 0$, so $\overline{I}(\vartheta) = \ell(I)$ and $\underline{I}(\vartheta) = 0$. Note that we could make a sum like the one in the last corollary to converge if we choose rational $\xi_{\nu,k}$, which is why the corollary requires that the convergence must hold for arbitrary $\xi_{\nu,k}$.

Proposition 17.4 *Let* $f_k \in \mathcal{R}(I)$ *be a uniformly bounded monotonically increasing sequence of functions. Then there is a bounded function* f *on* I *such that as* $k \to \infty$, $f_k(x) \nearrow f(x)$, *for all* $x \in I$.

Although we would hope $\int_I f_k(x)dx \to \int_I f(x)dx$, for Riemann integration such a limit might not belong to $\mathcal{R}(I)$. The Lebesgue theory of integration remedies this defect, which we will discuss next. Now, associated to the Riemann integral is a notion of the size of a set S, called its ***content***.

Definition 17.5 If $S \subseteq I$, define the ***characteristic function***

$$\chi_S(x) = 1, \; if \; x \in S, \; 0 \, if \; x \notin S$$

We define the ***upper content*** cont^+ and the ***lower content*** cont^- by

$$\text{cont}^+(S) = \overline{I}(\chi_S), \quad \text{cont}^-(S) = \underline{I}(\chi_S)$$

We say S ***has content***, or ***is contented***, if these quantities are equal, or equivalently if and only if $\chi_S \in \mathcal{R}(I)$, in which case we denote the common value $m(S) = \int_I \chi_S(x)dx$.

From the definition we have that

$$\text{cont}^+(S) = \inf\left\{\sum_{k=1}^{N} \ell(J_k) : S \subseteq J_1 \cup \cdots \cup J_N\right\}$$

where J_k are intervals. Here we require S to be a union of a finite collection of intervals.

The key to the construction of the Lebesgue measure is to cover a set S by a countable set of intervals.

Definition 17.6 The ***outer measure*** on $S \subseteq I$ will be defined by

$$m^*(S) := \inf\left\{\sum_{k\geq 1} \ell(J_k) : S \subseteq \bigcup_{k\geq 1} J_k\right\}$$

where J_k are intervals.

Evidently, $m^*(S) \leq \text{cont}^+(S)$ since we are taking an infimum over a larger set. Note that if $S = I \cap \mathbb{Q}$, then $\chi_S = \vartheta$. In this case $\text{cont}^+(S) = \ell(I)$, but $m^*(S) = 0$. In a sense zero is the “right" measure of this set - countable sets should be considered as “small" in $\mathbb{R}$. We continue with a few more properties of the Riemann integral.

Proposition 17.5 *If* $a < b < c$, $f : [a,c] \to \mathbb{R}$, $f_1 = f|_{[a,b]}$, $f_2 = f|_{[b,c]}$, *then*

$$f \in \mathcal{R}([a,c]) \iff f_1 \in \mathcal{R}([a,b]) \text{ and } f_2 \in \mathcal{R}([b,c])$$

and if this holds,

$$\int_a^c f dx = \int_a^b f_1 dx + \int_b^c f_2 dx$$

Proof Since any partition of $[a,c]$ has a refinement for which b is an endpoint, we may consider without loss of generality $\mathcal{P} = \mathcal{P}_1 \cup \mathcal{P}_2$, where $\mathcal{P}_1$ is a partition of $[a,b]$ and $\mathcal{P}_2$ is a partition of $[b,c]$. Then

$$\overline{\mathcal{P}}(f) = \overline{I}_{\mathcal{P}_1}(f_1) + \overline{\mathcal{P}_2}(f_2), \quad \underline{I}_{\mathcal{P}}(f) = \underline{I}_{\mathcal{P}_1}(f_1) + \underline{I}_{\mathcal{P}_2}(f_2)$$

so

$$\overline{I}_{\mathcal{P}}(f) - \underline{I}_{\mathcal{P}}(f) = [\overline{I}_{\mathcal{P}_1}(f_1) - \underline{I}_{\mathcal{P}_1}(f_1)] + [\overline{I}_{\mathcal{P}_2}(f_2) - \underline{I}_{\mathcal{P}_2}(f_2)]$$

Since both terms in braces are ≥ 0, we have equivalence of Riemann integrability. Then the second result follows from the two equalities above for the upper and lower sums upon taking infimums or supremums. □

Let $I = [a,b]$. If $f \in \mathcal{R}(I)$, then $f \in \mathcal{R}([a,x])$ for all $x \in [a,b]$, and we can consider the function

$$g(x) = \int_a^x f(t)dt$$

If $a \leq x_0 \leq x_1 \leq b$, then

$$g(x_1) - g(x_0) = \int_{x_0}^{x_1} f(t)dt$$

so, if $||f||_I \leq M$, then

$$|g(x_1) - g(x_0)| \leq M|x_1 - x_0|$$

In other words, if $f \in \mathcal{R}(I)$, then g is Lipschitz continuous on I. To finish this section we prove the Fundamental Theorems of Calculus.

Theorem 17.1 (Fundamental Theorem of Calculus Part I) *If $f \in C([a, b])$, then the function g defined above is differentiable at each point $x \in (a, b)$, and*

$$g'(x) = f(x)$$

Proof For $h > 0$ we have

$$\frac{1}{h}[g(x + h) - g(x)] = \frac{1}{h}\int_x^{x+h} f(t)dt$$

If f is continuous at x, then for any $\varepsilon > 0$ there exists $\delta > 0$ such that $|f(t) - f(x)| \leq \varepsilon$ whenever $|t - x| \leq \delta$. Thus, the right side is within ε of $f(x)$ whenever $h \in (0, \delta]$. Thus, the desired limit exists as $h \searrow 0$. A similar argument treats $h \nearrow 0$. □

Theorem 17.2 (Fundamental Theorem of Calculus Part II) *If G is differentiable and $G'(x)$ is continuous on $[a, b]$, then*

$$\int_a^b G'(t)dt = G(b) - G(a)$$

Proof Consider the function $g(x) = \int_a^x G'(t)dt$. We have $g \in C([a, b])$, $g(a) = 0$, and by FTOCI, $g'(x) = G'(x)$ for all $x \in (a, b)$. Thus $f(x) = g(x) - G(x)$ is continuous on $[a, b]$, and $f'(x) = 0$ for all $x \in (a, b)$. By the Mean Value Theorem this implies f is constant on $[a, b]$. Granted this, since $f(a) = g(a) - G(a) = -G(a)$, we have $f(x) = -G(a)$ for all $x \in [a, b]$, so $g(x) = G(x) - G(a)$ for all $x \in [a, b]$. Taking $x = b$ yields the result. □

The hypothesis of this result can be weakened from $G' \in C([a, b])$ to $G' \in \mathcal{R}([a, b])$ using Darboux's theorem and a telescoping sum. We recall the Mean Value Theorem:

Theorem 17.3 (Mean Value Theorem) *Let $f : [\alpha, \beta] \to \mathbb{R}$ be continuous, and assume f is differentiable on (α, β). Then there exists $\xi \in (\alpha, \beta)$ such that*

$$f'(\xi) = \frac{f(\beta) - f(\alpha)}{\beta - \alpha}$$

Proof Consider $g(x) = f(x) - h(x - a)$, where $h = \frac{f(\beta)-f(\alpha)}{\beta-\alpha}$. Then $g(\alpha) = g(\beta)$. We claim $g'(\xi) = 0$ for some $\xi \in (\alpha, \beta)$. Indeed, since $[\alpha, \beta]$ is compact, g must assume a maximum and a minimum on $[\alpha, \beta]$. If $g(\alpha) = g(\beta)$ is the minimum, then the maximum must occur in (α, β), at some ξ. Then $g'(\xi)$ must vanish at such a point. It follows that $f'(\xi) = h = \frac{f(\beta)-f(\alpha)}{\beta-\alpha}$. □

Note 17.1 If G is a function that is differentiable on (a, b) and G' is continuous on (a, b), we say G is a C^1 function and write $G \in C^1((a, b))$. Inductively we say $G \in C^k((a, b))$ provided $G' \in C^{k-1}((a, b))$. Similarly, define $C^k([a, b])$.

Chapter 18

Lebesgue Measure on a Line

In this chapter we restrict ourselves to the concept of the Lebesgue measure of subsets of $\mathbb{R}$.

Definition 18.1 If S is a subset of an interval $I = [a, b]$, then we define the ***outer measure*** of S by

$$m^*(S) = \inf\left\{\sum_{k\geq 0} \ell(J_k) : S \subseteq \bigcup_{k\geq 0} J_k\right\}, \quad J_k \text{ intervals}$$

Note that the definition works equally well with all open or closed intervals in I. We can let J_k be intervals in $\mathbb{R}$, or we can require $J_k \subseteq I$. In particular, if $O \subseteq (a, b)$ is open, then O is a disjoint union of a countable collection of open intervals O_k, and

$$m^*(O) = \sum_{k\geq 0} \ell(O_k)$$

Furthermore, for any $S \subseteq (a, b)$,

$$m^*(S) = \inf\{m^*(O) : O \subseteq S, O \text{ open}\}$$

using the fact that $\mathbb{R}$ is second countable (has a countable base).

If $C = \{c_1, c_2, \ldots\}$ is a countable subset of I, we can write $C \subseteq \bigcup_{k\geq 1} J_k(\varepsilon)$, where $J_k(\varepsilon)$ is an open interval of length $2^{-k}\varepsilon$, centered at c_k. Thus $m^*(C) \leq \sum_{k\geq 1} 2^{-k}\varepsilon = \varepsilon$, so

$$C \subseteq I \text{ countable} \implies m^*(C) = 0$$

Note that if $\{J_{1,k} : k \geq 0\}$ covers S_1 and $\{J_{2,k} : k \geq 0\}$ covers S_2, then $\{J_{1,k}, J_{2,k} : k \geq 0\}$ is a cover of $S_1 \cup S_2$, so

$$m^*(S_1 \cup S_2) \leq m^*(S_1) + m^*(S_2)$$

This subadditivity property is shared by the upper content, but outer measure is distinguished from the upper content by its σ-subadditivity, or countable subadditivity:

Proposition 18.1 *If $\{S_j : j \geq 0\}$ is a countable family of subsets of I, then*

$$m^*\left(\bigcup_j S_j\right) \leq \sum_j m^*(S_j)$$

Proof Pick $\varepsilon > 0$. Each S_j has a countable cover $\{J_{j,k} : k \geq 0\}$, by intervals such that $m^*(S_j) \geq \sum_k \ell(J_{j,k}) - 2^{-j}\varepsilon$. Then $\{J_{j,k} : j, k \geq 0\}$ is a countable cover of $\bigcup_j S_j$ by intervals, so $m^*\left(\bigcup_j S_j\right) \leq \sum_j m^*(S_j) + 2\varepsilon$, for all $\varepsilon > 0$. Letting $\varepsilon \to 0$ we obtain the result. □

We aim to find a collection $\mathfrak{L}$ on which m^* is countably additive on disjoint subsets. First, consider $S_1 = K$ to be compact and $S_2 = I\backslash K$. Note that the outer measure of a compact set restricts to an upper content since all countable covers can be reduced to finite subcovers of smaller length. That is

$$m^*(K) = \inf\left\{\sum_{k=1}^{N} \ell(J_k) : K \subseteq \bigcup_{k=1}^{N} J_k\right\}$$

for K compact, where J_k are open intervals. This implies that given $\varepsilon > 0$, we can pick a finite collection of disjoint open intervals $\{J_k : 1 \leq k \leq N\}$ such that $O = \bigcup_{k=1}^{N} J_k \supseteq K$ and such that we have

$$m^*(K) \leq m^*(O) = \sum_{k=1}^{N} \ell(J_k) \leq m^*(K) + \varepsilon$$

Lemma 18.1 *Given $\varepsilon > 0$, we can construct $O = \bigcup_{k=1}^{N} J_k \supseteq K$ such that*

$$m^*(O\backslash K) \leq \varepsilon$$

Proof Start with the O described above. Then $O\backslash K = \mathcal{A}$ is open (since $\mathbb{R}$ is Hausdorff so K is closed), so write $\mathcal{A} = \bigcup_{k\geq 1} \mathcal{A}_k$, a countable disjoint union of open intervals. We need $\sum_{k\geq 1} \ell(\mathcal{A}_k) \leq \varepsilon$, after possibly shrinking O.

To do this, pick M large enough that $\sum_{k>M} \ell(\mathcal{A}_k) \leq \varepsilon/2$, which can be done since the series must converge to $\ell(\mathcal{A})$, and so is Cauchy. Let $\mathcal{A}_k^\# \subseteq \mathcal{A}_k$ be a closed interval with the same center as $\mathcal{A}_k$, such that $\ell(\mathcal{A}_k^\#) \geq \ell(\mathcal{A}_k) - \varepsilon/2M$. We replace O by $O\backslash \bigcup_{k=1}^{M} \mathcal{A}_k^\#$; note that this set is open, and still covers K since all the $\mathcal{A}_k^\#$ are in $\mathcal{A}$. The lemma then follows. □

Proposition 18.2 *If $K \subseteq I$ is compact, then*

$$m^*(K) + m^*(I\backslash K) = \ell(I)$$

Proof To begin, we note that if $O = \bigcup_{k=1}^{N} J_k$ is a cover of K satisfying the conditions of the previous Lemma, then $I\backslash O$ is a finite disjoint union of intervals, say $I\backslash O = \bigcup_{j=1}^{\nu} J'_j$, and $m^*(I\backslash O) = \sum_{j=1}^{\nu} \ell(J'_j)$, so

$$m^*(O) + m^*(I\backslash O) = \ell(I)$$

Furthermore, $I\backslash K = (I\backslash O) \cup (O\backslash K)$, so by the Lemma and our result on sub-additivity

$$m^*(I\backslash K) \leq m^*(I\backslash O) + \varepsilon$$

It follows that

$$m^*(K) + m^*(I\backslash K) \leq m^*(O) + m^*(I\backslash O) + \varepsilon = \ell(I) + \varepsilon$$

for all $\varepsilon > 0$ (To Be Continued) □

Chapter 19

Basic Measure Spaces

Chapter 20

L^p Spaces

Chapter 21

Radon Measures

Index

ℓ_n^p spaces, 102, 104

Abel's Theorem, 61
Accumulation point, 15, 113
Affine map, 147
Affine maps, 166
Algebras, 155
  See also Normed algebras
Arzéla-Ascoli Theorem, 162

Baire's Category Theorem, 131, 132
Balls, 18, 97, 108
Banach Fixed Point Theorem, 120
Banach space, 133
Base, 128
Binary Operation, 3
Bolzano-Weierstrass, 12, 16, 116
Boundary, 113
Bounded, 116
Bounded operators, 138
Bounds, 14

Cantor's Theorem, 24
Cardinality, 21
Cauchy, 10, 105
Closed, 15
Closed sets, 111
Closure, 15, 114
Compactness, 123
Completeness, 14, 105
  See also Completion
Concave, 36
Connected, 152
Content, 81, 209
Continuity, 16, 136
Continuum Hypothesis, 23
Contraction, 120
Convergence, 10, 15, 96, 105
Convex, 35
Countable Product, 130
Countably infinite, 21
Covering, 19
  See also Open covering
Critical point, 31

Darboux's Theorem, 76, 209
Differentiation, 25, 40
Dini's Theorem, 157
Directional derivative, 166
Divisor, 9

Equicontinuity, 160
Equivalence relation, 7
  See also Equivalence class
Euclidean Space, 95

First Category, 131
FTC I, 77, 211
FTC II, 79, 211
Function spaces, 101
Functional Sequence, 150
Fundamental Theorem of Arithmetic, 10

General MVT, 174

Heine-Borel, 16, 128
  See also Generalized Heine-Borel

Higher order derivative, 165
Homeomorphism, 140

IFT, 42, 179
ImFT, 180
Inner product space, 95
Integers, 8
Integral, 70
Interior, 110
Isolated Point, 16
Isolated point, 113
Isometry, 146
IVT, 17, 153

Lattice, 158
Lebesque measure, 82
Lindelöf, 18
Lower sum, 67, 207

Metric, 96
Metric space, 98
MVT, 32, 42, 211
  See also Cauchy MVT

NLS, 99
Norm, 95
Norms
  See also Equivalent norms
Nowhere dense, 131
Nowhere differentiable, 55

Open, 15
Open Cover, 122
Outer measure, 210, 213

Partition, 67, 207
Path connected, 153
Peano Arithmetic, 3
Picard's Theorem, 121
Pointwise convergence, 45
Power series, 56
  See also Radius of convergence
Product Metric, 129

Rationals, 9
Reals, 13
Refinement, 68
Riemann Integrable, 208
Riemann sum, 75
Rolle's Theorem, 32

Schröder-Bernstein Theorem, 21
Second category, 131
Second countable, 18
Separable, 18, 97
Separating, 156
Sequence, 10
Sequentially Compact, 123
  See also Limit point compact
Sequentially compact, 97
Stone's Theorem, 157
Subspace, 112, 153
Successor, 4

Tangent line, 25
Taylor's Theorem, 60
Totally bounded, 125

Uniform Boundedness Principal, 133
Uniform continuity, 148
Uniform Convergence, 47
Uniform convergence, 150
  See also Cauchy criterion
Uniform norm, 47
Upper sum, 67, 207

Weierstrass Approximation Theorem, 157
Weierstrass M-Test, 54, 151
Well-Ordering Property, 7

ZFC, 22