A Universal Approximation Theorem for Tychonoff Spaces with Application to Spaces of Probability and Finite Measures

by

Daniel Richard

A thesis submitted in partial fulfillment of the requirements for the degree of

Master of Science

in

Statistical Machine Learning

Department of Mathematical and Statistical Sciences
University of Alberta

Abstract

Universal approximation refers to the property of a collection of functions to approximate continuous functions. Past literature has demonstrated that neural networks are dense in continuous functions on compact subsets of finite-dimensional spaces, and this document extends those findings to non-compact and infinite dimensional spaces using homeomorphism methods. The first result herein is a universal approximation theorem for Tychonoff spaces, which is where the input to the neural network comes from some Tychonoff space. The resulting theorem shows that neural networks can arbitrarily approximate uniformly continuous functions (with respect to the sup metric) associated with a unique uniformity. The result relies on constructing a homeomorphism from a collection of real-valued functions defined on the same space that collectively separate and strongly separate points. The Tychonoff space is shown to be metrizable in the case where only a countable number of such functions is required. The second result, as a product of the Tychonoff result, is a universal approximation theorem for spaces of positive-finite measures. The motivation for our second result comes from particle filtering with the goal of making a decision based on the state distribution. We also provide some discussion showing that neural networks on positive-finite measures are a generalization of deep sets.

Acknowledgements

I would like to thank my supervisor Michael Kouritzin for providing me with this very interesting thesis topic. Thanks to his inspiration and past works, I have learned more about mathematics than I ever thought I could. It has been rewarding to play the role of mathematician and extract eternal truths into existence. I cannot thank him enough for this.

Also, I would like to thank Beatrice-Helen Vritsiou for her ongoing support throughout my days as a graduate student. Her ability to patiently explain complex concepts is uncanny and is something I wish to aspire to.

I am also grateful for the graduate committee members Martha White and Adam Kashlak for being available in relatively short notice. They are both fantastic professors and I always recommend their courses to new students.

I would also like to thank Devon Upton for the interesting conversations we shared in "the room" and on Zoom. These conversations helped me maintain my sanity throughout the pandemic.

Lastly, but certainly not least, I am thankful for my parents and sister for supporting me no matter what. This work would not have been possible without them.

Table of Contents

List of Symbols

Basic Sets and Functions

$\doteq$	Define by equality.
∞	Countable infinity.
$\subset$, $\supset$	Set containment including equality.
$\cup$, $\cap$	Set union and intersection.
$\setminus$	Set difference. $A \setminus B = \{x \in A : x \notin B\}$.
$\emptyset$	Empty set.
$\mathbb{N}$	Natural numbers excluding 0.
$\mathbb{R}$	Real numbers.
$\mathscr{R}_0(A)$	Collection of all finite subsets of A.
$f \circ g$	Function composition of g and f.

$f\colon A \to B$	f is a function mapping elements of A to elements of B.
$f^{-1}\colon B \to A$	Inverse mapping of $f\colon A \to B$.
$a \mapsto b$	a "maps to" b. Defining a function pointwise.
$f(A)$, $f^{-1}(B)$	Image and Pre-image of f.
$f\vert_A$, $\mathcal{D}\vert_A$	Restriction of f to domain A. $\mathcal{D}\vert_A = \{f\vert_A : f \in \mathcal{D}\}$.
$x \vee y$, $x \wedge y$	Binary max and min operators; equivalent to $\max\{x, y\}$ and $\min\{x, y\}$.
$B(X, Y)$	Collection of bounded functions from X to Y.

Cartesian Products

$A \times B$	Cartesian product of sets A and B.
$\prod_{i\in I} Y_i$	Cartesian product of the collection $\{Y_i\}_{i\in I}$.
Y^I	Equivalent to $\prod_{i\in I} Y$.
Y^n	Equivalent to $\prod_{i=1}^{n} Y$ for $n \in \mathbb{N} \cup \{\infty\}$.
π_{I_0}, π_i	Projection functions on Cartesian products. See (2.2).
$\bigotimes \mathcal{D}$	Function simultaneously evaluating all functions in $\mathcal{D}$ into the Cartesian product. See (2.4).

Topological Spaces

$\mathscr{O}(X)$, $\mathscr{C}(X)$	Collection of open and closed sets on topological space X.
$\mathscr{O}_X(A)$	Subspace topology induced on A by X.
$\mathscr{O}_{\mathcal{M}}(X)$	Topology induced on X by a collection of functions. See Definition 3.3.1.
$\mathscr{O}_\rho(X)$	Topology on X generated by the metric ρ.
$\mathscr{S}_{\mathcal{D}}(A)$, $\mathscr{S}_{\mathcal{D}}(A; \{\mathcal{S}_f\}_{f\in\mathcal{D}})$	Subbasis on A induced from collection of functions $\mathcal{D}$. See (3.35) and (3.36).
$\mathscr{B}_{\mathcal{D}}(A)$	Topological basis induced on A by the collection of functions $\mathcal{D}$. See Proposition 3.3.4.
$\mathscr{B}_\rho(X)$	Topological basis on X generated by the metric ρ.
$\mathscr{B}[\mathcal{S}]$	Topological basis generated from subbasis $\mathcal{S}$. See 3.3.
$\mathfrak{Q}(X, \mathcal{T})$	The sequential topology on X generated from the topology $\mathcal{T}$. See Definition 3.4.6.
$C(X;Y)$, $C(X)$	Continuous functions on topological space X to topological space Y. $C(X)$ is implies $Y = \mathbb{R}$ with standard topology.
$C_B(X;Y)$, $C_B(X)$	Bounded functions in $C(X;Y)$ and $C(X)$.
$\mathfrak{cl}[A]$, $\bar{A}$	Topological closure of $A \subset X$ for some topological space X.

Uniform Spaces

$\Delta(X)$	Diagonal of X. Defined as $\{(x,x) : x \in X\}$.
A^{-1}	Inverse relation of A. Defined as $\{(y,x) : (x,y) \in A\}$.
$A \circ B$	Composition of relations A and B. Defined as $\{(x,y) :$ for some $z \in X, (x,z) \in A$ and $(z,y) \in B\}$.
$\mathscr{U}(X)$	The uniformity on the uniform space X.
$\mathscr{U}_X(A)$	The relative uniformity on A induced by X making A a uniform subspace of X.
$\mathscr{U}_d(X)$	The metric uniformity X generated by the metric d.
$D[x]$	Defined as $\{y \in X \mid (x,y) \in D\}$ for $D \subset X \times X$.
$C_U(X;Y)$, $C_U(X)$	Uniformly continuous functions from uniform space X to uniform space Y. $C_U(X)$ implies $Y = \mathbb{R}$ with standard uniformity.
$\mathfrak{S}_{\mathcal{M}}$	Uniformity associated with the collection of functions $\mathcal{M}$. See Notation 5.3.15.

Measure Spaces

$\sigma(\mathcal{C})$	σ-algebra generated by $\mathcal{C}$. See Proposition 6.1.3.
$\mathfrak{B}(E)$	Borel sets of topological space E. Equivalent to $\sigma(\mathscr{O}(E))$.

$M(X;Y)$, $M(X)$	Measurable functions on measurable space X to measurable space Y. $M(X)$ implies $Y=\mathbb{R}$ with σ-algebra $\mathfrak{B}(\mathbb{R})$.
$M_B(X;Y)$, $M_B(X)$	Bounded functions in $M(X;Y)$ and $M(X)$.
$\Rightarrow$	Weak convergence of positive-finite measures. See Definition 6.2.7.
$\mathscr{M}^+(E)$	Collection of positive-finite measures on measurable space E.
$\mathscr{P}(E)$	Collection of probability measures on measurable space E.
$\int_E f\, d\mu$	Lebesgue integral.
f^*	The mapping $f^*(\mu) \mapsto \int_E f\, d\mu$.
$\mathcal{D}^*$	Defined as $\{f^* : f \in \mathcal{D}\}$.
$\mathcal{T}^W$	The weak topology of positive-finite measures. See Definition 6.2.5.
$\mathcal{T}^{WC}$	The topology of weak convergence of positive-finite measures. See Definition 6.2.10.
$\mathfrak{W}_{g_0}[\mathcal{M}]$	Collection of functions on positive-finite measures generated by the collection of functions $\mathcal{M} \cup \{g_0\}$. See Proposition 6.3.5.

Chapter 1

Introduction

1.1 Machine Learning

In statistical machine learning, practitioners are often faced with the task of making predictions based on past data examples.

Let X and $\mathbb{R}$ be the sets of possible predictor values and target values, respectively; $p(x, y)$ be the population probability distribution over $X \times \mathbb{R}$; and suppose $D = \{(x_i, y_i)\}_{i=1}^n \subset X \times \mathbb{R}$ represent the past data examples sampled independently from $p(x, y)$. The goal is to find a "good" predictor function $f \colon X \to \mathbb{R}$ such that $f(x_i)$ is "close" to y_i for each $i = 1, \ldots, n$ from a collection of possible predictor functions, denoted $\mathcal{M}$, as defined by the practitioner. Whether a predictor function is good is a judgement for the practitioner to make; however, often one tries to choose the function that minimizes some kind of mean error. Suppose $r \colon \mathbb{R} \times \mathbb{R} \to [0, \infty)$ is the selected error function, then the mean error for a predictor function across $p(x, y)$ could be defined as

$$\widetilde{A}(f) = \int_{X \times \mathbb{R}} r(f(x), y)\, dp(x, y), \tag{1.1}$$

for which the best predictor function g with respect to $p(x, y)$ is defined as

$$\widetilde{g} = \underset{f \in \mathcal{M}}{\operatorname{argmin}} \{\widetilde{A}(f)\}, \tag{1.2}$$

which may or may not be unique.

The next main question that arises is: how does the practitioner find $\widetilde{g}$? There are a few practical issues to work through which are:

1. $p(x, y)$ is unknown, and
2. Can each $f \in \mathcal{M}$ be represented on a computer?

Fortunately, there are solutions. For (1), the practitioner has the data sample to work with, so they can use the empirical distribution in place of the population distribution. As such, the average error for a predictor function across D may be calculated as

$$A_n(f) = \frac{1}{n}\sum_{i=1}^{n} r(f(x_i), y_i), \tag{1.3}$$

where the best predictor function g with respect to D is given as

$$\widehat{g}_n = \underset{f\in\mathcal{M}}{\operatorname{argmin}}\{A_n(f)\}. \tag{1.4}$$

For (2), the practitioner can choose a parameterized computer-workable function f_θ where θ embodies the parameters controlling the behaviour of the predictor function. Typically the parameters are real numbers, so suppose $\Theta \subset \mathbb{R}^d$, and define a new collection of computer-workable functions as $\mathcal{M}_\Theta = \{f_\theta : \theta \in \Theta\}$ that, ideally, closely resembles $\mathcal{M}$ from (1.2). Now the search for the best prediction function becomes a search for $\widehat{\theta}$ defined as follows:

$$\widehat{\theta} = \underset{\theta\in\Theta}{\operatorname{argmin}}\{A_n(f_\theta)\}, \tag{1.5}$$

and so $f_{\widehat{\theta}} \in \mathcal{M}_\Theta$.

Another question to ask is: how different are $f_{\widehat{\theta}}$ and $\widehat{g}_n$? If the difference is small, then the practitioner can willingly use $f_{\widehat{\theta}}$ as a replacement for $\widehat{g}_n$ as the best predictor function; however, if the difference is quite large, then they should choose a more sophisticated class of parameterized functions. Ultimately, the answer depends on how closely $\mathcal{M}_\Theta$ is to $\mathcal{M}$ in some way. Consider the following definition:

Definition 1.1.1 (Uniform Dense)**.** Let $\mathcal{F}$ and $\mathcal{G}$ be collections of real valued functions with common domain X. We say $\mathcal{F}$ is *uniform dense* in $\mathcal{G}$ if and only if for each $g \in \mathcal{G}$ and $\epsilon > 0$, there exists an $f \in \mathcal{F}$ such that

$$\sup \{|f(x) - g(x)| : x \in X\} < \epsilon. \tag{1.6}$$

In addition, if $\mathcal{F} \subset \mathcal{G}$, then $\mathcal{F}$ is said to be a *uniform dense subset of* $\mathcal{G}$.

It then follows that if $\mathcal{M}_\Theta$ is uniform dense in $\mathcal{M}$, then the practitioner can rest assured that f_θ can be made close enough to $\widehat{g}_n$ for some setting of θ. Often the practitioner wishes for $\mathcal{M}_\Theta$ to be uniform dense within a collection of continuous functions of interest; in which case, $\mathcal{M}_\Theta$ is said to have the universal approximation property.

Cybenko 1989 showed that neural networks have the universal approximation property. In particular, he showed that functions of following form:

$$x \mapsto \sum_{j=1}^{n} \beta_j \sigma(a_j' x - \theta_j) \qquad\qquad a_j \in \mathbb{R}^k; \beta_j, \theta_j \in \mathbb{R}, \tag{1.7}$$

where $'$ denotes transpose (so $a_j' x$ is the dot product of a and x) and σ is a real to real valued function[1], are uniform dense in the continuous functions defined on $[0,1]^k$.

In what follows, let $C(X)$ be the collection of real valued continuous functions on X, and $f|_A$ be the restriction of f to the subset $A \subset X$.

Definition 1.1.2 (Uniform Dense on Compacts)**.** Let X be a topological space. Then, $\mathcal{F} \subset C(X)$ is said to be *uniform dense on compacts of* X if for each compact $K \subset X$, $\{f|_K : f \in \mathcal{F}\}$ is uniform dense in $C(K)$.

Hornik 1991 extended the work of Cybenko 1989 to all the compact subsets of $\mathbb{R}^k$. We repeat the theorem here.

[1]Additionally, σ must be discriminatory, which is to say that $\int_{[0,1]^k} \sigma(a_j' x - \theta_j)\, d\mu = 0$ for each $a_j \in \mathbb{R}^k$ and $\theta_j \in \mathbb{R}$ implies $\mu = 0$.

Theorem 1.1.3. *If* $\sigma \colon \mathbb{R} \to \mathbb{R}$ *is continuous, bounded and non-constant, then the following functions*

$$\bigcup_{n\in\mathbb{N}} \left\{ x \mapsto \sum_{j=1}^{n} \beta_j \sigma(a_j' x - \theta_j) : a_j \in \mathbb{R}^k; \beta_j, \theta_j \in \mathbb{R} \right\} \tag{1.8}$$

are uniform dense on compacts of $\mathbb{R}^k$.

Proof. See Theorem 2 from Hornik 1991. ∎

There are many collections of functions that are uniform dense of the compacts[2] of $\mathbb{R}^k$; however, (for this work) we need not focus on the variety of such collections, but rather that they exist.

A question to consider is: what universal approximation results can be had when the function domain is not $\mathbb{R}^n$? Suppose X is a topological space, and for the moment that $h \colon X \to \mathbb{R}^n$ is a homeomorphism. Given a compact set $K \subset X$, it follows that $h|_K \colon K \to h(K)$ is a homeomorphism to the compact set $h(K) \subset \mathbb{R}^n$, so we see that for any $g \in C(K)$, we have $g = g \circ h|_K^{-1} \circ h|_K$ and $g \circ h|_K^{-1} \in C(h(K))$. Letting $\mathcal{A}_n$ denote the neural networks from Theorem 1.1.3, given $\epsilon > 0$ there is $f \in \mathcal{A}_n$ such that

$$\begin{aligned} \epsilon &> \sup \left\{ |f(p) - g \circ h|_K^{-1}(p)| : p \in h(K) \right\} && (1.9) \\ &= \sup \left\{ |f \circ h|_K(q) - g(q)| : q \in K \right\}, && (1.10) \end{aligned}$$

which is to say that $\{p \mapsto f \circ h|_K(p) : f \in \mathcal{A}_n\}$ is uniform dense in $C(K)$. As K is arbitrary, it follows that $\{p \mapsto f \circ h(p) : f \in \mathcal{A}_n\}$ is uniform dense on compacts of X.

The above example suggests that homeomorphisms are a useful tool to achieving universal approximation results for non-real input values; however, there is a few other questions to answer which are:

1. Is there a uniform dense result for when X is not compact?

2. What if X is infinite dimensional? That is, there is no such homeomorphism $h \colon X \to \mathbb{R}^n$ for some $n \in \mathbb{N}$.

[2]This includes deep neural networks (see Kidger and Lyons 2020 Theorem 3.2).

To motivate the above questions further with an example, consider the following hidden Markov model:

$$X_0 = x_0 \sim p(x_0) \tag{1.11}$$

$$X_i | X_{i-1} = x_i \sim p(x_i \mid x_{i-1}) \tag{1.12}$$

$$Y_i | X_i = y_i \sim p(y_i \mid x_i) \tag{1.13}$$

for $i = 1, \ldots, n$ and $x_i, y_i \in \mathbb{R}$. The X_i are hidden (non-observed) random variables, while the Y_i are observed for $i < n$. The goal is then to compute the conditional distribution for the hidden variables given the observations $p(x_i \mid y_i, i < n)$. Particle filtering is a common technique for computing the distribution by representing it as $m \in \mathbb{N}$ number of weighted particles. For $j = 1, \ldots, m$ let $d_j \in \mathbb{R}$ and $L_j > 0$ be the particle value and likelihood (given y_i for $i < n$), respectively, for the j-th particle. Then the unnormalized measure based on the particles is given as

$$\mu(A) = \sum_{j=1}^{m} L_j I_{d_j}(A), \tag{1.14}$$

where $A \subset \mathbb{R}$ is some measurable set and I_c is the indicator function ($I_c(A) \mapsto 1$ if $c \in A$ and is 0 otherwise), can be used to represent the target distribution as

$$p(x_i \in A \mid y_i, i < n) \approx \frac{\mu(A)}{\mu(\mathbb{R})}, \tag{1.15}$$

where $\approx$ means "approximately". Both μ and $p(x_i \in A \mid y_i, i < n)$ give us information about the hidden variable X_i given the past y_i's, but how can we best utilize this information? Suppose that we want to learn a decision function with a positive-finite measure as an input. It is then natural to ask: what class of decision functions is appropriate for us to approximate continuous functions of positive-finite measures? We are unable to use our homeomorphism tricks exactly as done previously because the input space is infinite dimensional. However, Ma et al. 2020 (section 3.4) has shown some successful empirical work on this matter by evaluating the moment generating function for the input measure μ at various points, then passing those points into a neural network. We can compute the moment generating function for μ evaluated at $v \in \mathbb{R}$ as below:

$$M_\mu(v) = \int e^{v \cdot z} \, d\mu(z) \tag{1.16}$$

$$= \sum_{i=1}^{m} L_j \cdot e^{v \cdot d_j}. \tag{1.17}$$

Hence, their decision functions look like

$$\mu \mapsto f\left(M_\mu(v_1), \ldots, M_\mu(v_k)\right), \tag{1.18}$$

where $f \in \mathcal{A}_k$ is a neural network on $\mathbb{R}^k$. In Ma et al. 2020, the $v_1, \ldots, v_k$ are implemented as parameters, so the points in which they evaluate the moment generating function are learned. It is the goal of this document to provide some theoretical justification for this practise by providing a universal approximation theorem for positive-finite measures.

Chapters 2 through 6 provide the necessary mathematical background. Chapter 2 is focused on notation for sets and functions. Chapter 3 is about topological spaces which provide the framework for continuity. Chapter 4 is about point separation ideas which we will use to construct homeomorphisms on infinite dimensional spaces. Chapter 5 introduces uniform spaces for the purpose of identifying the class of functions for which our universal approximation theorems hold: they are exactly the uniformly continuous functions associated to a unique uniformity. Chapter 6 discusses measure spaces and topological spaces of measures as well as some homeomorphism ideas specific to measures.

Chapter 7 is where our universal approximation theorems are kept. The main results are Theorem 7.2.2 for Tychonoff spaces and Theorem 7.3.2 for spaces of positive-finite measures.

Chapter 8 concludes with some practical discussion related to the learning process, as well as thoughts about how neural networks with measures as inputs can be seen as a generalization of deep sets.

Chapter 2

Basic Notation

This brief chapter provides basic notation and concepts for sets and functions that will be used extensively throughout the following chapters.

2.1 Sets and Functions

We use "$\doteq$" when defining an object via equality. We will use $\emptyset$, $\mathbb{N}$, and $\mathbb{R}$ to represent the empty set, positive integers, and real numbers, respectively. We use ∞ to indicate countable infinity.

"$\subset$" and "$\supset$" are used to denote set containment including equality; that is $A \subset B$ if and only if $x \in A$ implies $x \in B$. We use "$\setminus$" for set difference, so $A \setminus B \doteq \{x \in A : x \notin B\}$.

Given a set X, we denote $\mathscr{R}_0(X)$ as the set of *all non-empty finite subsets* of X.

Let X and Y be sets. We use the standard notation $f\colon X \to Y$ to denote a function f with domain X and range Y. When defining a function pointwise, it is useful to use the notation "$\mapsto$" to indicate point assignment, so $f(p) \mapsto q$ implies $f(p) = q$. However, "$\mapsto$" allows us to define mappings without referring to f; for example, if the context is $\mathbb{R}$ to $\mathbb{R}$ functions, then $\{p \mapsto ap^2 : a \in (0, \infty)\}$ is the set of parabolas with a minimum at 0. Should f be invertible, then f^{-1} is defined as its inverse.

Given two functions $f\colon X \to Y$ and $g\colon Y \to U$, the composition of g and

f is denoted $g \circ f \colon X \to U$.

Suppose $A \subset X$ and $B \subset Y$. We refer to $f(A) \doteq \{f(x) \in Y : x \in A\}$ and $f^{-1}(B) \doteq \{x \in X : f(x) \in B\}$ as the *image* of A under f and *pre-image* of B under f, respectively. The following fact about pre-image is indispensable.

Fact 2.1.1. *Suppose $S_i \subset Y$ for each $i \in I$, and f a function mapping elements of X to elements of Y. The following are true:*

1. $f^{-1}\left(\bigcap_{i\in I} S_i\right) = \bigcap_{i\in I} f^{-1}(S_i)$, *and*

2. $f^{-1}\left(\bigcup_{i\in I} S_i\right) = \bigcup_{i\in I} f^{-1}(S_i)$.

The *restriction* of f to the domain A is a function denoted $f|_A \colon A \to Y$ defined as $f|_A(x) \doteq f(x)$ for each $x \in A$. If $\mathcal{D}$ is a collection of functions with common domain X, then $\mathcal{D}|_A \doteq \{f|_A : f \in \mathcal{D}\}$ is the collection $\mathcal{D}$ restricted to the domain A.

We use the symbols $\vee$ and $\wedge$ to represent binary operations of max and min functions, respectively. That is, $x \vee y \mapsto \max\{x, y\}$ and $x \wedge y \mapsto \min\{x, y\}$.

2.2 Cartesian Products

We let $A \times B \doteq \{(x, y) : x \in A, y \in B\}$ denote the Cartesian product of non-empty sets A and B. Given an index set I and collection of non-empty sets $\{Y_i\}_{i\in I}$ we denote the Cartesian product of the collection as $\prod_{i\in I} Y_i \doteq \{(x_i)_{i\in I} : x_j \in Y_j \text{ for each } j \in I\}$. In the common case where $Y_i = Y$ for each $i \in I$, we use the alternative notation $Y^I \doteq \prod_{i\in I} Y$. When the cardinality of I is $n \in \mathbb{N} \cup \{\infty\}$, it is often convenient to instead use the notation Y^n rather than Y^I.

We use the Cartesian product Y^X to represent the collection of all functions from X to Y and $B(X, Y) \subset Y^X$ is the bounded functions from X to Y.

Given an index set I, non-empty $I_0 \subset I$, and collection of non-empty sets $\{Y_i\}_{i\in I}$ we define the I_0 projection function on $\prod_{i\in I} Y_i$ as

$$\pi_{I_0} \colon \prod_{i\in I} Y_i \to \prod_{i\in I_0} Y_i \tag{2.1}$$

$$\pi_{I_0}\left((x_i)_{i\in I}\right) \mapsto (x_i)_{i\in I_0}. \tag{2.2}$$

In the common case I_0 has cardinality of 1, then $I_0 = \{j\}$ for some $j \in I$, so we then define $\pi_j \doteq \pi_{\{j\}}$. Suppose $f_i \colon X \to Y_i$ is a mapping for each $i \in I$ and let $\mathcal{D} = \{f_i : i \in I\}$, then we define the function

$$\bigotimes \mathcal{D} \colon X \to \prod_{i\in I} Y_i \tag{2.3}$$

$$\bigotimes \mathcal{D}(x) \mapsto (f_i(x))_{i\in I}, \tag{2.4}$$

which simultaneously evaluates all of the functions in $\mathcal{D}$.

Chapter 3

Topological Spaces

Universal approximation is about continuous functions. And continuous functions live in topology. So it should be of no surprise that we start our mathematical background here and that this Chapter is the largest.

The first section starts with introducing basic notation and terminology regarding topological spaces including topological bases and nets. This leads to the topic of homeomorphisms, whose glue-like properties will be vital for this work. Next, we introduce some common types and classifications of topological spaces that we will encounter in future Chapters. Then we discuss how topologies may be generated from any collection of functions with common domain, and the resulting topology ensures the collection of functions are continuous. Our last section is about sequential spaces which make the verification of continuity much easier by only having to study convergent sequences as opposed to nets.

3.1 Standard Topological Notions

We will make use of standard topological definitions from Munkres 2000 and Willard 2004. A *topology* on a set X is a collection $\mathcal{T}$ of subsets of X having the following three properties:

1. $\{\emptyset, X\} \subset \mathcal{T}$.

2. $\mathcal{C} \subset \mathcal{T}$ implies $\bigcup_{S \in \mathcal{C}} S \in \mathcal{T}$.

3. $\mathcal{C}_0 \in \mathscr{R}_0(\mathcal{T})$ implies $\bigcap_{S \in \mathcal{C}_0} S \in \mathcal{T}$.

The ordered pair $(X, \mathcal{T})$ is called a *topological space*, the elements of $\mathcal{T}$ are called the *open* sets of X, and $\{X \setminus S : S \in \mathcal{T}\}$ are called the *closed* sets of X. When no confusion arises, the ordered pair notation will be suppressed and we will simply refer to X as a topological space.

For later convenience, we introduce alternative notation helpful for referring to related topologies. Let X be a topological space and $A \subset X$ be nonempty. By $\mathscr{O}(X)$ and $\mathscr{C}(X)$ we denote the families of all open and closed subsets of X, respectively. That is, $\mathscr{O}(X, \mathcal{T}) = \mathcal{T}$ and $\mathscr{C}(X, \mathcal{T}) = \{X \setminus S : S \in \mathcal{T}\}$. Additionally,

$$\mathscr{O}_X(A) \doteq \{O \cap A : O \in \mathscr{O}(X)\} \tag{3.1}$$

denotes the *subspace topology* of A induced from X.

A function f mapping elements of topological spaces X to Y is *continuous* if $f^{-1}(A) \in \mathscr{O}(X)$ for each $A \in \mathscr{O}(X)$. The collection of continuous functions from X to Y is denoted $C(X; Y)$ or $C((X, \mathcal{T}_X); (Y, \mathcal{T}_Y))$ depending on the need for clarity. In the case where Y is the real numbers $\mathbb{R}$ with standard topology, then the abbreviated notation $C(X)$ or $C(X, \mathcal{T}_X)$ may instead be used.

Suppose $\mathcal{T}$ and $\mathcal{T}'$ are topologies on X and $\mathcal{T} \subset \mathcal{T}'$. We then say $\mathcal{T}$ is *coarser* than $\mathcal{T}'$ or, equivalently, $\mathcal{T}'$ is *finer* than $\mathcal{T}$. It is clear then that $\mathcal{T} \subset \mathcal{T}'$ implies $C((X, \mathcal{T}); Y) \subset C((X, \mathcal{T}'); Y)$; that is, finer topologies (on domain spaces) emit richer classes of continuous functions.

A *basis* for a topology on X is a collection $\mathcal{B}$ of subsets of X (called *basis elements*) such that

1. For each $x \in X$, there is a $B \in \mathcal{B}$ such that $x \in B$.

2. For any two basis elements B_1 and B_2, if $x \in B_1 \cap B_2$, then there is a $B_3 \in \mathcal{B}$ such that $x \in B_3 \subset B_1 \cap B_2$.

The *topology* $\mathcal{T}$ *generated by* $\mathcal{B}$ is defined as follows: $A \subset X$ is in $\mathcal{T}$ if for each $x \in A$, there is a basis element $B \in \mathcal{B}$ such that $x \in B \subset A$.

We now reiterate the following three facts from Munkres 2000 (pages 78 - 81).

Fact 3.1.1. *If $\mathcal{B}$ is a basis for a topology $\mathcal{T}$, then $\mathcal{T}$ is the collection of all unions of elements of $\mathcal{B}$.*

Fact 3.1.2. *Let $(X, \mathcal{T})$ be a topological space. Suppose that $\mathcal{C} \subset \mathcal{T}$ such that for each $U \in \mathcal{T}$ and each $x \in U$, there is an element of $C \in \mathcal{C}$ such that $x \in C \subset U$. Then $\mathcal{C}$ is a basis for $\mathcal{T}$.*

Fact 3.1.3. *Let $\mathcal{B}$ and $\mathcal{B}'$ be bases for the topologies $\mathcal{T}$ and $\mathcal{T}'$, respectively, on X. Then the following are equivalent:*

1. *$\mathcal{T}'$ is finer than $\mathcal{T}$.*
2. *For each $x \in X$ and each basis element $B \in \mathcal{B}$ containing x, there is a basis element $B' \in \mathcal{B}'$ such that $x \in B' \subset B$.*

A *subbasis* $\mathcal{S}$ for a topology on X is a collection of subsets of X whose union equals X. That is,

$$\bigcup_{V \in \mathcal{S}} V = X. \tag{3.2}$$

The topology generated from a subbasis $\mathcal{S}$ is generated from the following basis:

$$\mathscr{B}[\mathcal{S}] \doteq \left\{ \bigcap_{U \in \mathcal{U}_0} U : \mathcal{U}_0 \in \mathscr{R}_0[\mathcal{S}] \right\}. \tag{3.3}$$

Alternatively, it is useful to think of topologies in terms of neighborhoods and neighborhood bases, which we define below.

Definition 3.1.4 (Neighborhood, Neighborhood Base)**.** If X is a topological space and $x \in X$, a *neighborhood* of x is a set U which contains an open set V containing x. A *neighborhood base* at x is a collection of neighborhoods $\mathcal{N}_x$ at x with the property: if U is any neighborhood of x, then $U \supset V$ for some $V \in \mathcal{N}_x$.

Proposition 3.1.5. *Let X be a set and $\mathcal{N}_x$ be a collection of subsets of X for each $x \in X$ that satisfy the following properties:*

- *$V \in \mathcal{N}_x$ implies $x \in V$,*
- *$V_1, V_2 \in \mathcal{N}_x$ implies there is some $V_3 \in \mathcal{N}_x$ such that $V_3 \subset V_1 \cap V_2$, and*
- *$V \in \mathcal{N}_x$ implies there is some $V_0 \in \mathcal{N}_x$ such that if $y \in V_0$, then there is some $W \in \mathcal{N}_y$ with $W \subset V$.*

Suppose $\mathcal{T}$ is a collection of subsets of X, that satisfy: $U \in \mathcal{T}$ if and only if $x \in U$ implies there is a $V_x \in \mathcal{N}_x$ such that $V_x \subset U$. Then $\mathcal{T}$ is a topology on X and, for each $x \in X$, $\mathcal{N}_x$ is a neighborhood base at x.

Proof. Proven by Willard 2004 Theorem 4.5. ∎

Definition 3.1.6 (First-countable)**.** If each $p \in X$ has a countable neighborhood base, then X is said to be *first-countable*.

3.1.1 Nets and Limit Points

Next, we bring forth some properties of nets and limit points in topological spaces, which are yet another way to think about topological spaces.

Definition 3.1.7 (Limit point, Closure)**.** Suppose X is a topological space, $A \subset X$, $p \in X$, and $\mathcal{N}_p$ be the neighborhood system at p. We say p is a *limit point* of A if $A \cap O \neq \emptyset$ for each $O \in \mathcal{N}_p$. We define the *closure* of A as the intersection of all closed sets containing A and is denoted as $\bar{A}$ or $\mathfrak{cl}[A]$.

Proposition 3.1.8. *Let A be a subset of a topological space X and let A' be the set of all limit points of A. Then*

1. *$\bar{A} = A \cup A'$,*
2. *A is closed if and only if $A = \bar{A}$.*

Proof. See Munkres 2000 Theorem 17.6 and Corollary 17.7. ∎

Definition 3.1.9 (Directed sets)**.** A set Λ is a *directed set* if and only if there is a relation $\leq$ on Λ satisfying:

- $\lambda \leq \lambda$, for each $\lambda \in \Lambda$,
- if $\lambda_1 \leq \lambda_2$ and $\lambda_2 \leq \lambda_3$ then $\lambda_1 \leq \lambda_3$,
- if $\lambda_1, \lambda_2 \in \Lambda$ then there is some $\lambda_3 \in \Lambda$ with $\lambda_1 \leq \lambda_3$, $\lambda_2 \leq \lambda_3$.

Definition 3.1.10 (Nets, Subnets, Sequence)**.** A *net* in a set X is a function $P : \Lambda \to X$, where Λ is some directed set. The point $P(\lambda)$ is usually denoted x_λ and we may prefer the notation $(x_\lambda)_{\lambda\in\Lambda}$ or simply (x_λ) when there is no confusion regarding the directed set.

Given a directed set M, a *subnet* of P is the composition $P \circ \phi$, where $\phi \colon M \to \Lambda$ satisfies:

- $\phi(\mu_1) \leq \phi(\mu_2)$ whenever $\mu_1 \leq \mu_2$, and
- for each $\lambda \in \Lambda$, there is some $\mu \in M$ such that $\lambda \leq \phi(\mu)$.

Like for nets, often $P \circ \phi(\mu)$ will be denoted as x_{λ_μ}.

A *sequence* is a net whose directed set has the cardinality of the natural numbers.

Definition 3.1.11 (Net convergence)**.** Let $(x_\lambda)_{\lambda\in\Lambda}$ be a net in a topological space X. Then $(x_\lambda)_{\lambda\in\Lambda}$ *converges* to $p \in X$ (denoted $x_\lambda \to p$) if and only if for each neighborhood U of p, there is some $\lambda_0 \in \Lambda$ such that $\lambda \geq \lambda_0$ implies $x_\lambda \in U$. We then say p is a limit of $(x_\lambda)_{\lambda\in\Lambda}$ and is denoted as $\lim_{\lambda\in\Lambda} x_\lambda$ when the limit is unique.

Proposition 3.1.12. *Let (x_λ) be a net in a topological space X and suppose $x \in X$. The following statements are true:*

1. *If $x_\lambda = x$ for each λ, then $x_\lambda \to x$.*
2. *If $x_\lambda \to x$, then every subnet of (x_λ) converges to x.*
3. *If every subnet of (x_λ) has a subnet converging to x, then (x_λ) converges to x.*

Proof. (1.) Each neighborhood of x contains $x = x_\lambda$, so it is clear that $x_\lambda \to x$. (2.) Pick a neighborhood U of x and subnet (x_{λ_μ}) of (x_λ). There is some $\lambda_0 \in \Lambda$ such that $\lambda \geq \lambda_0$ implies $x_\lambda \in U$. By Definition 3.1.10, for a subnet we can choose μ_0 such that $\lambda_0 \leq \lambda_{\mu_0}$. It then follows that $x_{\lambda_\mu} \in U$ for all $\mu \geq \mu_0$, so $x_{\lambda_\mu} \to x$.

(3.) Suppose (x_λ) does not converge to x. Then, for some neighborhood U of x, for each λ there is $\lambda_0 \geq \lambda$ such that $x_{\lambda_0} \notin U$. That is, there is a subnet (x_{λ_μ}) such that $x_{\lambda_\mu} \notin U$ for all μ. It then follows that a further subnet of (x_{λ_μ}) cannot converge x, which is a contradiction. ■

It is the following three propositions that demonstrate the importance of nets within the study of general topological spaces. In particular, nets perfectly characterize closed sets, continuity, and compactness.

Proposition 3.1.13. *Suppose X is a topological space and $A \subset X$. Then $x \in \bar{A}$ if and only if there is a net (x_λ) in A with $x_\lambda \to x$.*

Proof. See Willard 2004 Theorem 11.7. ■

Proposition 3.1.14. *Suppose X and Y are topological spaces and $f\colon X \to Y$. Then f is continuous at $p \in X$ if and only if $f(x_\lambda) \to f(p)$ whenever $x_\lambda \to p$.*

Proof. See Willard 2004 Theorem 11.8. ■

Proposition 3.1.15. *A topological space is compact if and only if each net has a convergent subnet.*

Proof. See Willard 2004 Theorems 17.4 and 11.5. ■

3.1.2 Homeomorphisms

Much of the work in later chapters is dedicated to the construction of homeomorphisms. For now, all we do is define them and list some useful facts about them.

Definition 3.1.16 (Bijection, Inverse, Homeomorphism, Embedding)**.** Suppose X and Y are topological spaces and $f\colon X \to Y$. If, for each $q \in Y$ there is a unique $x \in X$ such that $f(x) = y$, then f is a *bijection*. If f is a bijection, then its inverse is denoted $f^{-1}\colon Y \to X$ and defined as $f^{-1}(y) = x$ if and only if $f(x) = y$. We call f a *homeomorphism* if it is a continuous bijection with continuous inverse. If $f\colon X \to f(X)$ is a homeomorphism then f is called an *embedding* of X into Y.

Proposition 3.1.17. *Suppose X and Y are topological spaces and $f\colon X \to Y$ is a bijection. Then the following are equivalent:*

1. *f is a homeomorphism,*

2. *$A \subset X$ is open in X if and only if $f(A)$ is open in Y,*

3. *$A \subset X$ is closed in X if and only if $f(A)$ is closed in Y,*

4. *$A \subset X$ implies $f(\mathfrak{cl}[A]) = \mathfrak{cl}[f(A)]$,*

5. *For any net (x_λ) and point p in X, $x_\lambda \to p$ if and only if $f(x_\lambda) \to f(p)$.*

Proof. See Willard 2004 Theorem 7.9 for (1 - 4). (5) follows from Proposition 3.1.14. ■

3.2 Common Topologies

Here, we present some common types of topologies that will dominate our analysis. Recall that, for a topological space X, $\mathscr{O}(X)$ and $\mathscr{C}(X)$ refer to the open and closed sets of X, respectively.

Definition 3.2.1 (Metric, Metric space, Metric topology, Metrizable space)**.** A *metric* for a set X is a real-valued function $\rho\colon X \times X \to \mathbb{R}$ with the following properties holding for any $x, y, z \in X$:

1. $\rho(x, y) = \rho(y, x)$,

2. $\rho(x, y) = 0$ implies $x = y$,

3. $\rho(x, y) \leq \rho(x, z) + \rho(z, y)$.

The ordered pair (X, ρ) is called a *metric space*. The *metric topology* on X, denoted $\mathscr{O}_\rho(X)$, is the topology generated from the following basis

$$\mathscr{B}_\rho(X) \doteq \Big\{ \{x \in X : \rho(x, y) < \epsilon\} : y \in X, \epsilon > 0 \Big\}. \tag{3.4}$$

A topological space $(X, \mathcal{T})$ is *metrizable* if there exists a metric such that $\mathscr{O}_\rho(X) = \mathcal{T}$.

Remark 3.2.2. It follows trivially that for any subset $A \subset X$, $\mathscr{B}_\rho(A)$ is a basis for the subspace topology on A induced by X. So subspaces of metrizable spaces are metrizable.

Definition 3.2.3 (Standard topology of real numbers)**.** The mapping given by $(x, y) \mapsto |x - y|$ for each $x, y \in \mathbb{R}$ is called the *standard metric* and the topology it generates on $\mathbb{R}$ is called the *standard topology.*

Remark 3.2.4. Unless otherwise stated, it is assumed $\mathbb{R}$ is equipped with the standard topology.

Interestingly, different metrics may generate the same metric topologies. We will illustrate this concept in the following example.

Example 3.2.5. Let $d_1(x, y) \doteq |x - y|$ be the standard metric on $\mathbb{R}$. The arctan metric defined as $d_2(x, y) \doteq |\arctan x - \arctan y|$ for each $x, y \in \mathbb{R}$, is indeed a metric on $\mathbb{R}$ that generates the standard topology. To prove this, recall that for each $x \in \mathbb{R}$

$$\arctan x = \int_0^x \frac{1}{1+z^2}\, dz. \tag{3.5}$$

Therefore, we have

1. $d_2(x, y) = \left|\int_y^x \frac{1}{1+z^2}\, dz\right| = \left|-\int_x^y \frac{1}{1+z^2}\, dz\right| = \left|\int_x^y \frac{1}{1+z^2}\, dz\right| = d_2(y, x)$,
2. $\left|\int_y^x \frac{1}{1+z^2}\, dz\right| = 0$ implies $x = y$ since $\frac{1}{1+z^2} > 0$, and
3. $$\begin{aligned} d_2(x, y) &= |(\arctan x - \arctan z) - (\arctan y - \arctan z)| \\ &\leq |\arctan x - \arctan z| + |\arctan y - \arctan z| \\ &= d_2(x, z) + d_2(y, z), \end{aligned}$$

which demonstrates that d_2 is a metric on $\mathbb{R}$.

We now work to show that d_2 generates the standard topology; that is, $\mathscr{O}_{d_1}(\mathbb{R}) = \mathscr{O}_{d_2}(\mathbb{R})$. This task is accomplished by making use of Fact 3.1.3 twice.

Since $\frac{1}{1+z^2} \leq 1$ for each $z \in \mathbb{R}$, we have

$$
\begin{aligned}
|\arctan x - \arctan y| &= \left| \int_y^x \frac{1}{1+z^2}\, dz \right| && (3.6)\\
&\leq \left| \int_y^x 1\, dz \right| && (3.7)\\
&= |x - y|, && (3.8)
\end{aligned}
$$

that is, $d_2(x,y) \leq d_1(x,y)$ for each $x, y \in \mathbb{R}$. Therefore, we have

$$\{x \in \mathbb{R} : d_1(x,y) < \epsilon\} \subset \{x \in \mathbb{R} : d_2(x,y) < \epsilon\} \tag{3.9}$$

for each $y \in \mathbb{R}$ and $\epsilon > 0$.

Suppose $U \in \mathscr{B}_{d_2}(\mathbb{R})$ and $z \in U$. Then U has the following form

$$U = \{x \in \mathbb{R} : |\arctan x - \arctan y| < \epsilon\} \tag{3.10}$$

for some $y \in \mathbb{R}$ and $\epsilon > 0$. Since $z \in U$, we have

$$
\begin{aligned}
\{x \in \mathbb{R} : d_2(x,y) < \epsilon\} &\supset \{x \in \mathbb{R} : d_2(x,z) + d_2(z,y) < \epsilon\} && (3.11)\\
&= \{x \in \mathbb{R} : d_2(x,z) < \epsilon - d_2(z,y)\} && (3.12)\\
&\supset \{x \in \mathbb{R} : d_1(x,z) < \epsilon - d_2(z,y)\} && (3.13)
\end{aligned}
$$

which is a basis element of $\mathscr{B}_{d_1}(\mathbb{R})$. Therefore, $\mathscr{O}_{d_1}(\mathbb{R}) \supset \mathscr{O}_{d_2}(\mathbb{R})$ by Fact 3.1.3.

We will again use Fact 3.1.3; but first, we bring forth some more properties of arctan and d_2. Let $g\colon \mathbb{R} \to \mathbb{R}$ be defined as $g(z) \doteq \frac{1}{1+z^2}$. If $|x| < |y|$ we have

$$x^2 < y^2 \tag{3.14}$$

$$\implies \frac{1}{1+x^2} > \frac{1}{1+y^2} \tag{3.15}$$

$$\implies g(x) > g(y) \tag{3.16}$$

which shows that for any closed interval $[a, b]$, we have

$$\inf_{z\in[a,b]} g(z) = \min\{g(a), g(b)\} > 0. \tag{3.17}$$

It then follows that

$$d_2(x, y) = \left|\int_y^x g(z)\, dz\right| \tag{3.18}$$

$$\geq \left|\int_y^x \inf_{w\in[x,y]} \{g(w)\}\ dz\right| \tag{3.19}$$

$$= \left|\int_y^x \min\{g(x), g(y)\}\, dz\right| \tag{3.20}$$

$$= |(x - y) \min\{g(x), g(y)\}| \tag{3.21}$$

$$= |x - y\}| \min\{g(x), g(y)\} \tag{3.22}$$

$$= d_1(x, y) \min\{g(x), g(y)\} \tag{3.23}$$

$$\geq d_1(x, y) \min\{g(a), g(b)\} \tag{3.24}$$

that is, $d_2(x, y) \geq d_1(x, y) \min\{g(a), g(b)\}$ for each $x, y \in [a, b]$.

Letting $k_{y,\epsilon} = \min\{g(y - \epsilon), g(y + \epsilon)\}$, we have

$$\{x \in \mathbb{R} : k_{y,\epsilon} d_1(x, y) < \epsilon\} \supset \{x \in \mathbb{R} : d_2(x, y) < \epsilon\} \tag{3.25}$$

for each $y \in \mathbb{R}$ and $\epsilon > 0$.

We now return to finish off the proof using Fact 3.1.3. Suppose $V \in \mathscr{B}_{d_1}(\mathbb{R})$ and $z \in V$. Then V has the following form

$$V = \{x \in \mathbb{R} : |x - y| < \epsilon\} \tag{3.26}$$

for some $y \in \mathbb{R}$ and $\epsilon > 0$. Since $z \in V$, we have

$$\{x \in \mathbb{R} \colon d_1(x, y) < \epsilon\} \supset \{x \in \mathbb{R} : d_1(x, z) + d_1(z, y) < \epsilon\} \tag{3.27}$$

$$= \{x \in \mathbb{R} : d_1(x, z) < \epsilon - d_1(z, y)\} \tag{3.28}$$

$$\supset \{x \in \mathbb{R} : d_2(x, z) < k_{y,\epsilon}(\epsilon - d_1(z, y))\} \tag{3.29}$$

which is a basis element of $\mathscr{B}_{d_2}(\mathbb{R})$. Therefore, $\mathscr{O}_{d_1}(\mathbb{R}) \subset \mathscr{O}_{d_2}(\mathbb{R})$ by Fact 3.1.3, which completes the proof.

Definition 3.2.6 (Cauchy sequence, complete space)**.** Let (X, d) be a metric space. A sequence of points $(x_n)_{n=1}^{\infty}$ in X is said to be a *Cauchy sequence* in (X, d) if it has the property that given $\epsilon > 0$, there is an integer N such that $d(x_n, x_m) < \epsilon$ whenever $n, m \geq N$. A metric space is said to be *complete* if every Cauchy sequence converges.

In what follows, let $B(X, Y)$ and $C_B(X, Y)$ be the collection of bounded and bounded continuous functions, respectively, from X to Y.

Definition 3.2.7 (sup metric)**.** If (Y, d) is a metric space and $B(X, Y)$ is the bounded functions from the set X to Y, then the *sup metric* is defined on $B(X, Y)$ as

$$\rho(f, g) = \sup\{d(f(p), g(p)) : p \in X\}. \tag{3.30}$$

Proposition 3.2.8. *Let X be a topological space and let (Y, d) be a complete metric space. Then $C_B(X, Y)$ is closed and complete in $B(X, Y)$ equipped with the sup metric.*

Proof. See Munkres 2000 and explanation on page 270. ∎

Definition 3.2.9 (Hausdorff space)**.** A topological space X is called a *Hausdorff space* if the following is true: If $p, q \in X$, and $p \neq q$, then there exist sets $U, V \in \mathscr{O}(X)$ with $p \in U$ and $q \in V$ such that $U \cap V = \emptyset$.

Remark 3.2.10. Clearly, it follows that if $(X, \mathcal{T})$ is a Hausdorff space and $\mathcal{T} \subset \mathcal{T}'$, then $(X, \mathcal{T}')$ is a Hausdorff space too (assuming $\mathcal{T}'$ is still a valid topology on X).

Proposition 3.2.11. *Suppose (x_λ) is a net in a Hausdorff space X, such that $x_\lambda \to p \in X$. Then, $\lim_\lambda x_\lambda = p$ (i.e. it is the unique limit).*

Proof. Suppose $x_\lambda \to q \in X$. As X is Hausdorff, we can choose open neighbourhoods O_p, O_q such that $O_p \cap O_q = \emptyset$. As (x_λ) converges to both p and q, there are indices λ_p, λ_q such that $x_\lambda \in O_p$ and $x_\lambda \in O_q$ for all $\lambda > \max\{\lambda_p, \lambda_q\}$, which is a contradiction since the intersection is empty. ■

Proposition 3.2.12. *Let (X, ρ) be a metric space. The metric topology of X is a Hausdorff space.*

Proof. Suppose $p, q \in X$ and $p \neq q$. Let $\zeta = \rho(p, q)$ and note that $\zeta > 0$. Then choose open sets $U = \left\{x \in X : \rho(x, p) < \frac{\zeta}{2}\right\}$ and $V = \left\{x \in X : \rho(x, q) < \frac{\zeta}{2}\right\}$ and assume $z \in U$. We then have

$$\begin{aligned} \zeta = \rho(p, q) &\leq \rho(p, z) + \rho(q, z) && (3.31) \\ &< \frac{\zeta}{2} + \rho(q, z) && (3.32) \\ \implies \frac{\zeta}{2} &< \rho(q, z), && (3.33) \end{aligned}$$

so $z \notin V$. It follows that $U \cap V = \emptyset$. ■

Proposition 3.2.13. *Finite subsets of Hausdorff spaces are closed.*

Proof. See Munkres 2000 Theorem 17.8. ■

Definition 3.2.14 (Completely Regular, Tychonoff space)**.** A topological space X is called *completely regular* if and only if for each $A \in \mathscr{C}(X)$ and point $p \in X \setminus A$ there exists a continuous function $f \colon X \to [0, 1]$ such that $f|_A = 0$ and $f(p) = 1$. If X is also Hausdorff, then X is called a *Tychonoff space*.

Proposition 3.2.15. *Subspaces of Tychonoff spaces are Tychonoff.*

Proof. See Munkres 2000 Theorem 33.2 ■

Proposition 3.2.16. *Let (X, ρ) be a metric space. The metric topology of X is a Tychonoff space.*

Proof. By Proposition 3.2.12, we need only show that X is completely regular. Given $A \in \mathscr{C}(X)$ and $p \in X \setminus A$, we choose the continuous function $f_{A,p} \colon X \to \mathbb{R}$ defined as

$$f_{A,p}(x) \doteq 1 \wedge \frac{\inf\{\rho(x,y) : y \in A\}}{\inf\{\rho(p,y) : y \in A\}}, \tag{3.34}$$

which has the required properties that $x \in A$ implies $f_{A,p}(x) = 0$ and $f_{A,p}(p) = 1$. ■

3.3 Topologies Induced from Collections of Functions

Interestingly, we can define a topology in terms of a collection of functions. In fact, this is how one of the topologies on measures is defined in Chapter 6.

Definition 3.3.1. Let X be a set and $A \subset X$. For an index set I, let Y_i be a topological space and $f_i \colon X \to Y_i$ for each $i \in I$. The *topology induced by* $\mathcal{D} \doteq \{f_i\}_{i \in I}$ *on* $A \subset X$, denoted by $\mathscr{O}_{\mathcal{D}}(A)$, is the one generated from the subbasis

$$\mathscr{S}_{\mathcal{D}}(A) \doteq \left\{ f_i^{-1}(O) \cap A : O \in \mathscr{O}(Y_i), i \in I \right\}. \tag{3.35}$$

Remark 3.3.2. For any $O \in \mathscr{O}(Y_i)$ and $f_i \in \mathcal{D}$, we have $f_i^{-1}(O) \cap A \in \mathscr{S}_{\mathcal{D}}(A) \subset \mathscr{O}_{\mathcal{D}}(A)$. Therefore, $f|_A \in C((A, \mathscr{O}_{\mathcal{D}}(A)), Y_i)$. That is, for any collection of functions $\mathcal{D}$, we are able to generate a topology on A such that $\mathcal{D}|_A$ are continuous. Further, $\mathscr{O}_{\mathcal{D}}(A)$ is the coarsest topology such that $\mathcal{D}|_A$ are continuous.

Proposition 3.3.3. *Let X be a set and $A \subset X$. For an index set I, let Y_i be a topological space with subbasis $\mathcal{S}_i$ and $f_i \colon X \to Y_i$ for each $i \in I$. Letting $\mathcal{D} \doteq \{f_i\}_{i \in I}$, the following collection of sets*

$$\mathscr{S}_{\mathcal{D}}(A; \{\mathcal{S}_i\}_{i \in I}) \doteq \left\{ f_i^{-1}(O) \cap A : O \in \mathcal{S}_i, i \in I \right\} \tag{3.36}$$

is a subbasis on $A \subset X$ that generates $\mathscr{O}_{\mathcal{D}}(A)$.

Proof. First, we show $\mathscr{S}_{\mathcal{D}}(A;\{\mathcal{S}_i\}_{i\in I})$ is a subbasis for a topology on A.

$$\bigcup_{V\in\mathscr{S}_{\mathcal{D}}(A;\{\mathcal{S}\}_{i\in I})} V = \bigcup_{i\in I}\bigcup_{O\in\mathcal{S}_i} f_i^{-1}(O)\cap A \tag{3.37}$$

$$= \bigcup_{i\in I} f_i^{-1}\left(\bigcup_{O\in\mathcal{S}_i} O\right)\cap A \tag{3.38}$$

$$= \bigcup_{i\in I} f_i^{-1}(Y_i)\cap A \tag{3.39}$$

$$= \bigcup_{i\in I} X\cap A \tag{3.40}$$

$$= X\cap A \tag{3.41}$$

$$= A. \tag{3.42}$$

Next, we employ Fact 3.1.3 twice to show $\mathscr{S}_{\mathcal{D}}(A;\{\mathcal{S}_i\}_{i\in I})$ generates $\mathscr{O}_{\mathcal{D}}(A)$. Let $\mathcal{T}$ denote the topology generated from the subbasis $\mathscr{S}_{\mathcal{D}}(A;\{\mathcal{S}_i\}_{i\in I})$. By (3.3), $\mathscr{B}[\mathscr{S}_{\mathcal{D}}(A;\{\mathcal{S}_i\}_{i\in I})]$ and $\mathscr{B}[\mathscr{S}_{\mathcal{D}}(A;\{\mathscr{O}(Y_i)\}_{i\in I})]$ are bases for $\mathcal{T}$ and $\mathscr{O}_{\mathcal{D}}(A)$, respectively.

Since $\mathcal{S}_i\subset\mathscr{O}(Y_i)$, it is clear that $\mathscr{B}[\mathscr{S}_{\mathcal{D}}(A;\{\mathcal{S}_i\}_{i\in I})]\subset\mathscr{B}[\mathscr{S}_{\mathcal{D}}(A;\{\mathscr{O}(Y_i)\}_{i\in I})]$, so $\mathcal{T}\subset\mathscr{O}_{\mathcal{D}}(A)$.

Conversely, suppose $U\in\mathscr{B}[\mathscr{S}_{\mathcal{D}}(A;\{\mathscr{O}(Y_i)\}_{i\in I})]$ and $x\in U$. Then, for some $I_0\in\mathscr{R}_0[I]$, U has the form

$$U=\bigcap_{i\in I_0} f_i^{-1}(O_i)\cap A \tag{3.43}$$

where $O_i\in\mathscr{O}(Y_i)$ for each $i\in I_0$. For each $i\in I_0$, we have $f_i(x)\in O_i$; hence, there is a basis element $B_i\in\mathscr{B}[\mathcal{S}_i]$ such that $f_i(x)\in B_i\subset O_i$. By (3.3), $B_i=\bigcap_{V\in\mathscr{V}_i}V$ for some $\mathscr{V}_i\in\mathscr{R}_0[\mathcal{S}_i]$. It then follows that

$$x\in\bigcap_{i\in I_0} f_i^{-1}(B_i)=\bigcap_{i\in I_0} f_i^{-1}\left(\bigcap_{V\in\mathscr{V}_i}V\right)\cap A \tag{3.44}$$

$$=\bigcap_{i\in I_0}\bigcap_{V\in\mathscr{V}_i} f_i^{-1}(V)\cap A \tag{3.45}$$

$$\in\mathscr{B}[\mathscr{S}_{\mathcal{D}}(A;\{\mathcal{S}_i\}_{i\in I})], \tag{3.46}$$

so $\mathcal{T} \supset \mathscr{O}_{\mathcal{D}}(A)$ by Fact 3.1.3, which completes the proof. ■

An important case to consider for the topology $\mathscr{O}_{\mathcal{D}}(A)$ is when $\mathcal{D}$ is a collection of real-valued functions; that is, when $\mathcal{D} \subset \mathbb{R}^X$ where $\mathbb{R}$ is given the standard topology. Recall that a basis for the standard topology on $\mathbb{R}$ is given as

$$\mathscr{B}_\rho(\mathbb{R}) = \Big\{ \big\{ b \in \mathbb{R} : |a - b| < \epsilon \big\} : a \in \mathbb{R}, \epsilon > 0 \Big\}, \tag{3.47}$$

where ρ is the standard metric. Therefore, a subbasis for $\mathscr{O}_{\mathcal{D}}(A)$ is given as

$$\begin{aligned} \mathscr{S}_{\mathcal{D}}(A; \mathscr{B}_\rho(\mathbb{R})) &\doteq \big\{ f^{-1}(O) \cap A : O \in \mathscr{B}_\rho(\mathbb{R}), f \in \mathcal{D} \big\} && (3.48) \\ &= \Big\{ f^{-1}\big(\{ b \in \mathbb{R} : |a - b| < \epsilon \} \big) \cap A : a \in \mathbb{R}, \epsilon > 0, f \in \mathcal{D} \Big\} && (3.49) \\ &= \Big\{ \big\{ x \in X : |a - f(x)| < \epsilon \big\} \cap A : a \in \mathbb{R}, \epsilon > 0, f \in \mathcal{D} \Big\} && (3.50) \end{aligned}$$

which then generates the basis $\mathscr{B}[\mathscr{S}_{\mathcal{D}}(A; \mathscr{B}_\rho(\mathbb{R}))]$ given by the following sets

$$\Big\{ \bigcap_{i=1}^{n} \big\{ x \in X : |a_i - f_i(x)| < \epsilon_i \big\} \cap A : a_i \in \mathbb{R}, \epsilon_i > 0, f_i \in \mathcal{D}, n \in \mathbb{N} \Big\}. \tag{3.51}$$

Proposition 3.3.4. *Suppose X is a topological space, $A \subset X$, and $\mathcal{D} \subset \mathbb{R}^X$ where $\mathbb{R}$ is given the standard topology. The following collection of sets*

$$\Big\{ \big\{ x \in A : \max_{1 \le i \le n} |f_i(y) - f_i(x)| < \epsilon \big\} : y \in A, \epsilon > 0, f_i \in \mathcal{D}, n \in \mathbb{N} \Big\}, \tag{3.52}$$

denoted as $\mathscr{B}_{\mathcal{D}}(A)$, is a basis for $\mathscr{O}_{\mathcal{D}}(A)$.

Proof. We wish to employ Fact 3.1.2. Notice the following

$$\begin{aligned} &\big\{ x \in A : \max_{1 \le i \le n} |f_i(y) - f_i(x)| < \epsilon \big\} && (3.53) \\ &\qquad = \bigcap_{i=1}^{n} \big\{ x \in A : |f_i(y) - f_i(x)| < \epsilon \big\} && (3.54) \end{aligned}$$

$$= \bigcap_{i=1}^{n} \left\{ x \in X : |f_i(y) - f_i(x)| < \epsilon \right\} \cap A. \tag{3.55}$$

Therefore, $\mathscr{B}_{\mathcal{D}}(A) \subset \mathscr{B}[\mathscr{S}_{\mathcal{D}}(A; \mathscr{B}_\rho(\mathbb{R}))] \subset \mathscr{O}_{\mathcal{D}}(A)$ (where $\mathscr{S}_{\mathcal{D}}(A; \mathscr{B}_\rho(\mathbb{R}))$ and $\mathscr{B}[\mathscr{S}_{\mathcal{D}}(A; \mathscr{B}_\rho(\mathbb{R}))]$ are given by (3.48) and (3.51), respectively). Now suppose $U \in \mathscr{B}[\mathscr{S}_{\mathcal{D}}(A; \mathscr{B}_\rho(\mathbb{R}))]$ and $y \in U$. Then U has the form

$$U = \bigcap_{i=1}^{n} \left\{ x \in X : |a_i - f_i(x)| < \epsilon_i \right\} \cap A \tag{3.56}$$

for some $f_i \in \mathcal{D}$, $a_i \in \mathbb{R}$, $\epsilon_i > 0$, and $n \in \mathbb{N}$. Since $y \in U$, we have $f_i(y) \in (a_i - \epsilon_i, a_i + \epsilon_i)$ for each $i = 1, \ldots, n$. Define ϵ_y^* as follows

$$\epsilon_y^* \doteq \min_{1 \le i \le n} \{\epsilon_i - |a_i - f_i(y)|\}. \tag{3.57}$$

Then

$$V_y = \bigcap_{i=1}^{n} \left\{ x \in X : |f_i(y) - f_i(x)| < \epsilon_y^* \right\} \cap A \tag{3.58}$$

$$= \left\{ x \in A : \max_{1 \le i \le n} |f_i(y) - f_i(x)| < \epsilon_y^* \right\} \tag{3.59}$$

is a basis element of $\mathscr{B}_{\mathcal{D}}(A)$ such that $y \in V_y \subset U$. ■

3.3.1 Product Topology

The product topology is defined on Cartesian products and is generated via the projection functions. Notation for Cartesian products and projection functions was provided in Chapter 2.2.

Definition 3.3.5 (Product Topology, Product Space)**.** Let I be an index set and suppose Y_i is a topological space for each $i \in I$. We define the *product topology* on the Cartesian product $\prod_{i \in I} Y_i$ as $\mathscr{O}_{\mathcal{D}}\left(\prod_{i \in I} Y_i\right)$ where $\mathcal{D} = \{\pi_i\}_{i \in I}$[1]. A Cartesian product equipped with the product topology is called a *product space*.

[1] π_i are the projection functions defined in 2.2.

Remark 3.3.6. It is assumed that Cartesian products of topological spaces are product spaces unless otherwise noted.

Proposition 3.3.7. *Let X be a topological space; I be an index set; Y_i be a topological space for each $i \in I$; $f_i \colon X \to Y_i$ be a mapping for each $i \in I$; $\mathcal{D} = \{f_i : i \in I\}$; and recall the notation $\bigotimes \mathcal{D}$ defined by (2.4). Then the following statements are true:*

1. *For each $J \subset I$, $\pi_J \colon \prod_{i \in I} Y_i \to \prod_{i \in J} Y_i$ is continuous.*

2. *$\bigotimes \mathcal{D} \colon X \to \prod_{i \in I} Y_i$ is continuous if and only if f_i is continuous for each $i \in I$.*

3. *$\bigotimes \mathcal{D} \colon X \to \prod_{i \in I} Y_i$ is continuous at $x \in X$ if and only if f_i is continuous at $x \in X$ for each $i \in I$.*

4. *$\bigotimes \mathcal{D} \colon X \to \prod_{i \in I} Y_i$ is an embedding if and only if it is injective and $\mathscr{O}(X) = \mathscr{O}_{\mathcal{D}}(X)$.*

5. *A net (x_λ) in $\prod_{i \in I} Y_i$ converges to p if and only if for each $i \in I$, $\pi_i(x_\lambda) \to \pi_i(p)$ in Y_i.*

6. *If Y_i is Hausdorff for each $i \in I$, then $\prod_{i \in I} Y_i$ is Hausdorff.*

7. *If Y_i is Tychonoff for each $i \in I$, then $\prod_{i \in I} Y_i$ is Tychonoff.*

Proof. See Munkres 2000 Theorems 18.1 and 19.6 for (1-3), 19.4 for (6), and 33.2 for (7). See Willard 2004 Theorem 8.12 for (4) and 11.9 for (5). ∎

Proposition 3.3.8. *Suppose $N \in \mathbb{N} \cup \{\infty\}$. The following is a metric on $\mathbb{R}^N$ that generates the product topology:*

$$\rho(x, y) \mapsto \sum_{n=1}^{N} 2^{-n} \big(|x_n - y_n| \wedge 1 \big) \tag{3.60}$$

for each $x, y \in \mathbb{R}^{\mathbb{N}}$. That is, $\mathbb{R}^N$ is metrizable.

Proof. See the proof of Willard 2004 Theorem 24.11. ∎

3.4 Sequential Spaces

Sequential spaces are topological spaces whose properties are often reduced to the checking of sequences, rather than having to deal with nets. As we shall see, a sequential space can be generated from any topology and, in fact, this is what we do in Chapter 6 when defining a topology on a set of measures. The generated sequential space shares the same convergent sequences as the original space. Some nice sources for study of sequential spaces can be found in Vermeeren 2010 and Antosik, Boehme, and Mohanadi 1985.

Definition 3.4.1 (Eventually in, Sequentially Open, Sequential Space)**.** Suppose X is a topological space, $A \subset X$, and (x_n) is a sequence in X. We say (x_n) is *eventually in* A if there is an $N \in \mathbb{N}$ such that $n \geq N$ implies $x_n \in A$. The set A is said to be *sequentially open* if every sequence in X that converges to a point in A is eventually in A. X is a *sequential space* if every sequentially open set is open.

Proposition 3.4.2. *Open sets are sequentially open.*

Proof. Suppose A is open in a topological space X, $p \in A$, and $x_n \to p$ in X. Clearly, A is a neighborhood of p. So by the definition of net convergence (x_n) is eventually in A. ■

Definition 3.4.3 (Sequentially Continuous)**.** Given topological spaces X and Y, a function $f : X \to Y$ is *sequentially continuous* if for any sequence (x_n) and point $p \in X$ such that $x_n \to p$ we have $f(x_n) \to f(p)$.

One of the nice properties about sequential spaces is that continuity is the same as sequential continuity as we see next.

Proposition 3.4.4. *Let X be a sequential space and Y be a topological space. Then $f : X \to Y$ is continuous if and only if f is sequentially continuous.*

Proof. Sequences are nets, so continuity of f implies sequential continuity by Proposition 3.1.14. In the reverse direction, let A be open in Y. We must show $f^{-1}(A)$ is sequentially open in X. Let (x_n) be a sequence in X converging to $p \in f^{-1}(A)$. As f is sequentially continuous, we see that $(f(x_n))$ is a sequence in Y converging to $f(p) \in A$. By Proposition 3.4.2, A is sequentially open, so $(f(x_n))$ is eventually in A, which implies (x_n) is eventually in $f^{-1}(A)$, so $f^{-1}(A)$ is sequentially open. ■

The collection of sequentially open sets defines a finer topology than the original space.

Proposition 3.4.5. *Suppose $(X, \mathcal{T})$ is a topological space. The collection of sequentially open sets, denoted $\mathcal{T}_s$, is a topology on X. Further $\mathcal{T} \subset \mathcal{T}_s$.*

Proof. First, we show $\mathcal{T}_s$ is a topology on X. Clearly, $\{\emptyset, X\} \subset \mathcal{T}_s$. Suppose $\mathcal{C} \subset \mathcal{T}_s$ and let $A_{\mathcal{C}} = \bigcup_{A \in \mathcal{C}} A$. If $p \in A_{\mathcal{C}}$, then there is some $A_0 \in \mathcal{C}$ such that $p \in A_0$. Therefore if (x_n) converges to p, then it is eventually in A_0 and $A_{\mathcal{C}}$. Next, suppose $\mathcal{C}_0 \in \mathscr{R}_0(\mathcal{T}_s)$ and let $A_{\mathcal{C}_0} = \bigcap_{A \in \mathcal{C}_0} A$. If $p \in A_{\mathcal{C}_0}$, then $p \in A$ for each $A \in \mathcal{C}_0$. Therefore if (x_n) converges to p, then we can define N_A such that $x_n \in A$ for each $n \geq N_A$ and choose $N = \max\{N_A : A \in \mathcal{C}_0\}$ implying $x_n \in A_{\mathcal{C}_0}$ when $n \geq N$.

Open sets are sequentially open by Proposition 3.4.2, so $\mathcal{T} \subset \mathcal{T}_s$. ■

Definition 3.4.6 (Generating a sequential space)**.** Given a topological space $(X, \mathcal{T})$, we denote $\mathfrak{Q}(X, \mathcal{T})$ as the sequential topology on X generated from the topology $\mathcal{T}$ according to Proposition 3.4.5. That is, $\mathfrak{Q}(X, \mathcal{T})$ is the collection of sequentially open sets with respect to $(X, \mathcal{T})$.

Remark 3.4.7. It should be clear that a topological space $(X, \mathcal{T})$ is a sequential space if and only if $\mathcal{T} = \mathfrak{Q}(X, \mathcal{T})$.

The next result shows that a topological space X and its generated sequential space share the same convergent sequences.

Proposition 3.4.8. *Suppose $(X, \mathcal{T})$ is a topological space and let $\mathcal{T}_s = \mathfrak{Q}(X, \mathcal{T})$. Then, a sequence (x_n) converges to $p \in X$ with respect to $(X, \mathcal{T})$ if and only if it converges to p with respect to $(X, \mathcal{T}_s)$.*

Proof. Proposition 3.4.5 says $\mathcal{T} \subset \mathcal{T}_s$, so convergence in $(X, \mathcal{T}_s)$ implies convergence in $(X, \mathcal{T})$ by Definition 3.1.11. Conversely, suppose (x_n) converges to p with respect to $(X, \mathcal{T})$ and let $A_p \in \mathcal{T}_s$ be an open neighborhood of p

with respect to $(X, \mathcal{T}_s)$. Since $A_p \in \mathcal{T}_s$, it is a sequential open set with respect to $(X, \mathcal{T})$ by Definition 3.4.6, implying (x_n) is eventually in A_p. Therefore, $x_n \to p$ with respect to $(X, \mathcal{T}_s)$. ■

Now, we would like to better understand when we are working with a sequential space. First we define a sequential limit point, which is analogous to nets except for the sequential setting.

Definition 3.4.9 (Sequential Limit Point)**.** Let X be a topological space and $A \subset X$. Then, p is a *sequential limit point* of A if there exists a sequence in A converging to p.

Remark 3.4.10. Every sequence is a net, so it should be clear that sequential limit points are always limit points.

Proposition 3.4.11. *Let X be a first-countable*[2] *topological space. Then,*

1. *A point $p \in X$ is a limit point of $A \subset X$ if and only if p is a sequential limit point of A.*

2. *X is a sequential space.*

Proof. See Munkres 2000 Theorem 30.1 for (1.). We show (1.) implies (2.). By Definition 3.4.1 of a sequential space, we must show sequential sets are open. Let $B \subset X$ be sequentially open and we argue that $X \setminus B$ is closed. Let p be a limit point of $X \setminus B$, so by (1.) there must be a sequence (x_n) in $X \setminus B$ converging to $p \in X$. If $p \in B$, then (x_n) must be eventually in B; however, $x_n \in X \setminus B$ for all n, so a contradiction has been reached. Therefore, $X \setminus B$ contains all of its limit points, which is to say that it is closed. ■

Proposition 3.4.12. *Metrizable spaces are first-countable.*

Proof. See Munkres 2000 pages 130/131. ■

The previous two results turn out to be quite important. It is often the case in universal approximation that we are at least working with a metric space, so Propositions 3.4.12 and 3.4.11 tell us we can check topological properties

[2] First-countable was defined in Definition 3.1.6

using sequences instead of nets. For example, Proposition 3.4.4 says continuity is reduced to sequential continuity.

Our universal approximation results rely on the construction of homeomorphisms, so it is natural to wonder if homeomorphisms preserve sequential spaces. The answer is yes and is shown in the next result.

Proposition 3.4.13. *Suppose $(X, \mathcal{T}_X)$ is a sequential space and $(Y, \mathcal{T}_Y)$ is a topological space. If $h\colon X \to Y$ is a homeomorphism, then Y is a sequential space.*

Proof. Suppose $A \subset Y$ is sequentially open. We must show $A \in \mathcal{T}_Y$. Let (x_n) be a sequence converging to $p \in f^{-1}(A)$. By Proposition 3.1.17 (5), it follows then that $f(x_n)$ is a sequence converging to $f(p) \in A$, so $f(x_n)$ is eventually in A since A is sequentially open. Therefore, there is some $N \in \mathbb{N}$ such that $f(x_n) \in A$ when $n \geq N$, which implies $x_n \in f^{-1}(A)$ when $n \geq N$. That is, x_n is eventually in $f^{-1}(A)$, so $f^{-1}(A)$ is sequentially open and so it is open (as X is a sequential space). It then follows from Proposition 3.1.17 (2) that A is open. ∎

The next result is rather important for homeomorphism construction. It is often easier to confirm continuity in the forward direction of a bijective function than in the reverse direction. The following result makes the reverse direction a bit easier by reducing it to checking for sequential continuity.

Proposition 3.4.14. *Suppose X is a topological space; $\mathcal{M} \subset \mathbb{R}^X$ is countable; and $\bigotimes \mathcal{M}\colon X \to \bigotimes \mathcal{M}(X)$ is a bijection. Then $[\bigotimes \mathcal{M}]^{-1}$ is continuous if and only if, for any sequence (x_n) in X, we have $g(x_n) \to g(p)$ for all $g \in \mathcal{M}$ implies $x_n \to p$ in X.*

Proof. Since $\mathcal{M}$ is countable, Proposition 3.3.8 says $\mathbb{R}^{\mathcal{M}}$ is metrizable and so is $\bigotimes \mathcal{M}(X)$ as a subspace, so it is first-countable by Proposition 3.4.12. It then follows from Proposition 3.4.11 that we only need $[\bigotimes \mathcal{M}]^{-1}$ to be sequentially continuous to claim it is continuous. Suppose (y_n) is a sequence in $\bigotimes \mathcal{M}(X)$ converging to q. Since $\bigotimes \mathcal{M}$ is a bijection, define $x_n \doteq [\bigotimes \mathcal{M}]^{-1}(y_n)$ and $p \doteq [\bigotimes \mathcal{M}]^{-1}(q)$, so (x_n) is the corresponding sequence in X. By Proposition 3.3.7 (5), (y_n) converges to q in $\bigotimes \mathcal{M}(X)$ if and only if $(\pi_g(y_n))$ converges to

$\pi_g(q)$ for each $g \in \mathcal{M}$. However, $\pi_g(y_n) = g(x_n)$ and $\pi_g(q) = g(p)$. So we have $y_n \to q$ in $\bigotimes \mathcal{M}(X)$ if and only if $g(x_n) \to g(p)$ for all $g \in \mathcal{M}$, which is to say that $y_n \to q$ in $\bigotimes \mathcal{M}(X)$ implies $x_n \to p$ in X. So $[\bigotimes \mathcal{M}]^{-1}$ is sequentially continuous and hence is continuous. ■

Chapter 4

Point Separation

In the introductory chapter, we presented some universal approximation results on finite dimensional topological spaces by utilizing homeomorphisms. Our plan of action for handling infinite dimensional spaces is no different; however, the homeomorphisms becomes more complicated. This section provides many results about the necessary and sufficient conditions required to construct homeomorphisms on general Tychonoff spaces.

If X is a topological space, then we wish to identify a collection of real valued functions $\mathcal{M}$ such that $\bigotimes \mathcal{M} \colon X \to \mathbb{R}^{\mathcal{M}}$ is an embedding into a compact subset of $\mathbb{R}^{\mathcal{M}}$. As we shall see, our goal is achieved when $\mathcal{M}$ is said to both separate points (s.p.) and strongly separate points (s.s.p.) on X. In the case where $\mathcal{M}$ is countable, then the s.s.p. condition may be confirmed by checking if $[\bigotimes \mathcal{M}]^{-1}$ is sequentially continuous, which we instead say $\mathcal{M}$ determines sequential point convergence.

4.1 (Strong) Separation of Points

The main goal of this section is to introduce the strong separation of points (s.s.p.) property; however, first we need to know what it means to separate points (s.p.).

Definition 4.1.1 (Separation of Points)**.** Let $\mathcal{M}$ be a class of functions mapping A to B. If for every $x, y \in A$ with $x \neq y$ there exists $f \in \mathcal{M}$ such that

$f(x) \neq f(y)$, then $\mathcal{M}$ is said to *separate points* (s.p.).

The separating points property is going to be important for homeomorphism construction as its presence implies a bijection.

Proposition 4.1.2. *Let X be a topological space. If $\mathcal{M} \subset \mathbb{R}^X$ separates points on X, then:*

1. *$\bigotimes \mathcal{M} \colon X \to \bigotimes \mathcal{M}(X)$ is a bijection,*
2. *$\mathcal{M} \subset C(X)$ implies X is a Hausdorff space,*
3. *$\mathcal{M} \subset \mathcal{N} \subset \mathbb{R}^X$ implies $\mathcal{N}$ s.p. on X, and*
4. *$A \subset X$ implies $\mathcal{M}|_A$ s.p. on A.*

Proof. (1.) Suppose $p \neq q \in X$. $\mathcal{M}$ s.p. on X implies there is some $g \in \mathcal{M}$ such that $g(p) \neq g(q)$. It follows then that $\pi_g \circ \bigotimes \mathcal{M}(p) \neq \pi_g \circ \bigotimes \mathcal{M}(q)$, so $\bigotimes \mathcal{M}(p) \neq \bigotimes \mathcal{M}(q)$. That is, for each $r \in \bigotimes \mathcal{M}(X)$ there is a unique $p \in X$ such that $\bigotimes \mathcal{M}(p) = r$, so $\bigotimes \mathcal{M}$ is a bijection.

(2.) For any pair of points $p, q \in X$, there is some real valued continuous function, say f, such that $f(p) \neq f(q)$. Let $r = \frac{|f(p)-f(q)|}{3} > 0$, then sets $(f(p) - r, f(p) + r)$ and $(f(q) - r, f(q) + r)$ are open balls containing $f(p)$ and $f(q)$, respectively, and have empty intersection. Since f is continuous, $f^{-1}[(f(p) - r, f(p) + r)] \ni p$ and $f^{-1}[(f(p) - r, f(p) + r)] \ni q$ are open in X and have empty intersection due to Fact 2.1.1. It then follows that X is Hausdorff by Definition 3.2.9.

(3.) For $p \neq q \in X$, there is $g \in \mathcal{M} \subset \mathcal{N}$ such that $g(p) \neq g(q)$.

(4.) For $p \neq q \in A \subset X$, there is $g \in \mathcal{M}$ such that $g(p) \neq g(q)$, so $g|_A(p) \neq g|_A(q)$. ■

Now we define strong separation of points and provide some basic relevant properties.

Definition 4.1.3 (Strong Separation of Points)**.** Let $(X, \mathcal{T})$ be a topological space and $\mathcal{M} \subset \mathbb{R}^X$ be a collection of real valued mappings. Then $\mathcal{M}$ *strongly* *separates points* (s.s.p.) if, for every $x \in X$ and neighborhood O_x of x, there is a finite collection $\{g_1, \ldots, g_k\} \in \mathscr{R}_0(\mathcal{M})$ such that

$$\inf_{y \notin O_x} \max_{1 \leq i \leq k} |g_i(y) - g_i(x)| > 0. \tag{4.1}$$

Proposition 4.1.4. *Let X be a topological space; $\mathcal{M} \subset \mathbb{R}^X$ be a collection of real valued functions on X; and A be a subspace of X. The following properties hold:*

1. *If $\mathcal{M}$ s.s.p. and $\mathcal{M} \subset \mathcal{N} \subset \mathbb{R}^X$, then $\mathcal{N}$ also s.s.p.*

2. *If $\mathcal{M}$ s.s.p. on X, then $\mathcal{M}|_A$ s.s.p. on A.*

3. *Let $\mathcal{T}_1 \subset \mathcal{T}_2$ be topologies on X. If $\mathcal{M}$ s.s.p. on $(X, \mathcal{T}_2)$, then $\mathcal{M}$ s.s.p. on $(X, \mathcal{T}_1)$.*

Proof. First, we show 1. $\mathcal{M}$ s.s.p., so given $x \in X$ and neighborhood O_x there is a finite collection $\{g^1, \ldots, g^k\} \subset \mathcal{M} \subset \mathcal{N}$ that satisfy (4.1), hence $\mathcal{N}$ s.s.p.

Next, we show 2. Let $\mathcal{T}_X$ and $\mathcal{T}_A$ denote the topologies on X and A, respectively where $\mathcal{T}_A \doteq \{A \cap S \mid S \in \mathcal{T}_X\}$ is the subspace topology. Fix $x \in A$ and let $O_x^A \doteq A \cap O_x^X$ be a neighborhood of x with respect to $\mathcal{T}_A$, where $O_x^X \in \mathcal{T}_X$. (2) follows from observing that:

$$\inf_{y \notin O_x^A} \max_{1 \leq i \leq k} \Big| g_i|_A(y) - g_i|_A(x) \Big| \geq \inf_{y \notin O_x^X} \max_{1 \leq i \leq k} \Big| g_i(y) - g_i(x) \Big| \tag{4.2}$$

since $A \setminus O_x^A = A \setminus O_x^X \subset X \setminus O_x^X$.

Finally, we prove 3. If O_x^1 is a neighborhood of x with respect to topology $\mathcal{T}_1$, then there exists a $P_x^1 \in \mathcal{T}_1$ such that $O_x^1 \supset P_x^1$. Since $\mathcal{T}_1 \subset \mathcal{T}_2$, we have $P_x^1 \in \mathcal{T}_2$ and O_x^1 is also a neighborhood of x with respect to topology $\mathcal{T}_2$. Therefore, (4.1) is satisfied trivially for O_x^1, and so $\mathcal{M}$ s.s.p. on $(X, \mathcal{T}_1)$. ■

The next proposition provides an alternative means of defining the strong separation of points property, that is particularly useful when the topology of some space is defined by a collection of real valued functions like was shown in Definition 3.3.1.

Proposition 4.1.5. *Suppose $(X, \mathcal{T})$ is a topological space and $\mathcal{M} \subset \mathbb{R}^X$ is a collection of real valued functions on X. Then $\mathcal{M}$ s.s.p. on $(X, \mathcal{T})$ if and only if $\mathcal{T} \subset \mathscr{O}_{\mathcal{M}}(X)$.*

Proof. By Proposition 3.3.4, $\mathscr{B}_{\mathcal{M}}(X)$ is a basis for $\mathscr{O}_{\mathcal{M}}(X)$ with sets of the form

$$B_{q,\epsilon}(\mathcal{M}_0) \doteq \left\{p \in X : \max_{f \in \mathcal{M}_0} |f(q) - f(p)| < \epsilon\right\} \qquad \mathcal{M}_0 \in \mathscr{R}_0(\mathcal{M}). \tag{4.3}$$

If $\mathcal{M}$ s.s.p., then for each $O_q \in \mathcal{T}$ there exist $\mathcal{M}_0 \in \mathscr{R}_0(\mathcal{M})$ and $\epsilon > 0$ such that $\inf_{p \notin O_q}\{\max_{f \in \mathcal{M}_0} |f(q) - f(p)|\} = \epsilon$, which implies $B_{q,\epsilon}(\mathcal{M}_0) \subset O_q$. It then follows from Fact 3.1.3 that $\mathcal{T} \subset \mathscr{O}_{\mathcal{M}}(X)$.

Conversely, now assume $\mathcal{T} \subset \mathscr{O}_{\mathcal{M}}(X)$. For each neighborhood N_q of q, there is an $O_q \in \mathcal{T}$ such that $q \in O_q \subset N_q$. By assumption, $O_q \in \mathcal{T}$ implies $O_q \in \mathscr{O}_{\mathcal{M}}(X)$, so there is $\mathcal{M}_0 \in \mathscr{R}_0(\mathcal{M})$ and $\epsilon > 0$ such that $B_{q,\epsilon}(\mathcal{M}_0) \subset O_q$. It then follows that $\inf_{p \notin O_q}\{\max_{f \in \mathcal{M}_0} |f(q) - f(p)|\} \geq \epsilon > 0$, so $\mathcal{M}$ s.s.p. ∎

Now we provide a result that shows how s.p. relates to s.s.p. via a spaces Hausdorff property.

Proposition 4.1.6. *Let X be a topological space, $A \subset X$ be non-empty, and $\mathcal{D} \subset \mathbb{R}^X$. Then, the following statements are true:*

(a) If $\{\{x\} : x \in A\} \subset \mathscr{C}(X)$, especially if A is a Hausdorff subspace of X, then $\mathcal{D}$ strongly separating points on A implies $\mathcal{D}$ separating points on A.

(b) $\mathcal{D}$ separates points on A if and only if $(A, \mathscr{O}_{\mathcal{D}}(A))$ is a Hausdorff space.

Proof. We provide the proof from Dong and Kouritzin 2020 Proposition 9.2.1.

(a) The Hausdorff property of $(A, \mathscr{O}_X(A))$ (if any) implies $\{\{x\} : x \in A\} \subset \mathscr{C}(A, \mathscr{O}_X(A))$ by Proposition 3.2.13. We then prove (a) by contradiction. If $\mathcal{D}$ fails to separate points on A, then there exist distinct $x, y \in A$ such that $\bigotimes \mathcal{D}(x) = \bigotimes \mathcal{D}(y)$. Since $\{y\}$ is a closed set and $\mathcal{D}$ strongly separates points on A, there exist $\mathcal{D}_x \in \mathscr{R}_0(\mathcal{D})$ and $\epsilon \in (0, \infty)$ such that $y \in \{z \in A : \max_{f \in \mathcal{D}_x} |f(x) - f(z)| < \epsilon\} \subset A \setminus \{y\}$, Which is a contradiction.

(b - Sufficiency) follows by (a) (with $A = (X, \mathscr{O}_{\mathcal{D}}(X))$).

(b - Necessity) Let $x_1, x_2 \in A$ be distinct. Since $\mathcal{D}$ separates points on A, there exists an $f \in \mathcal{D}$ such that $\epsilon_0 \doteq |f(x_1) - f(x_2)| > 0$. Then, we define $O_i \doteq \{z \in A : |f(x_i) - f(z)| < \frac{\epsilon_0}{3}\} \in \mathscr{O}_{\mathcal{D}}(A)$ for each $i = 1, 2$ and observe that $x_1 \in O_1$, $x_2 \in O_2$ and $O_1 \cap O_2 = \emptyset$. ■

The last result for this section, shows that continuous functions that s.s.p. and s.p. only occur when working in a Tychonoff space. The result also is the reason why our universal approximation results are limited to Tychonoff spaces.

Proposition 4.1.7. *Let X be a topological space. Then, the following statements are equivalent:*

1. *X is a Tychonoff space.*
2. *$C(X)$ separates and strongly separates points on X.*
3. *$C_B(X)$ separates and strongly separates points on X.*

Remark 4.1.8. If $\mathcal{M} \subset C(X)$ s.p and s.s.p., then $C(X)$ s.p. and s.s.p. implying X is Tychonoff.

Proof. We provide the proof from Dong and Kouritzin 2020 Proposition 9.3.1.

($1 \to 2$) Suppose $O_p \in \mathscr{O}(X)$ is an open neighborhood of $p \in X$ and let $A = X \setminus O_p \in \mathscr{C}(X)$ so $p \notin A$. Since E is a Tychonoff space, there is a $f_{A,p} \in C(X; [0, 1])$ such that $f_{A,p}|_A = 0$ and $f_{A,p}(p) = 1$ by Definition 3.2.14. It then follows that $\{q \in X : |f_{A,p}(p) - f_{A,p}(q)| < \epsilon\} \subset O_p$ for each $\epsilon \in (0, 1)$. We then have $\mathscr{O}(X) \subset \mathscr{O}_{C(X)}(X)$ by Fact 3.1.3, which implies $C(X)$ s.s.p. on X by Proposition 4.1.5. Proposition 4.1.6 (a) implies $C(X)$ s.p. on X.

($2 \to 3$) Letting ρ be the standard metric on $\mathbb{R}$, it follows from Proposition 3.3.3 that, for any collection of real valued functions on X,
$\mathscr{S}_{\mathcal{D}}(X; \mathscr{B}_\rho(\mathbb{R})) = \{\{q \in X : a - r < f(q) < a + r\} : a \in \mathbb{R}, r > 0, f \in \mathcal{D}\}$ is a subbasis for $\mathscr{O}_{\mathcal{D}}(X)$. For each $f \in C(X)$, define the function
$g_{r,a,f} \doteq (f \vee (a - r)) \wedge (a + r)$, which is bounded, continuous, and satisfies $\{q \in X : a - r < g_{r,a,f}(q) < a + r\} = \{q \in X : a - r < f(q) < a + r\}$. So

$\mathscr{O}_{C_B(X)}(X) \supset \mathscr{O}_{C(X)}(X) \supset \mathscr{O}(X)$; and hence, $C_B(X)$ s.s.p. on X by Proposition 4.1.5. By Proposition 4.1.2, X is Hausdorff, so Proposition 4.1.6 (a) implies $C_B(X)$ s.p. on X.

($3 \to 1$) Pick $p \in X$ and $A \in \mathscr{C}(X)$ such that $p \notin A$. Since $C_B(X)$ s.s.p. on X, there exist $\mathcal{M}_0 \in \mathscr{R}_0(C_B(X))$ and $\epsilon > 0$ such that

$$p \in \left\{ q \in X : \max_{f \in \mathcal{M}_0} |f(p) - f(q)| < \epsilon \right\} \subset X \setminus A, \tag{4.4}$$

from which it follows that

$$h(q) \doteq 1 - \min\left\{ 1, \frac{\max_{f \in \mathcal{M}_0} |f(p) - f(q)|}{\epsilon} \right\} \tag{4.5}$$

is a continuous function from X to $[0, 1]$ such that $h|_A = 0$ and $h(p) = 1$. Hence, by Definition 3.2.14, X is a Tychonoff space. ∎

4.2 Homeomorphisms

Recall the main point of this chapter was to construct homeomorphisms, which we said were related to s.s.p. and s.p. properties. We have already seen that s.p. implies a bijection and, in addition, s.s.p. implies the bijection has a continuous inverse, which we will see shortly. First, there is yet another property to consider, which we define next.

Definition 4.2.1 (Determines Point Convergence, Determines Sequential Point Convergence)**.** Let X be a topological space and $\mathcal{M} \subset \mathbb{R}^X$. We say $\mathcal{M}$ *determines point convergence on* X if and only if, for any net (x_λ) and point p in X, we have $g(x_\lambda) \to g(p)$ for each $g \in \mathcal{M}$ implies $x_\lambda \to p$ in X. Similarly, we say $\mathcal{M}$ *determines sequential point convergence on* X if and only if, for any sequence (x_n) and point p in X, we have $g(x_n) \to g(p)$ for each $g \in \mathcal{M}$ implies $x_n \to p$ in X.

Proposition 4.2.2. *Let X be a topological space and $\mathcal{M} \subset \mathbb{R}^X$. Then, $\mathcal{M}$ determines point convergence on X implies $\mathcal{M}$ determines sequential point convergence on X.*

Proof. By Definition 3.1.10, sequences are nets. Hence, if $\mathcal{M}$ determines point convergence on X, (x_n) is a sequence in X, $p \in X$, and $g(x_n) \to g(p)$ for each $g \in \mathcal{M}$; then $x_n \to p$. So $\mathcal{M}$ determines sequential point convergence on X. ■

Below is a culmination of the different ways in which homeomorphisms are related to s.s.p. and determining point convergence properties. We will make use of this proposition repeatedly in future chapters.

Proposition 4.2.3. *Suppose X be a Hausdorff space or $C(X)$ s.p. on X and let $\mathcal{M} \subset \mathbb{R}^X$. Statements (1 - 3) are equivalent and imply (4). If $\mathcal{M}$ is countable then (1 - 4) are equivalent.*

1. *$\bigotimes \mathcal{M} \colon X \to \bigotimes \mathcal{M}(X)$ has a continuous inverse. If $\mathcal{M} \subset C(X)$, then $\bigotimes \mathcal{M} \colon X \to \bigotimes \mathcal{M}(X)$ is a homeomorphism.*
2. *$\mathcal{M}$ s.s.p. on X.*
3. *$\mathcal{M}$ determines point convergence on X.*
4. *$\mathcal{M}$ determines sequential point convergence on X.*

Proof. $(1 \leftrightarrow 2)$ See Blount and Kouritzin 2010 Lemma 1.

$(2 \leftrightarrow 3)$ See Blount and Kouritzin 2010 Lemma 4.

$(3 \to 4)$ Follows directly from Definition 4.2.1 as all sequences are nets.

$(4 \to 1$, $\mathcal{M}$ countable$)$ Follows from Proposition 3.4.14. ■

The next proposition provides a way in which we can tell if there is a countable collection that strongly separates points. Note that a topological space whose topology can be generated from a topological base of countable size is said to have a *countable base*.

Proposition 4.2.4. *If $(X, \mathcal{T})$ has a countable basis and $\mathcal{M} \subset C(X)$ s.s.p., then there is a countable collection $\{g_i\}_{i=0}^{\infty} \subset \mathcal{M}$ that s.s.p. Moreover, $\{g_i\}_{i=0}^{\infty}$ can be taken closed under either multiplication or addition if $\mathcal{M}$ is.*

Proof. Proven by Blount and Kouritzin 2010 Lemma 2. ■

Recall that Proposition 4.1.7 tells us the bounded continuous functions s.s.p. on Tychonoff spaces. Hence, if we have a Tychonoff space with a countable base, then we can be sure there is a countable collection of bounded continuous functions that are closed under multiplication and s.s.p..

4.3 Compactification

Compact sets are often nicer to deal with than perhaps our original space. This section provides some propositions telling us when we can view the original space as a subspace of a compact set. First we start with an interesting result demonstrating the "niceness" of compact sets.

Proposition 4.3.1. *Let X be a compact space and $\mathcal{M} \subset C(X)$. Then, X is a Hausdorff space and $\mathcal{M}$ strongly separates points on X if and only if $\mathcal{M}$ separates points on X.*

Proof. We provide the proof from Dong and Kouritzin 2020 Lemma 9.2.4.

Due to Proposition 4.1.2 (2) and Proposition 4.1.6 (a), we need only show $\mathcal{M}$ s.p. on the compact X implies $\mathcal{M}$ s.s.p. on X. Further, by Proposition 4.2.3, we need only show $\mathcal{M}$ s.p. implies $\mathcal{M}$ determines point convergence on X.

Let (x_λ) be a net such that $\lim_\lambda f(x_\lambda) = f(x)$ for all $f \in \mathcal{M}$. Since X is compact, Proposition 3.1.15 applies, and so there exists a subnet (x_{λ_μ}) and $p \in X$ such that $x_{\lambda_\mu} \to p$. Further, for each $f \in \mathcal{M}$, $\lim_\lambda f(x_\lambda) = f(x)$ implies $\lim_\mu f(x_{\lambda_\mu}) = f(x)$ and, since $x_{\lambda_\mu} \to p$ and f is continuous, we also have $\lim_\mu f(x_{\lambda_\mu}) = f(x) = f(p)$. As $\mathcal{M}$ s.p. on X, we can conclude $x = p$, so $x_{\lambda_\mu} \to x$. As every subnet has a subnet converging to x, we have shown $x_\lambda \to x$, so $\mathcal{M}$ determines point convergence and s.s.p. on X. ■

So when X is compact, it is enough to show $\mathcal{M} \subset C(X)$ s.p. on X in order to get the s.s.p. property, which is often much easier than trying to show $\mathcal{M}$ s.s.p. directly.

Now we work to show when we are in a subspace of a compact space.

Definition 4.3.2 (Compactification, Equivalent, Unique)**.** Suppose X is a topological subspace of a compact space S. Then, S is called a *compactification* of X if $\bar{X} = S$. If S and T are compactifications of X, then we say they are *equivalent up to homeomorphism* if there exists a homeomorphism $h\colon S \to T$ such that $h(p) = p$ for each $p \in X$. If every compactification of X with a proposed property property is equivalent, then it is said that S is *unique up to homeomorphism*.

Proposition 4.3.3. *Let X be a topological space and $\mathcal{M} \subset \mathbb{R}^X$ be a collection of bounded functions. Then, the following statements are equivalent:*

1. *$\mathcal{M} \subset C_B(X)$ separates and strongly separates points on X.*
2. *X admits a unique compactification S up to homeomorphism such that $\bigotimes \mathcal{M}$ extends to a homeomorphism between S and the closure of $\bigotimes \mathcal{M}(X)$ in $\mathbb{R}^{\mathcal{M}}$.*
3. *$\bigotimes \mathcal{M}$ is an imbedding of X in $\mathbb{R}^{\mathcal{M}}$.*

Proof. We provide the proof from Dong and Kouritzin 2020 Lemma 9.3.4.

($1 \to 2$) Kouritzin 2016 Theorem 6 (1 - 3) shows that there exists a compact S and homeomorphism $h\colon S \to \mathfrak{cl}[\bigotimes \mathcal{M}(X)]$ such that $h|_X = \bigotimes \mathcal{M}$. We show S is unique up to homeomorphism. Suppose T is another compactification of X such that $b\colon T \to \mathfrak{cl}[\bigotimes \mathcal{M}(X)]$ such that $b|_X = \bigotimes \mathcal{M}$. Then $b^{-1} \circ h\colon S \to T$ is a homeomorphism such that $b^{-1} \circ h(p) = p$ for each $p \in X$, which implies S and T are equivalent.

($2 \to 3$) Is proven directly as X is a subspace of S.

($3 \to 1$) By Proposition 3.3.7 (4), $\bigotimes \mathcal{M}$ is injective and $\mathscr{O}(X) = \mathscr{O}_{\mathcal{M}}(X)$. Given $p \neq q \in X$, $\bigotimes \mathcal{M}(p) \neq \bigotimes \mathcal{M}(q)$; hence, $\pi_f \circ \bigotimes \mathcal{M}(p) \neq \pi_f \circ \bigotimes \mathcal{M}(q)$ for some $f \in \mathcal{M}$ and $\pi_f \circ \bigotimes \mathcal{M} = f$, so $\mathcal{M}$ s.p. on X. Finally, $\mathcal{M}$ s.s.p. by Proposition 4.1.5. ∎

The compactification S and the associated homeomorphism will be vital when proving our universal approximation result for Tychonoff spaces in Chapter 7.

Interestingly, when there is a countable collection of functions that separate and strongly separate points, we can define a metric on the compactified space S.

Proposition 4.3.4. *Let X be a topological space; $N \in \mathbb{N} \cup \{\infty\}$; $\mathcal{M} = \{g_i\}_{i=1}^N \subset C(X)$ s.p. and s.s.p. on X; and let $\bar{h}\colon S \to \mathfrak{cl}[\bigotimes \mathcal{M}(X)]$ denote the extended homeomorphism mentioned in Proposition 4.3.3 (2). Then, S is metrized by the following metric:*

$$d(x,y) \mapsto \sum_{i=1}^{N} 2^{-i} \left(|\bar{g}_i(x) - \bar{g}_i(y)| \wedge 1\right) \qquad \forall x, y \in S \tag{4.6}$$

where $\bar{g}_i \doteq \pi_i \circ h$ for each $i = 1, \ldots, N$.

Proof. See Kouritzin 2016 Theorem 6 (4). ∎

Remark 4.3.5. As X is a subspace of S, the above metric is also valid for X.

4.4 Examples

Here we present some examples of functions that strongly separate points. First, we start off with a proposition that can make it easier to discover such collections of functions.

Proposition 4.4.1. *Let $(X, \mathcal{T})$ be a topological space and the members of $\mathcal{M}_0 \subset \mathbb{R}^X$ and $\mathcal{M} \subset \mathbb{R}^X$ are bounded. Suppose $\mathcal{M} \subset \bar{\mathcal{M}}_0$ (where the bar denotes closure under the sup metric[1]). If $\mathcal{M}$ separates points or strongly separates points, then $\mathcal{M}_0$ does also.*

Proof. We provide the proof from Dong and Kouritzin 2020 Corollary 9.2.3.

By Proposition 3.2.8, we have $\bar{\mathcal{M}}_0 \subset C_B(X, \mathscr{O}_{\mathcal{M}_0}(X); \mathbb{R})$; hence, $\mathcal{M} \subset C(X, \mathscr{O}_{\mathcal{M}_0}(X); \mathbb{R})$ and $\mathscr{O}_{\mathcal{M}}(X) \subset \mathscr{O}_{\mathcal{M}_0}(X)$.

So $\mathcal{M}$ s.p. on $(X, \mathcal{T})$ implies $(X, \mathscr{O}_{\mathcal{M}}(X))$ is a Hausdorff space by Proposition 4.1.6 (b). Then $\mathscr{O}_{\mathcal{M}}(X) \subset \mathscr{O}_{\mathcal{M}_0}(X)$ implies $(X, \mathscr{O}_{\mathcal{M}_0}(X))$ is also a Hausdorff space so $\mathcal{M}_0$ s.p. on $(X, \mathcal{T})$ too by Proposition 4.1.6.

[1] The sup metric was defined in Definition 3.2.7.

So $\mathcal{M}$ s.s.p. on $(X, \mathcal{T})$ implies $\mathcal{T} \subset \mathscr{O}_{\mathcal{M}}(X) \subset \mathscr{O}_{\mathcal{M}_0}(X)$ by Proposition 4.1.5, so $\mathcal{M}_0$ s.s.p. on $(X, \mathcal{T})$ too. ■

We would like to discuss how one may use the above result. By Proposition 4.1.5, $\mathcal{M}$ s.s.p. on $(X, \mathscr{O}_{\mathcal{M}}(X))$. So if the topology on X is defined by a collection of functions, then we only need a uniform dense collection $\mathcal{M}_0$ to get the s.s.p. property.

Example 4.4.2. Suppose X is a Tychonoff space. Proposition 4.1.7 tells us $C_B(X)$ s.s.p. and s.p. on X. So any uniform dense subset of $C_B(X)$ provides us with the s.s.p. and s.p. properties.

We can also use knowledge of homeomorphisms to quickly satisfy the strong separation of points property for simpler spaces.

Example 4.4.3. The projection functions s.s.p. and s.p. on $\mathbb{R}^I$ where I is any index set. To see this, let $\mathcal{M} = \{\pi_i : i \in I\}$, which then implies $\bigotimes \mathcal{M}$ is just the identity function on $\mathbb{R}^I$, so it is a homeomorphism.

The next example is interesting because each function is non-zero in a bounded region, so each function can be "turned off" depending on the input.

Example 4.4.4 (From Blount and Kouritzin 2010)**.** Suppose (X, d) is a metric space. Then the following collection of functions are continuous, s.p., and s.s.p. on X:

$$\{g_{q,k}(p) \mapsto (1 - kd(p, q)) \vee 0 : q \in X, k \in \mathbb{N}\}. \tag{4.7}$$

Finally, we conclude with Hibert spaces.

Example 4.4.5. Suppose X is a Hilbert space with inner product $\langle \cdot, \cdot \rangle \colon X \times X \to \mathbb{R}$ and a countable complete orthonormal basis $\mathcal{B}$. By Parseval's identity, for each $p \in X$ we have the following relationship:

$$p = \sum_{e \in \mathcal{B}} \langle p, e \rangle e. \tag{4.8}$$

So define $g_e(p) \mapsto \langle p, e\rangle$ and let $\mathcal{M} = \{g_e : e \in \mathcal{B}\}$. If $p \neq q$, then

$$p - q = \sum_{e\in\mathcal{B}} \langle p - q, e\rangle e \neq 0, \tag{4.9}$$

so, by the orthogonality of $\mathcal{B}$, there is some $e \in \mathcal{B}$ such that $\langle p-q, e\rangle = g_e(p) - g_e(q) \neq 0$, implying $\mathcal{M}$ s.p. on X. The topology on X is induced by the metric $d(p, q) \mapsto \sqrt{\langle p - q, p - q\rangle}$, and it is clear that the function $d_p(q) \mapsto d(p, q)$ is continuous as $\mathscr{O}_d(X)$ is generated by sets of the form

$$\{q \in X : d(p, q) < \epsilon\} = \{q \in X : d_p(q) < \epsilon\}. \tag{4.10}$$

Also, by expanding $d(q - p, e)$ and letting $p = 0$, one finds that

$$g_e(q) = \frac{1}{2}\left(d_0(q)^2 - d_e(q)^2 + 1\right), \tag{4.11}$$

which shows g_e is continuous. Hence we have that $\mathcal{M} \subset C(X)$. Now suppose (x_n) is a sequence in X such that $g_e(x_n) \to g_e(p)$ for each $e \in \mathcal{B}$. That is, we have $\langle x_n, e\rangle \to \langle p, e\rangle$. So by the completeness of $\mathcal{B}$, we have

$$\lim_{n\to\infty} x_n = \lim_{n\to\infty} \sum_{e\in\mathcal{B}} \langle x_n, e\rangle e = \sum_{e\in\mathcal{B}} \langle p, e\rangle e = p, \tag{4.12}$$

and implies $\mathcal{M}$ determines sequential point convergence, which is enough to conclude that $\mathcal{M}$ s.s.p. on X since it is countable by Proposition 4.2.3.

Chapter 5

Uniform Spaces

Uniform spaces are to uniform continuity as topological spaces are to continuity. In this section, we introduce tools to extend the concept of uniform continuity beyond metric spaces to the more general notion of uniform spaces. First, we motivate the use of uniform spaces through an example showing that uniform continuity cannot entirely be explained through topological spaces. Then, we provide a formal definition of uniform spaces and provide some introductory results as well as explore how uniform spaces are related to topological spaces. Finally, we show how uniformly continuous functions may be extended to compact sets.

5.1 Motivation

For metric spaces, uniformly continuous functions are defined according to the metrics associated with their input and output spaces. A uniformly continuous function between metric spaces is defined as below.

Definition 5.1.1 (Uniformly Continuous)**.** Let f be a function mapping between metric spaces (X, d_X) and (Y, d_Y). Then f is considered *uniformly continuous* if for every $\epsilon > 0$ there exists $\delta > 0$ such that

$$d_Y\left(f(p), f(q)\right) < \epsilon \tag{5.1}$$

for all p and q in X for which $d_X(p, q) < \delta$.

There are two questions that may naturally arise about the uniform continuity definition provided for metrics spaces:

1. Can it be extended to non-metric spaces?

2. Can it be entirely described through topological spaces?

The answer to the first question is yes and will be left to the next section. However, the answer to the second question is no, as we shall see in the following example.

Example 5.1.2. Recall the following metrics on $\mathbb{R}$ from Example 3.2.5:

$$d_1(x, y) \doteq |x - y| \qquad d_2(x, y) \doteq |\arctan x - \arctan y| \tag{5.2}$$

which both generate the standard topology. Now consider the real continuous function $f : \mathbb{R} \to \mathbb{R}$ defined as $f(x) \doteq x$ for each $x \in \mathbb{R}$. Is f uniformly continuous? It depends on the metric. We consider four cases:

(a) $f_a : (\mathbb{R}, d_1) \to (\mathbb{R}, d_1)$,

(b) $f_b : (\mathbb{R}, d_1) \to (\mathbb{R}, d_2)$,

(c) $f_c : (\mathbb{R}, d_2) \to (\mathbb{R}, d_1)$, and

(d) $f_d : (\mathbb{R}, d_2) \to (\mathbb{R}, d_2)$.

It follows trivially that f_a and f_d are uniformly continuous as $d_i(x, y) = d_i(f(x), f(y)) < \epsilon$ for $i = 1, 2$.

Now let us look at f_b. Recall from Example 3.2.5 lines (3.6 - 3.8) that $d_2(x, y) \leq d_1(x, y)$ for each $x, y \in \mathbb{R}$. Given $\epsilon > 0$, choose $\delta = \epsilon$ and it follows that $d_2(f(x), f(y)) = d_2(x, y) \leq d_1(x, y) < \delta = \epsilon$. So f_b is uniformly continuous.

Finally, we consider f_c. Suppose $n = 1, 2, \ldots$ and consider the sequences $x_n \doteq n$ and $y_n \doteq 2n$. Clearly, $d_1(x_n, y_n) = n \to \infty$. Also, by similar argument to lines (3.14 - 3.24), we have

$$
\begin{aligned}
d_2(x_n, y_n) &= \int_{x_n}^{y_n} \frac{1}{1+z^2}\, dz &(5.3)\\
&\leq \int_{x_n}^{y_n} \sup_{w\in[x_n,y_n]} \left\{\frac{1}{1+w^2}\right\} dz &(5.4)\\
&= \int_{x_n}^{y_n} \frac{1}{1+x_n^2}\, dz &(5.5)\\
&= \frac{(y_n - x_n)}{1+x_n^2} &(5.6)\\
&= \frac{n}{1+n^2} \to 0. &(5.7)
\end{aligned}
$$

That is, for any $\delta, M > 0$ there is an $N \in \mathbb{N}$ such that $d_1(x_N, y_N) > M$ and $d_2(x_N, y_N) < \delta$. Hence, f_c is *not* uniformly continuous.

The previous example demonstrated that uniform continuity is not entirely a topological property; which is to say that we cannot rely on the topologies of a function's domain and range to determine whether it is uniformly continuous. The next section will answer the question of: how can we extend the concept of uniform continuity to spaces without a metric?

5.2 Formal Introduction

This section shows how uniform continuity can be extended to non-metric spaces by introducing uniformities and uniform spaces. The majority of the material to be presented here can be found in sections 35-37 of Willard 2004.

It is common to characterize the continuity of a function pointwise; that is, a function is continuous if (and only if) it is continuous at each point in its domain. However, it is nonsensical to describe uniform continuity pointwise, since a function is uniformly continuous if mapped points are close *for any pair of domain points within a certain degree of closeness.* The particular location of the domain points is irrelevant. The main takeaway is that uniform continuity requires that we have some concept of closeness between points.

A metric d on a set X automatically provides closeness information as we can say that the set $\{(x,y) : d(x,y) < \epsilon;\ x, y \in X\}$ consists of point pairs that are within ϵ distance from each other. This is a hint that we may be able to extend the definition of uniform continuity beyond metric spaces by instead considering subsets of the Cartesian product $X \times X$. A *uniformity* does just this.

Before providing the definition for a uniformity, we first make note of some notation for subsets of $X \times X$.

Definition 5.2.1 (Diagonal)**.** If X is a set, then the *diagonal* of X is defined as $\Delta(X) \doteq \{(x,x) : x \in X\}$.

Definition 5.2.2 (Relation)**.** If A is a subset of $X \times X$, then A is called a *relation.* If $(x,y) \in A$, then x is *related* to y.

Remark 5.2.3. If x is related to y it is not necessarily the case that y is related to x.

Definition 5.2.4 (Inverse Relation, Symmetric)**.** If A is a relation, then the *inverse relation* of A is defined as $A^{-1} \doteq \{(y,x) \colon (x,y) \in A\}$. If $A = A^{-1}$, then A is *symmetric.*

Definition 5.2.5 (Composition of relations)**.** If A and B are subsets of $X \times X$, then the *composition* of A and B is defined as $A \circ B \doteq \{(x,y) : \text{for some } z \in X, (x,z) \in A \text{ and } (z,y) \in B\}$.

Remark 5.2.6. This is related to the idea of composition of functions for if $f \colon X \to X$ and $g \colon X \to X$ are functions, then $\{(x, g \circ f(x)) : x \in X\}$
$= \{(x,y) : \text{for some } z \in X, (x,z) \in X \times f(X) \text{ and } (z,y) \in X \times g(X)\}$.

We are finally ready to introduce the definitions of a uniformity and a uniform space.

Definition 5.2.7 (Uniformity)**.** A *uniformity*[1] on a set X is a collection $\mathfrak{D}$ of subsets of $X \times X$ which satisfy:

1. $A \in \mathfrak{D} \implies \Delta(X) \subset A$,

[1]There are equivalent definitions of a uniformity as explained by Willard 2004 chapters 35 and 36. The definition used in this document is for diagonal uniformities.

2. $A_1, A_2 \in \mathfrak{D} \Longrightarrow A_1 \cap A_2 \in \mathfrak{D}$,

3. $A \in \mathfrak{D} \Longrightarrow B \circ B \subset A$ for some $B \in \mathfrak{D}$,

4. $A \in \mathfrak{D} \Longrightarrow B^{-1} \subset A$ for some $B \in \mathfrak{D}$, and

5. $A \in \mathfrak{D}$ and $A \subset B \subset X \times X \Longrightarrow B \in \mathfrak{D}$.

Definition 5.2.8 (Uniform Space, Surroundings)**.** A set X together with uniformity $\mathfrak{D}$ form a *uniform space*. The members of $\mathfrak{D}$ are called *surroundings*. Given a uniform space X, we use the notation $\mathscr{U}(X)$ to denote the uniformity on X.

Looking at the definition of a uniformity, we can see the remnants of a metric. Condition 1 is an extension of $d(x, y) = 0$ if and only if $x = y$, condition 3 comes from the triangle inequality (see accompanying proof for Definition 5.2.15), and condition 4 is analogous to the symmetry of a metric ($d(x, y) = d(y, x)$).

Proposition 5.2.9. *Suppose X is a uniform space and $A \subset X$. The collection defined as $\mathscr{U}_X(A) \doteq \{D \cap (A \times A) : D \in \mathscr{U}(X)\}$ is a uniformity on A.*

Proof. We show $\mathscr{U}_X(A)$ is a uniformity on A. By definition, for each $E \in \mathscr{U}_X(A)$, there exists a $D \in \mathscr{U}(X)$ such that $E = D \cap (A \times A)$.

(Condition 1) $\Delta(X) \subset D$ implies $\Delta(X) \cap (A \times A) \subset D \cap (A \times A)$. Therefore, $\Delta(A) \subset E$ for each $E \in \mathscr{U}_X(A)$.

(Condition 2) Suppose $D_i \in \mathscr{U}(X)$, $E_i \in \mathscr{U}_X(A)$, and $E_i = D_i \cap (A \times A)$ for $i = 1, 2$. Then $E_1 \cap E_2 = D_1 \cap D_2 \cap (A \times A)$, and $D_1 \cap D_2 \in \mathscr{U}(X)$, so $E_1 \cap E_2 \in \mathscr{U}_X(A)$.

(Condition 3) $D \in \mathscr{U}(X)$ implies there is a $B \in \mathscr{U}(X)$ such that $B \circ B \subset D$. Suppose $(a, b) \in [B \cap (A \times A)] \circ [B \cap (A \times A)]$. Then there exists a $z^* \in X$ such that $(a, z^*) \in B \cap (A \times A)$ and $(z^*, b) \in B \cap (A \times A)$. However, $B \cap (A \times A) \subset B$; therefore, $(a, b) \in B \circ B$ and $(a, b) \in (A \times A) \circ (A \times A) = A \times A$. It follows that $[B \cap (A \times A)] \circ [B \cap (A \times A)] \subset B \circ B \cap (A \times A) \subset D \cap (A \times A)$.

(Condition 4) $D \in \mathscr{U}(X)$ implies there is a $B \in \mathscr{U}(X)$ such that $B^{-1} \subset D$. It follows that both $D \cap (A \times A)$ and $B \cap (A \times A)$ are elements of $\mathscr{U}_X(A)$, so we have $[B \cap (A \times A)]^{-1} = B^{-1} \cap (A \times A) \subset D \cap (A \times A)$.

(Condition 5) Suppose $E_1 \subset E_2 \subset A \times A$ and $E_1 \in \mathscr{U}_X(A)$; therefore, there is a $D \in \mathscr{U}(X)$ such that $E_1 = D \cap (A \times A)$. Clearly, $D \subset D \cup E_2$, so $D \cup E_2 \in \mathscr{U}(X)$, and it then follows that $(D \cup E_2) \cap (A \times A) \in \mathscr{U}_X(A)$. However, $(D \cup E_2) \cap (A \times A) = [E_2 \cap (A \times A)] \cup [D \cap (A \times A)] = E_2$. ■

Definition 5.2.10 (Relative Uniformity, Uniform Subspace)**.** Suppose X is a uniform space and $A \subset X$. Then $\mathscr{U}_X(A)$ (defined in Proposition 5.2.9) is called the *relative uniformity* induced on A by X. With this uniformity, A is called a *uniform subspace* of X.

Just like how a topological space may be generated from a topological basis, we can generate a uniformity from a uniform basis. We define a uniform basis next, and provide a means of identifying when a collection of subsets of $X \times X$ is indeed a uniform basis.

Definition 5.2.11 (Uniform Base)**.** $\mathfrak{E}$ is a *uniform base* for $\mathfrak{D}$ if and only if $\mathfrak{E} \subset \mathfrak{D}$ and each $D \in \mathfrak{D}$ contains some $E \in \mathfrak{E}$. A uniformity is generated from a base through repeated use of condition 5; that is $\mathfrak{D} = \{D \subset X \times X \mid E \subset D, E \in \mathfrak{E}\}$.

Proposition 5.2.12. *The symmetric surroundings form a uniform base.*

Proof. First, we demonstrate that, for any uniformity $\mathfrak{D}$, we have $D \in \mathfrak{D}$ implies $D^{-1} \in \mathfrak{D}$. From condition (4), we have $E^{-1} \subset D$ for some $E \in \mathfrak{D}$ and so $E \subset D^{-1}$. It follows that $D^{-1} \in \mathfrak{D}$ from condition (5).

Now suppose $D \in \mathfrak{D}$. It then follows from condition (2) that $D \cap D^{-1} \in \mathfrak{D}$ which is symmetric. ■

Proposition 5.2.13. *Suppose $\mathfrak{E}$ is a collection of subsets of $X \times X$. Then $\mathfrak{E}$ is a basis for some uniformity on X if and only if $\mathfrak{E}$ satisfies conditions (1), (3), and (4) as well as the below modified version of (2):*

$$A_1, A_2 \in \mathfrak{E} \Longrightarrow B \subset A_1 \cap A_2 \text{ for some } B \in \mathfrak{E}. \tag{5.8}$$

Proof. We first show that $\mathfrak{D} = \{D \subset X \times X \mid E \subset D, E \in \mathfrak{E}\}$ is a uniformity on X when $\mathfrak{E}$ satisfies all stated conditions. In what follows, let $A_1, A_2 \in \mathfrak{D}$ be arbitrary. Then, there exists $B_1, B_2 \in \mathfrak{E}$ such that $B_1 \subset A_1$ and $B_2 \subset A_2$.

(Condition 1) $\Delta(X) \subset B$ for any $B \in \mathfrak{E}$, so we have $\Delta(X) \subset B_1 \subset A_1$.

(Condition 2) By (5.8), there is a $B_3 \in \mathfrak{E}$ such that $B_3 \subset B_1 \cap B_2 \subset A_1 \cap A_2$. Since $B_3 \in \mathfrak{E}$ and $B_3 \subset A_1 \cap A_2$, it follows that $A_1 \cap A_2 \in \mathfrak{D}$.

(Condition 3) $B_1 \in \mathfrak{E}$ implies there is a $B_3 \in \mathfrak{E} \subset \mathfrak{D}$ such that $B_3 \circ B_3 \subset B_1 \subset A_1$.

(Condition 4) $B_1 \in \mathfrak{E}$ implies there is a $B_3 \in \mathfrak{E} \subset \mathfrak{D}$ such that $B_3^{-1} \subset B_1 \subset A_1$.

(Condition 5) Suppose A_3 is such that $A_1 \subset A_3 \subset X \times X$. Since $B_1 \subset A_1$ and $B_1 \in \mathfrak{E}$, it follows that $A_3 \in \mathfrak{D}$.

Next, we show that each condition is necessary for $\mathfrak{D}$ to be a uniformity.

(Condition 1) Choose $B \in \mathfrak{E}$ such that $\Delta(X) \not\in B$. Then $\mathfrak{E} \subset \mathfrak{D}$ implies $B \in \mathfrak{D}$ showing $\mathfrak{D}$ is not a uniformity.

(Condition 5.8) Choose $B_1, B_2 \in \mathfrak{E}$ such that there is no $E \in \mathfrak{E}$ such that $E \subset B_1 \cap B_2$. Each $A \in \mathfrak{D}$ contains some $E \in \mathfrak{E}$ which is not contained by $B_1 \cap B_2$, so $A \not\subset B_1 \cap B_2$.

(Condition 3) Since $B \subset A$ implies $B \circ B \subset A \circ A$, if there is no $E \in \mathfrak{E}$ such that $E \circ E \subset B$ for some $B \in \mathfrak{E}$, then there is no $D \in \mathfrak{D}$ such that $D \circ D \subset B$.

(Condition 4) Since $B \subset A$ implies $B^{-1} \subset A^{-1}$, if there is no $E \in \mathfrak{E}$ such that $E^{-1} \subset B$ for some $B \in \mathfrak{E}$, then there is no $D \in \mathfrak{D}$ such that $D^{-1} \subset B$. ■

We brought in the notion of a uniformity in order to extend the idea of a metric, so it is perhaps unsurprising that a metric may be used to generate a uniformity. We demonstrate this now with the following proposition and definition.

Proposition 5.2.14. *Given a metric d on X, the following sets form a base for a uniformity on X:*

$$D_\epsilon \doteq D_{X,d,\epsilon} \doteq \{(x, y) \in X \times X \mid d(x, y) < \epsilon\} \qquad \epsilon > 0. \tag{5.9}$$

Proof. We show the collection of sets does indeed form a basis for some uniformity.

(Condition 1) $\epsilon > 0$ implies $(x, x) \in D_\epsilon$ for any $x \in X$, so $\Delta(X) \subset D_\epsilon$.

(Condition 5.8) Suppose $\epsilon_2 > \epsilon_1 > 0$, then we have $D_{\epsilon_1} \cap D_{\epsilon_2} = D_{\epsilon_1}$ which is a basis element.

(Condition 3) Suppose $(a, b) \in D_{\epsilon_1} \circ D_{\epsilon_2}$, then there exists a point $z^* \in X$ such that $d(a, z^*) < \epsilon_1$ and $d(z^*, b) < \epsilon_2$. By the triangle inequality, we have $d(a, b) \leq d(a, z^*) + d(z^*, b) < \epsilon_1 + \epsilon_2$, so it follows that $D_{\epsilon_1} \circ D_{\epsilon_2} \subset D_{\epsilon_1+\epsilon_2}$. Hence, for any D_ϵ, we can choose the basis element $D_{\frac{\epsilon}{2}}$ to satisfy $D_{\frac{\epsilon}{2}} \circ D_{\frac{\epsilon}{2}} \subset D_\epsilon$.

(Condition 4) $d(x, y) = d(y, x)$, so $D_\epsilon^{-1} = D_\epsilon$ which is a basis element. ■

Definition 5.2.15 (Metric Uniformity)**.** A uniformity generated according to Proposition 5.2.14 is referred to as the *metric uniformity* generated by d on X and is denoted as $\mathscr{U}_d(X)$.

Definition 5.2.16 (Standard Uniformity on Real Numbers)**.** The *standard uniformity on* $\mathbb{R}$ is the metric uniformity generated from the standard metric[2].

Different metrics may or may not generate different metric uniformities. We showcase one example where the generated uniformities are the same and another where they are not.

Example 5.2.17. Suppose X is a set, d is a metric on X, and $a > 0$. Then $d_a(x, y) \doteq ad(x, y)$ defines a metric on X which generates the same metric uniformity as d. This follows from the fact that $\{(x, y) : d_a(x, y) < \epsilon\} = \{(x, y) : d(x, y) < \frac{\epsilon}{a}\}$, so the uniform bases are the same.

Example 5.2.18. Let d_1 and d_2 be defined as they were in Examples 3.2.5 and 5.1.2. That is,

$$d_1(x, y) \doteq |x - y| \qquad d_2(x, y) \doteq |\arctan x - \arctan y| \,. \tag{5.10}$$

Fix $\epsilon > 0$, and let $D = \{(x, y) : d_1(x, y) < \epsilon\}$. We claim $D \notin \mathscr{U}_{d_2}(\mathbb{R})$.

If D were a surrounding for $\mathscr{U}_{d_2}(\mathbb{R})$, then there would be a basis element $E = \{(x, y) : d_2(x, y) < \delta\} \in \mathscr{U}_{d_2}(\mathbb{R})$ with $\delta > 0$ such that $E \subset D$. By similar argument to lines (5.4 - 5.7) we can choose $a > 0$ such that $d_2(a, y) < \delta$ for all $y > a$ meaning that $\{(a, y) : y > a\} \subset E$.

[2]The standard metric was defined in Definition 3.2.3

Choosing $b > a + 2\epsilon$ implies $(a, b) \notin D$. Therefore, $E \not\subset D$, so $D \notin \mathscr{U}_{d_2}(\mathbb{R})$, and the uniformities are different.

Further yet, by line (3.9), we have $\{(x, y) : d_1(x, y) < \epsilon\} \subset \{(x, y) : d_2(x, y) < \epsilon\}$, so we have shown that $\mathscr{U}_{d_2}(\mathbb{R})$ is an exclusive subset of $\mathscr{U}_{d_1}(\mathbb{R})$.

Next, we present a minor result that will be useful in later sections when working with restrictions of metrics.

Proposition 5.2.19. *Suppose X is a uniform space with metric uniformity generated from the metric d_X. Let $A \subset X$ and define the metric d_A on A as $d_A(x, y) \mapsto d_X(x, y)$ for each $x, y \in A$. Then the metric uniformity on A generated by d_A is the subspace uniformity on A inherited from X ($\mathscr{U}_{d_A}(A) = \mathscr{U}_X(A)$).*

Proof. By definition, $\{D_{X,d_X,\epsilon} : \epsilon > 0\}$ is a uniform base for $\mathscr{U}_{d_X}(X)$. By the definition of a uniform base, it is easy to see that $\{(A \times A) \cap D_{X,d_X,\epsilon} : \epsilon > 0\}$ is a uniform base for $\mathscr{U}_X(A)$. However, $(A \times A) \cap D_{X,d_X,\epsilon} = D_{A,d_A,\epsilon}$, and $\{D_{A,d_A,\epsilon} : \epsilon > 0\}$ is a uniform base for $\mathscr{U}_{d_A}(A)$ by definition. So $\mathscr{U}_{d_A}(A)$ and $\mathscr{U}_X(A)$ have equivalent uniform bases. Hence, the uniformities are the same. ∎

5.2.1 Connection to Topological Spaces

Topological spaces are the means in which we characterize continuity, and we have mentioned that uniform spaces can be used to characterize uniform continuity. Hence, it is reasonable to ask how a uniformity relates to a topology. As it turns out, every uniformity defines a topology we call the uniform topology. We present this next.

Definition 5.2.20. Suppose X is a set and $D \subset X \times X$. It is then useful to introduce the set $D[x] \doteq \{y \in X \mid (x, y) \in D\}$.

Proposition 5.2.21. *Suppose $\mathfrak{D}$ is a uniformity on X. Then, for each $x \in X$, the following collection of sets:*

$$\mathcal{N}_x \doteq \{D[x] : D \in \mathfrak{D}\} \tag{5.11}$$

satisfy the properties of Proposition 3.1.5, forming a neighborhood base at x and defines a topology on X. The same topology is formed if a uniform base $\mathfrak{E}$ is used in place of $\mathfrak{D}$.

Proof. First, we show, for any uniform base $\mathfrak{E}$, $\mathcal{N}_x^{\mathfrak{E}} \doteq \{E[x] : E \in \mathfrak{E}\}$ satisfies the properties of Proposition 3.1.5 by adapting the proof provided by Willard 2004 Theorem 35.6. For any $E \in \mathfrak{E}$, we have $\Delta(X) \subset E$, therefore $x \in E[x]$. Next, for $E_1, E_2 \in \mathfrak{E}$, we can choose $E_3 \in \mathfrak{E}$ such that $E_3 \subset E_1 \cap E_2$, and observe that $E_3[x] \subset (E_1 \cap E_2)[x] = E_1[x] \cap E_2[x]$. Finally, $A \in \mathfrak{E}$ implies there is $B \in \mathfrak{E}$ such that $B \circ B \subset A$; therefore, $y \in B[x]$ implies $B[y] \subset A[x]$.

Let $\mathcal{T}_{\mathfrak{E}}$ and $\mathcal{T}_{\mathfrak{D}}$ represent the topologies generated by $\mathcal{N}_x^{\mathfrak{E}}$ and $\mathcal{N}_x^{\mathfrak{D}}$, respectively, according to Proposition 3.1.5. We show $\mathcal{T}_{\mathfrak{E}} = \mathcal{T}_{\mathfrak{D}}$. Clearly, $\mathcal{N}_x^{\mathfrak{E}} \subset \mathcal{N}_x^{\mathfrak{D}}$, so $\mathcal{T}_{\mathfrak{E}} \subset \mathcal{T}_{\mathfrak{D}}$. Now let $U \in \mathcal{T}_{\mathfrak{D}}$ and $x \in U$. It follows that there is $V_x \in \mathcal{N}_x^{\mathfrak{D}}$ such that $V_x \subset U$. However, $V_x \in \mathcal{N}_x^{\mathfrak{D}}$ implies there is a $D \in \mathfrak{D}$ such that $V_x = D[x]$. Since $D \in \mathfrak{D}$ there is $E \in \mathfrak{E}$ such that $E \subset D$. We then have $E[x] \in \mathcal{N}_x^{\mathfrak{E}}$ that satisfies $E[x] \subset V_x \subset U$, so $U \in \mathcal{T}_{\mathfrak{E}} = \mathcal{T}_{\mathfrak{D}}$. ■

Definition 5.2.22 (Uniform Topology, Uniformizable, Compatible)**.** The *uniform topology* is the one generated from a uniformity according to Proposition 5.2.21. We say a topological space is *uniformizable* if it is the uniform topology generated by some uniformity. Similarly, a uniform space and topological space are *compatible* with each other if the uniform topology is equivalent to the topology equipped to the topological space.

Remark 5.2.23. Unless otherwise mentioned, it is typical to assume that a given uniform space is also a topological space with its respective uniform topology.

Generally speaking, a uniformizable topology may be compatible with many uniformities. We already saw an example of this with the real numbers and standard topology. Example 3.2.5 showed that the metrics $d_1(x, y) \doteq |x - y|$ and $d_2(x, y) \doteq |\arctan x - \arctan y|$ each generate the standard topology, and Example 5.2.18 showed us that the uniformities generated by these metrics are in fact different. As we will see in Propostion 5.2.24, the uniform topologies generated by these metric uniformities are indeed the standard topology; hence, they are each compatible with the standard topology.

Conveniently, the uniform topology of a uniform subspace and metric uniformity correspond exactly with the topological subspace and metric topology, respectively.

Proposition 5.2.24. *The uniform topology generated by a metric uniformity is the metric topology.*

Proof. Suppose X is a set with metric d and recall that $\mathfrak{E} = \{\{(x, y) : d(x, y) < \epsilon\} : \epsilon > 0\}$ is a uniform base for the metric uniformity generated by d. Proposition 5.2.21 then states that $\mathcal{N}_x = \{D[x] : D \in \mathfrak{E}\}$ is a neighborhood base at x. Suppose $D \in \mathfrak{E}$, then $D[x] = \{y : d(x, y) < \epsilon\}$ is a topological basis element for the metric topology (see Definition 3.2.1 where the metric topology is defined). It then follows that

$$\mathscr{B}_d(X) = \bigcup_{x \in X} \mathcal{N}_x. \tag{5.12}$$

■

Proposition 5.2.25. *The uniform topology generated by the relative uniformity is the subspace topology.*

Proof. Note that if $\mathcal{N}_x$ is a neighborhood base for a topology on X, and $A \subset X$, then $\{U \cap A : U \in \mathcal{N}_x\}$ is a neighborhood base for the subspace topology on A. We then have

$$\begin{aligned} \{E[x] : E \in \mathscr{U}_X(A)\} &= \{(D \cap (A \times A))[x] : D \in \mathscr{U}(X)\} && (5.13) \\ &= \{D[x] \cap A : D \in \mathscr{U}(X)\}. && (5.14) \end{aligned}$$

■

Now that we have learned that a uniform space may be used to define a topological space, it is natural to wonder about the sorts of traits that can be said about the aforementioned uniform topology. The next result, from Willard 2004, is quite satisfying considering this document is primarily focused on Tychonoff spaces (defined as a completely regular Hausdorff space).

Proposition 5.2.26. *A topological space is uniformizable if and only if it is completely regular.*

Proof. Proved by Willard 2004 Theorem 38.2. ■

The main takeaway from Proposition 5.2.26 is that each Tychonofff space has at least one uniformity that is compatible with its topology.

In the previous section, Definition 5.1.1 defined uniform continuity for functions between metric spaces. One of the goals of this section is to extend uniform continuity to uniform spaces. We do this next.

Definition 5.2.27 (Uniformly Continuous)**.** Let X and Y be uniform spaces. A function $f : X \to Y$ is *uniformly continuous* if and only if for each $E \in \mathscr{U}(Y)$, there is some $D \in \mathscr{U}(X)$ such that $(x, y) \in D \Rightarrow (f(x), f(y)) \in E$.

The main results of this document are concerned with real valued functions; hence, it is unsurprising we introduce notation to denote real valued uniformly continuous functions.

Notation 5.2.28 (Collection of uniformly continuous functions)**.** Suppose $(X, \mathfrak{D})$ and $(Y, \mathfrak{E})$ are uniform spaces. We denote the collection of uniformly continuous functions from X to Y as $C_U((X, \mathfrak{D}); (Y, \mathfrak{E}))$, though often we will just say $C_U(X; Y)$ when there is no confusion over the particular uniformities. We reserve $C_U(X, \mathfrak{D})$ (and similarly $C_U(X)$) for when Y is the real numbers with standard uniformity.

For metric spaces, uniformly continuous functions are continuous functions. This relationship remains the same for uniform spaces when the topologies in question are generated from the uniformities. Also, restrictions of uniformly continuous functions to uniform subspaces are uniformly continuous.

Proposition 5.2.29. *Every uniformly continuous function is continuous with respect to the uniform topologies.*

Proof. Let X and Y be uniform spaces with respective uniform topologies $\mathcal{T}_X^U$ and $\mathcal{T}_Y^U$; $f\colon X \to Y$ be uniformly continuous; and $O \in \mathcal{T}_Y^U$. We show $f^{-1}(O) \in \mathcal{T}_X^U$.

Suppose $a \in f^{-1}(O)$. Then Proposition 5.2.21 says $\mathcal{M}_a = \{D[a] : D \in \mathscr{U}(X)\}$ is a neighborhood base at a and $\mathcal{N}_{f(a)} = \{E[f(a)] : E \in \mathscr{U}(Y)\}$ is a neighborhood base at $f(a)$. Since $f(a) \in O$ and $O \in \mathcal{T}_Y^U$, there is a $B \in \mathscr{U}(Y)$ such that $B[f(a)] \in \mathcal{N}_{f(a)}$ and $B[f(a)] \subset O$. The uniform continuity of f indicates there is an $A \in \mathscr{U}(X)$ such that $(x, y) \in A$ implies $(f(x), f(y)) \in B$, from which it then follows that $f(A[a]) \subset B[f(a)]$. We have then shown $A[a] \subset f^{-1}(O)$ and clearly $A[a] \in \mathcal{M}_a$, so $f^{-1}(O) \in \mathcal{T}_X^U$ by Definition 5.2.22. ■

Proposition 5.2.30. *Let X and Y be uniform spaces and suppose $f\colon X \to Y$ is uniformly continuous. If A is a uniform subspace of X, then $f|_A$ is uniformly continuous.*

Proof. Recall that $\mathscr{U}_X(A) \doteq \{(A \times A) \cap D \mid D \in \mathscr{U}(X)\}$ is the subspace uniformity on A inherited from X.

Let $E \in \mathscr{U}(Y)$. By the uniform continuity of f, there exists a $D \in \mathscr{U}(X)$ such that $(x, y) \in D$ implies $(f(x), f(y)) \in E$. Since $(A \times A) \cap D \subset D$, we also have $(x, y) \in (A \times A) \cap D$ implies $(f(x), f(y)) \in E$. Clearly, $(A \times A) \cap D \in \mathscr{U}_X(A)$, so $f|_A$ is uniformly continuous. ■

In the study of topological spaces we see domain spaces with finer topologies emit richer classes of continuous functions. This trend holds true for uniform spaces but instead with respect to the uniformities.

Proposition 5.2.31. *Suppose Y is a uniform space; $\mathfrak{D}_1$ and $\mathfrak{D}_2$ are uniformities on X; and $\mathfrak{D}_1 \subset \mathfrak{D}_2$. Then $C_U((X, \mathfrak{D}_1); Y) \subset C_U((X, \mathfrak{D}_2); Y)$.*

Proof. Follows directly from Definition 5.2.27. ■

> **Example 5.2.32.** We continue off from Example 5.2.18 where we showed that the uniformity generated by the arctan metric is contained within the standard uniformity on the real numbers. It follows then from Proposition 5.2.31 that each uniformly continuous function with domain uniformity generated by the arctan metric is also uniformly continuous if the domain were instead equipped with the standard uniformity. Further, Example 5.1.2 case (c) indicates that the containment is strict (a proper subset).

5.3 Extending to Compact Spaces

Now that we have put in the work to understand uniform spaces, we finally get to see their relevant properties required for the rest of this document. In particular, for some compact space X, we will want to better understand the collection of real-valued continuous functions on X restricted to some dense subspace A, which is denoted $C(X)|_A$. As it turns out, $C(X)|_A$ is in fact the uniformly continuous functions on A with a particular uniformity that is unique. The title of this section is inspired by the dual problem where one asks under what conditions can a uniformly continuous function $f\colon A \to \mathbb{R}$ be *extended* to X? Just as $C(X)|_A$ are restrictions of $C(X)$, $C(X)$ are extensions of $C(X)|_A$.

We have already defined Cauchy sequences in metric spaces (Definition 3.2.6). As this chapter continues to generalize via uniformities, it is perhaps then unsurprising that we instead discuss Cauchy nets in uniform spaces.

Definition 5.3.1 (Cauchy nets)**.** Let X be a uniform space. A net $(x_\lambda)_{\lambda\in\Lambda}$ in X is *Cauchy* if and only if for each $D \in \mathscr{U}(X)$, there is some $\lambda_0 \in \Lambda$ such that $(x_{\lambda_1}, x_{\lambda_2}) \in D$ whenever $\lambda_1, \lambda_2 \geq \lambda_0$.

A net converges or is Cauchy depending upon the topology and uniformity, respectively. As we learned from Proposition 5.2.21, each uniformity emits a topology called the uniform topology. In the case where the topology in question is the uniform topology, we have the following result.

Proposition 5.3.2. *Every convergent net is Cauchy.*

Proof. We provide the proof from Willard 2004 Theorem 39.2 with some added details. It follows from Propositions 5.2.12 and 5.2.13 that for any surrounding $D \in \mathscr{U}(X)$ there exists a symmetric $E \in \mathscr{U}(X)$ such that $E \circ E \subset D$. Now suppose $x_\lambda \to x$. For some λ_0, we have $x_\lambda \in E[x]$ for all $\lambda > \lambda_0$ (recall that $E[x]$ is a neighborhood of x by Definition 5.2.22). Finally, let λ_1 and λ_2 be larger than λ_0. It then follows that $(x_{\lambda_1}, x) \in E$ and $(x, x_{\lambda_2}) \in E$, so $(x_{\lambda_1}, x_{\lambda_2}) \in E \circ E \subset D$. ∎

Just as continuous functions map convergent nets to convergent nets, we have uniformly continuous functions map Cauchy nets to Cauchy nets.

Proposition 5.3.3. *Suppose $f\colon X \to Y$ is uniformly continuous and $(x_\lambda)_{\lambda\in\Lambda}$ is a Cauchy net in X. Then $(f(x_\lambda))_{\lambda\in\Lambda}$ is a Cauchy net in Y.*

Proof. Choose $E \in \mathscr{U}(Y)$. By the uniform continuity of f, there is some $D \in \mathscr{U}(X)$ such that $(x, y) \in D \Rightarrow (f(x), f(y)) \in E$. Since $(x_\lambda)_{\lambda\in\Lambda}$ is Cauchy, there is some $\lambda_0 \in \Lambda$ such that $(x_{\lambda_1}, x_{\lambda_2}) \in D$ whenever $\lambda_1, \lambda_2 \geq \lambda_0$. Therefore, $(f(x_{\lambda_1}), f(x_{\lambda_2})) \in E$ whenever $\lambda_1, \lambda_2 \geq \lambda_0$; hence, $(f(x_\lambda))_{\lambda\in\Lambda}$ is Cauchy. ■

We now have enough to answer the question of when can a uniformly continuous function defined on a dense subspace be extended to the closure. The answer is that we can create such an extension when the codomain is a *complete* uniform space, which we define next.

Definition 5.3.4 (Complete uniform space)**.** A uniform space is called *complete* if every Cauchy net converges.

Remark 5.3.5. $\mathbb{R}$ with its standard uniformity is complete.

Theorem 5.3.6. *Let $A \subset X$ be a uniform subspace, Y be a complete uniform space, and $f\colon A \to Y$ be uniformly continuous. For each $p \in \bar{A}$ choose a net (x^p_λ) in A such that $x^p_\lambda \to p$ and define the following function*

$$g(p) \doteq \lim_{\lambda} x^p_\lambda. \tag{5.15}$$

Then g extends f to $\bar{A}$ and is uniformly continuous.

Proof. See Willard 2004 Theorem 39.10. ■

At the beginning of this section, we mentioned a relationship between continuous functions on compact spaces and uniformly continuous functions on dense subspaces. Below is the result that allows us to make this connection. Recall from Definition 5.2.22 that a uniform space and topological space are compatible with each other if the uniform topology is equivalent to the topology equipped to the topological space.

Proposition 5.3.7. *Let X be a compact Hausdorff space. Then,*

1. *X has only one uniformity compatible with its topology, and*

2. every continuous function is uniformly continuous.

Proof. See Willard 2004 Theorems 36.18 and 36.19 and Corollary 36.20. ■

We now combine the past two propositions to reach the main result to carry forward from this chapter. In what follows, let $C_U(X)$ denote the set of real uniformly continuous functions on the uniform space X.

Proposition 5.3.8. *Let X be a compact Hausdorff space and A be a dense uniform subspace of X [i.e. $\bar{A} = X$]. Then $C(X)|_A = C_U(A)$.*

Proof. By Theorem 5.3.7, X being a compact Hausdorff space implies there is a unique uniformity on X compatible with its topology and $C_U(X) = C(X)$.

By Proposition 5.2.30, each uniformly continuous function restricted to a uniform subspace is uniformly continuous, so we have $C(X)|_A = C_U(X)|_A \subset C_U(A)$.

Likewise, by Theorem 5.3.6 and the completeness of $\mathbb{R}$, each real valued uniformly continuous function may be extended to the closure of its domain, so we have $C_U(A) \subset C_U(X)|_A = C(X)|_A$. ■

Remark 5.3.9. **A word of warning** needs to be mentioned when interpreting the presented proposition. Although X is only compatible with one uniformity, the topology of A may be compatible with many; and, as is often the case in future chapters, we only particularly care about A. We must remind ourselves that the uniformity defined on A is the subspace uniformity inherited from X (even though we may not care so much about X). See Example 5.3.10 for more on this discussion.

> **Example 5.3.10.** Here we construct an example to illustrate possible confusion that may arise when interpreting Proposition 5.3.8. Recall from Example 3.2.5 when defining the arctan metric we said $\arctan x = \int_0^x \frac{1}{1+z^2} dz$ for each $x \in \mathbb{R}$. Also, it is well known that $\lim_{x\to\pm\infty} \arctan x = \pm\frac{\pi}{2}$. This knowledge can be used to create a metric for the compact space $\bar{\mathbb{R}} \doteq \mathbb{R} \cup \{-\infty, \infty\}$ defined as
>
> $$\bar{d}_2(x, y) = \left| \lim_{p\to x} \arctan p - \lim_{q\to y} \arctan q \right|, \tag{5.16}$$

which is equivalent to the arctan metric when restricted to $\mathbb{R}$.

$\mathbb{R}$ is a dense subspace of $\bar{\mathbb{R}}$, so Proposition 5.3.8 applies, but the caution lies in the interpretation of $C_U(\mathbb{R})$. In particular, $\mathbb{R}$ inherits its uniformity from $\bar{\mathbb{R}}$, so what exactly is this uniformity? It is **not** the standard uniformity as defined in Definition 5.2.16, but instead is equivalent to the uniformity generated by the arctan metric. That is, we should add the metric to the notation as such: $C_U(\mathbb{R}, \mathscr{U}_{\bar{d}_2}(\mathbb{R}))$, or else confusion may arise. We know from Example 5.2.32 that $C_U(\mathbb{R}, \mathscr{U}_{\bar{d}_2}(\mathbb{R}))$ is a strict subset of $C_U(\mathbb{R}, \mathscr{U}_{d_1}(\mathbb{R}))$ ($\mathbb{R}$ with standard uniformity), so a mistake in interpretation would be costly.

As we show next, the uniformly continuous functions mentioned in Proposition 5.3.8 share many properties with continuous functions defined on compact spaces.

Proposition 5.3.11. *Under the conditions of Proposition 5.3.8, $C_U(A)$ is closed under addition and multiplication and $C_U(A) \subset C_B(A)$.*

Proof. $C(X) = C_B(X)$ is closed under addition and multiplication and $C(X)|_A = C_U(A)$. ∎

Remark 5.3.12. The conditions of Proposition 5.3.8 can not be removed entirely as the following example shows. Consider $\mathbb{R}$ with its standard uniformity and the identity function $f\colon \mathbb{R} \to \mathbb{R}$ defined as $f(x) \mapsto x$. Then $f \in C_U(\mathbb{R})$; however, $f^2 \notin C_U(\mathbb{R})$.

Proposition 5.3.13. *Suppose X and Y are compact uniform spaces; $A \subset X$ and $B \subset Y$ are dense uniform subspaces in X and Y, respectively; and $h\colon X \to Y$ is a homeomorphism such that $h(A) = B$. Then $C_U(A, \mathscr{U}_X(A)) = \{f \circ h : f \in C_U(B, \mathscr{U}_Y(B))\}$.*

Proof. From Proposition 5.3.8, we know that $C_U(A) = C(X)|_A$ and $C_U(B) = C(Y)|_B$, thus we need only show $C(X) = \{f \circ h : f \in C(Y)\}$.

Given $f \in C(Y)$, $f \circ h\colon X \to \mathbb{R}$ is a composition of continuous functions, so it is continuous; hence $C(X) \supset \{f \circ h : f \in C(Y)\}$.

Further, by symmetry we must have $C(Y) \supset \{g \circ h^{-1} : g \in C(X)\}$. It then follows that

$$C(X) \supset \{f \circ h : f \in C(Y)\} \tag{5.17}$$
$$\supset \{f \circ h : f \in \{g \circ h^{-1} : g \in C(X)\}\} \tag{5.18}$$
$$= \{g \circ h^{-1} \circ h : g \in C(X)\} \tag{5.19}$$
$$= C(X) \tag{5.20}$$

■

Remark 5.3.14. The compactness criteria is rather important as it implies we are working with a very particular uniformity that is tied to continuous functions on compact spaces. Generally speaking, homeomorphisms do not preserve uniform continuity. Consider the four homeomorphisms presented in Example 5.1.2, which are pointwise and topologically equivalent (the topologies of their domain and codomains is the same); however, f_c is not uniformly continuous. Therefore, $f_a \circ f_c$ is not uniformly continuous.

Now for a result that involves point separation properties from the previous chapter.

Notation 5.3.15. Let X be a topological space; $\mathcal{M} \subset C_B(X)$ s.p. and s.s.p. on X; and S be the unique compactification (up to homeomorphism) of X described in Proposition 4.3.3 statement 2. As S is compact, there is a unique uniformity compatible with the topology on S by Proposition 5.3.8, which we denote as $\mathfrak{S}_{\mathcal{M}}(S)$ and similarly the inherited uniform subspace on X from S is $\mathfrak{S}_{\mathcal{M}}(X)$. However, typically X is the intended focus of attention, so we will often exclude X from the notation in order to keep things tidy. So will use $\mathfrak{S}_{\mathcal{M}}$ when there is not risk of confusion.

Proposition 5.3.16. *Suppose X is a topological space, and let $\mathcal{M} \subset C_B(X)$ separate and strongly separate points on X. Then $C_U(X, \mathfrak{S}_{\mathcal{M}})$ separates and strongly separates points on X and $\mathcal{M} \subset C_U(X, \mathfrak{S}_{\mathcal{M}})$.*

Proof. By Proposition 4.3.3, $\bigotimes \mathcal{M}$ extends to a homeomorphism $\widehat{h}_{\mathcal{M}} \colon S \to \mathfrak{cl}[\bigotimes \mathcal{M}(X)]$, where $\mathfrak{cl}[\cdot]$ denotes closure in $\mathbb{R}^{\mathcal{M}}$, and S is compact.

For each $g \in \mathcal{M}$, define $\widehat{g} \doteq \pi_g \circ \widehat{h}_{\mathcal{M}}$ (π is the projection function). Then $\widehat{g}$ is a continuous extension of g to the compact set S. Since X is a dense uniform subspace of the compact Hausdorff space S, we have by Proposition 5.3.8 that $C_U(X, \mathfrak{S}_{\mathcal{M}}) = C(S)|_X \ni \widehat{g}|_X = g$. This holds for all $g \in \mathcal{M}$, hence we have $\mathcal{M} \subset C_U(X, \mathfrak{S}_{\mathcal{M}})$. It follows from Proposition 4.1.4 that $C_U(X, \mathfrak{S}_{\mathcal{M}})$ separates and strongly separates points on X. ■

Remark 5.3.17. If $\mathcal{M} = C_B(X)$, then $C_U(X) = C_B(X)$ and the compactified space S is the Stone–Čech compactification.

Chapter 6

Measure Spaces and Spaces of Measures

Measure spaces provide the backbone for probability theory. It is assumed the reader has some familiarity with the material, so we will not dwell too much on the basics; however, relevant parts will be covered to serve as a refresher as well as to establish notation for following chapters. The first section covers the very basics of measure spaces and notation. Section two introduces a few topologies on the set of positive-finite measures (which induces a topology on the probability measures via the subspace topology): these are (1) the weak topology and (2) the topology of weak convergence. The two topologies are related in the sense that the topology of weak convergence is the sequential topology generated from the weak topology, so they share the same set of convergent sequences. Lastly, we present some results indicating when a collection of functions s.s.p. on these spaces which will be of particular importance for our universal approximation theorems.

6.1 Measure Spaces

In this section, we introduce notation and some basic facts about measure spaces. The majority of the material can be found Dudley 2002.

Definition 6.1.1 (σ-algebra, Measurable Space)**.** Let X be a nonempty set, and Σ be a collection of subsets of X. Then Σ is called a *σ-algebra* if the following conditions hold:

1. $\emptyset, X \in \Sigma$,

2. $\{A_i\}_{i=1}^{\infty} \subset \Sigma$ implies $\bigcup_{i=1}^{\infty} A_i \in \Sigma$, and

3. $A \in \Sigma$ implies $X \setminus A \in \Sigma$.

Σ makes X a *measurable space*.

Remark 6.1.2. By De Morgan's laws, conditions (2) and (3) imply $\bigcap_{i=1}^{\infty} A_i \in \Sigma$ when $\{A_i\}_{i=1}^{\infty} \subset \Sigma$.

Proposition 6.1.3. *Suppose $\mathcal{C}$ is a collection of subsets of a set X, and define $\sigma(\mathcal{C})$ as the intersection of all σ-algebra's on X that contain $\mathcal{C}$. Then $\sigma(\mathcal{C})$ is a σ-algebra on X. We say $\sigma(\mathcal{C})$ is the σ-algebra generated by $\mathcal{C}$.*

Proof. Every σ-algebra on X contains $\emptyset$ and X, so $\{\emptyset, X\} \subset \sigma(\mathcal{C})$.

In what follows, let $S(X, \mathcal{C})$ denote the collection of all σ-algebra's on X that contain $\mathcal{C}$. If $\{A_i\}_{i=1}^{\infty} \subset \sigma(\mathcal{C})$, then $\{A_i\}_{i=1}^{\infty} \subset \Sigma$ for each $\Sigma \in S(X, \mathcal{C})$, so $\bigcup_{i=1}^{\infty} A_i \in \Sigma$ for each $\Sigma \in S(X, \mathcal{C})$; hence, $\bigcup_{i=1}^{\infty} A_i \in \sigma(\mathcal{C})$. Likewise, if $A \in \sigma(\mathcal{C})$, then $A \in \Sigma$ for each $\Sigma \in S(X, \mathcal{C})$, so $X \setminus A \in \Sigma$ for each $\Sigma \in S(X, \mathcal{C})$; hence, $X \setminus A \in \sigma(\mathcal{C})$. ■

Definition 6.1.4 (Borel σ-algebra, Borel sets)**.** Suppose X is a topological space. Then $\mathfrak{B}(X) \doteq \sigma(\mathscr{O}(X))$, where $\sigma(\cdot)$ is defined as in Proposition 6.1.3, is called the *Borel σ-algebra* on X. The elements of $\mathfrak{B}(X)$ are called *Borel sets*.

Remark 6.1.5. If X is a uniform space, then $\mathscr{O}(X)$ is assumed to be the uniform topology, so $\mathfrak{B}(X)$ is well defined. Further, it should be clear that uniformities with equivalent uniform topologies also have equivalent Borel σ-algebras.

Definition 6.1.6 (Measurable Functions)**.** Let X and Y be measurable spaces with respective σ-algebra's Σ_X and Σ_Y. A function $f\colon X \to Y$ is *measurable* if $B \in \Sigma_Y$ implies $f^{-1}(B) \in \Sigma_X$. We use $M(X, Y)$ to denote the collection of measurable functions, but may use the suppressed notation $M(X)$ when Y is the real numbers with $\Sigma_Y = \mathfrak{B}(\mathbb{R})$. Similarly, $M_B(X, Y)$ (or $M_B(X)$ when suppressed) will denote bounded measurable functions.

Proposition 6.1.7. *Suppose X and Y are measurable spaces with respective σ-algebra's Σ_X and Σ_Y; and $\mathcal{C} \subset \Sigma_Y$ is such that $\sigma(\mathcal{C}) = \Sigma_Y$. Then a function $f\colon X \to Y$ is measurable if and only if $B \in \mathcal{C}$ implies $f^{-1}(B) \in \Sigma_X$.*

Proof. See Dudley 2002 Theorem 4.1.6. ■

Proposition 6.1.8. *Suppose X and Y are topological spaces. Then $C(X,Y) \subset M(X,Y)$ where X and Y are equipped with their respective Borel σ-algebra's.*

Proof. Suppose $f \in C(X,Y)$. As f is continuous, we have $B \in \mathscr{O}(Y)$ implies $f^{-1}(B) \in \mathscr{O}(X)$. However, $\mathscr{O}(X)$ are Borel sets of X and $\mathfrak{B}(Y) = \sigma(\mathscr{O}(Y))$, so f is measurable by Proposition 6.1.7. ■

Definition 6.1.9 (Measure, Measure Space)**.** Suppose Σ is a σ-algebra on a set X. Then $\mu\colon \Sigma \to [0,\infty]$ is called a *measure* on Σ if it satisfies the following properties:

1. $\mu(\emptyset) = 0$, and
2. for any countable collection $\{A_i\}_{i=1}^{\infty} \subset \Sigma$ of mutually disjoint sets (i.e. $A_i \cap A_j = \emptyset$ when $i \neq j$), we have

$$\mu\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mu(A_i). \tag{6.1}$$

The set X with σ-algebra Σ and measure μ is called a *measure space*.

Definition 6.1.10 (Positive-Finite Measure, Probability Measure)**.** If (X,Σ,μ) is a measure space that satisfies $0 < \mu(X) < \infty$ and $\mu(A) \geq 0$ for each $A \in \Sigma$, then μ is called a *positive-finite measure*. Additionally, if $\mu(X) = 1$, then μ is a *probability measure*.

We conclude this section with a few results demonstrating that positive-finite measures can be thought of as scaled probability measures.

Proposition 6.1.11. *If μ is a positive-finite measure on the measurable space (X,Σ) and $c > 0$, then $\nu \doteq c \cdot \mu$ is a positive-finite measure on the same measurable space.*

Proof. First, show ν is a measure. $\nu(\emptyset) = c \cdot \mu(\emptyset) = c \cdot 0 = 0$. If $\{A_i\}_{i=1}^{\infty} \subset \Sigma$ is a countable collection of mutually disjoint sets then

$$\nu\left(\bigcup_{i=1}^{\infty} A_i\right) = c \cdot \mu\left(\bigcup_{i=1}^{\infty} A_i\right) \tag{6.2}$$

$$= c \cdot \sum_{i=1}^{\infty} \mu(A_i) \tag{6.3}$$

$$= \sum_{i=1}^{\infty} c \cdot \mu(A_i) \tag{6.4}$$

$$= \sum_{i=1}^{\infty} \nu(A_i). \tag{6.5}$$

So ν is a measure on (X, Σ). ν is positive-finite since for $c > 0$ we have

$$\mu(X) < \infty \iff c \cdot \mu(X) < c \cdot \infty = \infty \tag{6.6}$$

and

$$\mu(A) \geq 0 \iff c \cdot \mu(A) \geq c \cdot 0 = 0. \tag{6.7}$$

■

Proposition 6.1.12. *If μ is a positive-finite measure on the measurable space (X, Σ), then $\nu \doteq \frac{\mu}{\mu(X)}$ is a probability measure on the same measurable space.*

Proof. Since μ is a positive-finite measure, we have $\mu(X) > 0$ implies $\frac{1}{\mu(X)} > 0$ and so by Proposition 6.1.11, $\nu = \frac{1}{\mu(X)} \cdot \mu$ is a positive-finite measure on (X, Σ). Also, $\nu(X) = \frac{1}{\mu(X)} \cdot \mu(X) = 1$, so it is a probability measure on (X, Σ). ■

6.2 Spaces of Measures

There are two topologies on finite measures that we will examine: the *weak topology* and the *topology of weak convergence*. Subsequently, this implies two topologies on probability measures provided through the subspace topology.

First, we introduce some notation that will commonly used when working with measures.

Notation 6.2.1 (Collections of Measures)**.** We denote $\mathscr{M}^+(E, \Sigma)$ as the collection of all finite-measures on the measurable space (E, Σ). Like wise, $\mathscr{P}(E, \Sigma)$ is the collection of all probability measures. Often the σ-algebra will be dropped from the notation when E is a topological space, in which case $(E, \mathfrak{B}(E))$ is the measurable space.

Remark 6.2.2. Clearly, we have $\mathscr{P}(E) = \{\mu : \mu(E) = 1, \mu \in \mathscr{M}^+(E)\}$, so $\mathscr{P}(E) \subset \mathscr{M}^+(E)$.

Notation 6.2.3 (Lebesgue Integral)**.** Given a measure space (E, Σ, μ) and $f \in M(E)$, we use the common notation $\int_E f \, d\mu$ for the Lebesgue integral developed in chapter 4 of Dudley 2002.

Notation 6.2.4 (Functionals)**.** Given a measurable space (E, Σ) and $f \in M_B(E)$, we define the functional f^* as the mapping given as

$$f^*(\mu) \mapsto \int_E f \, d\mu \tag{6.8}$$

for each $\mu \in \mathscr{M}^+(E)$. If $\mathcal{D} \subset M_B(E)$, then $\mathcal{D}^* \doteq \{f^* \colon f \in \mathcal{D}\}$.

Now we are ready to introduce the first of the two topologies, which is defined based on a collection of functionals.

Definition 6.2.5 (Weak Topology of Finite Measures)**.** Let $(E, \mathcal{T})$ be a topological space. Then the *weak topology* on $\mathscr{M}^+(E)$ is defined as $\mathscr{O}_{C_B(E)^*}(\mathscr{M}^+(E))$ and is then denoted as $\mathcal{T}^W$.

Remark 6.2.6. By definition, $\mathcal{T}^W$ is the coarsest topology on $\mathscr{M}^+(E)$ such that $C_B(E)^* \subset C(\mathscr{M}^+(E))$.

Before getting to the second topology, we first define a common convergence criteria for spaces of measures.

Definition 6.2.7 (Weak Convergence, Weak Limit Point)**.** Let E be a topological space. A sequence of measures $\{\mu_n\}_{n\in\mathbb{N}} \subset \mathscr{M}^+(E)$ is said to *converge* *weakly to* $\mu \in \mathscr{M}^+(E)$, written $\mu_n \Rightarrow \mu$, if $\mu_n \to \mu$ with respect to the weak topology $\mathcal{T}^W$. This type of convergence is called *weak convergence.* μ is a *weak limit point* of $\Gamma \subset \mathscr{M}^+(E)$ if there exists a sequence in Γ that converges weakly to μ.

The above definitions might sound familiar. Way back when we discussed sequential spaces (Chapter 3.4), we defined a sequential limit point of a set A as being a point p in a topological space X such that there exists a sequence in A converging to p. So a weak limit point is the same as a sequential limit point. However, weak convergence can also be understood in terms of $C_B(E)^*$. As the below result shows, $C_B(E)^*$ determines sequential point convergence.

Proposition 6.2.8. *$\mu_n \Rightarrow \mu$ if and only if $f^*(\mu_n) \to f^*(\mu)$ holds for every $f \in C_B(E)$.*

Proof. Assume $\mu_n \Rightarrow \mu$. $C_B(E)^*$ are continuous in the weak topology and any sequence is a net; hence, by proposition 3.1.14, we have $\mu_n \to \mu$ implies $f^*(\mu_n) \to f^*(\mu)$ for each $f \in C_B(E)$.

Conversely, given $\mu \in \mathscr{M}^+(E)$ and neighborhood $N_\mu \in \mathcal{T}^W$, there exist $\mathcal{M}_0 \in \mathscr{R}_0(C_B(E))$ and $\epsilon > 0$ such that

$$\mu \in \left\{ \nu \in \mathscr{M}^+(E) : \max_{f \in \mathcal{M}_0} |f^*(\mu) - f^*(\nu)| < \epsilon \right\} \tag{6.9}$$

$$= \bigcap_{f \in \mathcal{M}_0} \left\{ \nu \in \mathscr{M}^+(E) : |f^*(\mu) - f^*(\nu)| < \epsilon \right\} \subset N_\mu, \tag{6.10}$$

since $C_B(E)^*$ s.s.p. by Definition 6.2.5. Since we have $f^*(\mu_n) \to f^*(\mu)$ for every $f \in \mathcal{M}_0$, there exists some $M_{f,\epsilon} \in \mathbb{N}$ such that
$f^*(\mu_n) \in (f^*(\mu) - \epsilon, f^*(\mu) + \epsilon)$ for each $n \geq M_{f,\epsilon}$. Thus, we can choose $M_\epsilon = \max_{f \in \mathcal{M}_0}\{M_{f,\epsilon}\} \in \mathbb{N}$, which guarantees that $\mu_n \in N_\mu$ for all $n \geq M_\epsilon$. So $\mu_n \Rightarrow \mu$. ■

Often, it is of particular use to find a subcollection $\mathcal{M} \subset C_B(E)$ that need only be checked to conclude $\mu_n \Rightarrow \mu$. The following proposition gives us conditions on $\mathcal{M}$ so we can do just that when working with probability measures.

Proposition 6.2.9. *Suppose that E is a topological space; $\{P_n\} \cup \{P\} \subset \mathscr{P}(E)$; and $\mathcal{M} \subset C_B(E)$ is countable, s.p., s.s.p., is closed under multiplication, and*

$$\int_E g\, dP_n \to \int_E g\, dP \qquad \forall g \in \mathcal{M}. \tag{6.11}$$

Then $P_n \Rightarrow P$.

Proof. See Blount and Kouritzin 2010 Theorem 6. ∎

Now we define our second topology based on the weak convergence criteria.

Definition 6.2.10 (Topology of Weak Convergence)**.** The *topology of weak convergence* on $\mathscr{M}^+(E)$ is defined as the sequentially open sets of $\mathscr{M}^+(E)$ generated from the weak topology[1]. We denote the topology of weak convergence as $\mathcal{T}^{WC}$.

As the topology of weak convergence is the sequential space generated from the weak topology, we would like to restate some results from our study of sequential spaces in Chapter 3.4, but within the context of positive-finite measures. Also, in the case where E is a metrizable space, the two topologies are equivalent and allows us to view continuity in terms of sequences rather than nets.

Proposition 6.2.11. *The following are true:*

1. *$\mathcal{T}^W$ and $\mathcal{T}^{WC}$ share the same convergent sequences.*

2. *$\mathcal{T}^W \subset \mathcal{T}^{WC}$.*

3. *$f \in C(\mathscr{M}^+(E), \mathcal{T}^{WC})$ if and only if f is sequentially continuous; that is, $\mu_n \Rightarrow \mu$ implies $f(\mu_n) \to f(\mu)$.*

4. *If E is a metrizable space, then $(\mathscr{M}^+(E), \mathcal{T}^W)$ is metrizable and $\mathcal{T}^W = \mathcal{T}^{WC}$.*

[1] See Definition 3.4.5

Proof. By Definition 6.2.10, $(\mathscr{M}^+(E), \mathcal{T}^{WC})$ is a sequential space. So we have:

(1.) Follows from Proposition 3.4.8.

(2.) Follows from Proposition 3.4.5.

(3.) Follows from Proposition 3.4.4.

(4.) The development of the Prohorov metric on $(\mathscr{P}(E), \mathcal{T}^W)$ is discussed in chapter 3 of Ethier and Kurtz 1986 and is extended to $(\mathscr{M}^+(E), \mathcal{T}^W)$ in Chapter 9 problem 6. The equality of the topologies then follows from Proposition 3.4.11 as metric spaces are sequential spaces. ■

6.3 Strong Separation of Measures

Here, we showcase some results that will be useful in identifying collections of functionals that separate and strongly separate points on spaces of measures. First, we start off with some results about sequences of measures which apply to both $\mathcal{T}^W$ and $\mathcal{T}^{WC}$ as they share the same convergent sequences.

Proposition 6.3.1 (Dong and Kouritzin 2020; Fact 10.1.19)**.** *Let E be a topological space. Then, the following statements are true:*

1. $\mu_1 = \mu_2$ *in* $\mathscr{M}^+(E)$ *if and only if* $\frac{\mu_1}{\mu_1(E)} = \frac{\mu_2}{\mu_2(E)}$ *in* $\mathscr{P}(E)$ *and* $\mu_1(E) = \mu_2(E)$.

2. $\mu_n \Rightarrow \mu$ *if and only if* $\lim_{n\to\infty} \mu_n(E) = \mu(E)$ *and*

$$\frac{\mu_n}{\mu_n(E)} \Rightarrow \frac{\mu}{\mu(E)} \text{ in } \mathscr{P}(E). \tag{6.12}$$

Proof. (1) Follows from $\mu(A) = \mu(E)\frac{\mu}{\mu(E)}(A)$ for each $A \in \mathfrak{B}(E)$ and $\mu \in \mathscr{M}^+(E)$.

(2.) Assume $\mu_n \Rightarrow \mu$. $1 \in C_B(E)$ implies

$$\lim_{n\to\infty} \mu_n(E) = \lim_{n\to\infty} \int_E 1\, d\mu_n = \int_E 1\, d\mu = \mu(E). \tag{6.13}$$

Then

$$\lim_{n\to\infty} f^*\left(\frac{\mu_n}{\mu_n(E)}\right) = \lim_{n\to\infty} \frac{f^*(\mu_n)}{\mu_n(E)} = \frac{f^*(\mu)}{\mu(E)} = f^*\left(\frac{\mu}{\mu(E)}\right) \tag{6.14}$$

holds for each $f \in C_B(E)$. Conversely, we have for each $f \in C_B(E)$ that

$$\lim_{n\to\infty} f^*(\mu_n) = \lim_{n\to\infty} \mu_n(E) f^*\left(\frac{\mu_n}{\mu_n(E)}\right) = \mu(E) f^*\left(\frac{\mu}{\mu(E)}\right) = f^*(\mu). \quad (6.15)$$

■

The above result immediately implies a one to one relationship regarding the determining sequential point convergence property, which we state next.

Proposition 6.3.2 (Dong and Kouritzin 2020; 10.1.20)**.** *Let E be a topological space and $1 \in \mathcal{M} \subset M_B(E)$. Then:*

1. *$\mathcal{M}^*$ separates points on $\mathscr{M}^+(E)$ if and only if $\mathcal{M}^*$ separates points on $\mathscr{P}(E)$.*
2. *$\mathcal{M}^*$ determines sequential point convergence on $\mathscr{M}^+(E)$ if and only if $\mathcal{M}^*$ determines sequential point convergence on $\mathscr{P}(E)$.*

Proof. Follows from Proposition 6.3.1. ■

In the case where we are working with the topology of weak convergence, a useful homeomorphism can be defined that relates positive-finite measures with probability measures.

Proposition 6.3.3. *Equip $\mathscr{M}^+(E)$ and $\mathscr{P}(E)$ with the topology of weak convergence and consider the function $H\colon \mathscr{M}^+(E) \to (0,\infty) \times \mathscr{P}(E)$ defined as $H(\mu) \mapsto \left(\mu(E), \frac{\mu}{\mu(E)}\right)$. Then H is a homeomorphism with inverse $H^{-1}(c, P) \mapsto c \cdot P$.*

Proof. Proposition 6.3.1 (1) shows that H is a bijection, so all is left is to demonstrate continuity. $\mathscr{M}^+(E)$ and $\mathscr{P}(E)$ with the topology of weak convergence are sequential spaces by definition. Further, $(0,\infty)$ is metrizable as a subspace of $\mathbb{R}$, so it is also a sequential space. Theorem 4 from Antosik, Boehme, and Mohanadi 1985 says that the product of two sequential spaces is sequential when one of the spaces is locally compact. Conveniently, $(0,\infty)$ is locally compact, so we can conclude that $(0,\infty) \times \mathscr{P}(E)$ is a sequential space. Hence, H is a mapping between sequential spaces, so its continuity properties are reduced to checking for sequential continuity, which has already been established by Proposition 6.3.1 (2). So H is a homeomorphism when $\mathscr{M}^+(E)$ and $\mathscr{P}(E)$ are equipped with the topology of weak convergence. ■

The above is particularly important for our universal approximation results where we will want bounded functionals. This boundedness property is satisfied when the measure is a probability measure. To see this, let $f \in C_B(E)$, so there is some $M > 0$ such that $M = \sup\{|f(q)| : q \in E\}$, from which it follows that

$$\left|\int f\, dP\right| < \int M\, dP \tag{6.16}$$

$$=M, \tag{6.17}$$

implying $f^*(P)$ is bounded for any $P \in \mathscr{P}(E)$.

Next, we provide a result demonstrating the difficulty in trying to find a collection of functionals that s.s.p. on the topology of weak convergence.

Proposition 6.3.4. *$C_B(E)^*$ always s.s.p. on $(\mathscr{M}^+(E), \mathcal{T}^W)$; however, $C_B(E)^*$ s.s.p. on $(\mathscr{M}^+(E), \mathcal{T}^{WC})$ if and only if $\mathcal{T}^{WC} = \mathcal{T}^W$.*

Proof. Proposition 4.1.5 says that $\mathcal{M}$ s.s.p. on $(X, \mathcal{T})$ if and only if $\mathcal{T} \subset \mathscr{O}_{\mathcal{M}}(X)$. So the result follows then from Definition 6.2.5 and Proposition 6.2.11 (1), where we have $\mathscr{O}_{C_B(E)^*}(\mathscr{M}^+(E)) = \mathcal{T}^W \subset \mathcal{T}^{WC}$. ■

With the above in mind, we will restrict our attention to when E is a metric space, so that the weak topology and topology of weak convergence are the same and we will only need to consider converging sequences rather than nets (see Proposition 6.2.11). Now we have a nice proposition telling us when we have the s.s.p. property on our spaces of measures.

Theorem 6.3.5. *Suppose that E is a topological space; $g_0 \in C_B((0,\infty))$ s.p. and s.s.p. on $(0,\infty)$; and $\mathcal{M} = \{1\} \cup \{g_i\}_{i=2}^N \subset C_B(E)$ is countable, s.p., s.s.p., and is closed under multiplication. Further, define $\mathfrak{W}_{g_0}[\mathcal{M}] \doteq \{\mu \mapsto g^*(\frac{\mu}{\mu(E)}) : g \in \mathcal{M}, g \neq 1\} \cup \{\mu \mapsto g_0(\mu(E))\}$. Then*

1. *$\mathscr{M}^+(E)$ is metrizable and $\mathcal{T}^{WC} = \mathcal{T}^W$,*

2. $\mathfrak{W}_{g_0}[\mathcal{M}] \subset C_B(\mathcal{M}^+(E))$

3. $\mathfrak{W}_{g_0}[\mathcal{M}]$ *determines sequential point convergence on* $\mathscr{M}^+(E)$,

4. $\mathfrak{W}_{g_0}[\mathcal{M}]$ *s.s.p and s.p. on* $\mathscr{M}^+(E)$,

5. $\bigotimes \mathfrak{W}_{g_0}[\mathcal{M}] \colon \mathscr{M}^+(E) \to \bigotimes \mathfrak{W}_{g_0}[\mathcal{M}](\mathscr{M}^+(E))$ *is a homeomorphism.*

Proof. (1.) By Proposition 4.3.4, E is metrizable, so $\mathscr{M}^+(E)$ is metrizable (hence, Hausdorff) and $\mathcal{T}^{WC} = \mathcal{T}^W$ by Proposition 6.2.11.

(2.) Justification for the boundedness of $\mu \mapsto g^*(\frac{\mu}{\mu(E)})$ was provided in (6.17). Boundedness of $\mu \mapsto g_0(\mu(E))$ comes from the boundedness of g_0. We get $\mathfrak{W}_{g_0}[\mathcal{M}] \subset C_B(E)^* \subset C(\mathcal{M}^+(E))$ since g_0 and $\mathcal{M}$ are bounded continuous functions on E.

(3.) From Proposition 6.2.9, we have that $\{g_i^*\}_{i=2}^N$ determines sequential point convergence on $\mathscr{P}(E)$. Also, clearly $g_0 \in C_B((0,\infty))$ determines sequential point convergence on $(0,\infty)$ as it s.p. and s.s.p. on $(0,\infty)$. It then follows that the following collection determines sequential point convergence on $(0,\infty) \times \mathscr{P}(E)$

$$\{g_0 \circ \pi_1\} \cup \{g_i^* \circ \pi_2\}_{i=2}^N. \tag{6.18}$$

The rest follows from the homeomorphism $H \colon \mathscr{M}^+(E) \to (0,\infty) \times \mathscr{P}(E)$ presented in Proposition 6.3.3.

(4.) S.s.p. is implied by (1.), (3.), Proposition 4.2.3 ($4 \to 2$), and that $\mathfrak{W}_{g_0}[\mathcal{M}]$ is countable. (1.) implies $\mathscr{M}^+(E)$ is Hausdorff and combined with Proposition 4.1.6, implies $\mathfrak{W}[\mathcal{M}]$ s.p..

(5.) Follows from (2.), (4.), and Proposition 4.2.3 ($2 \to 1$). ∎

We finish this chapter with a statement about uniformly continuous functions.

Proposition 6.3.6. *Suppose E is a topological space; $g_0 \in C_B((0,\infty))$ s.p. and s.s.p. on $(0,\infty)$; and let $\mathcal{M} \subset C_B(E)$ be countable, s.p. and s.s.p. on E. Then $\mathfrak{W}_{g_0}[C_U(E, \mathfrak{S}_{\mathcal{M}})]$[2] s.p. and s.s.p. on $\mathscr{M}^+(E)$.*

[2] $\mathfrak{S}_{\mathcal{M}}$ was defined in Notation 5.3.15.

Proof. By Propositions 5.3.16 and 5.3.11, $C_U(E, \mathfrak{S}_{\mathcal{M}}(E))$ separate and strongly separate points on E, is closed under multiplication, and $\mathcal{M} \subset C_U(E, \mathfrak{S}_{\mathcal{M}})$. As $\mathcal{M}$ is countable, say it is of size $N \in \mathbb{N} \cup \{\infty\}$ and $\mathcal{M} = \{g_i\}_{i=1}^N$. Define the following

$$\mathcal{N}_i = \left\{ \prod_{g \in \mathcal{C}_0} g : \mathcal{C}_0 \in \mathscr{R}_0 \left(\{g_j\}_{j=1}^i \right) \right\}, \tag{6.19}$$

which is finite for each i, so $\mathcal{N} = \bigcup_{i=1}^N \mathcal{N}_i$ is countable, closed under multiplication, and contains $\mathcal{M}$. Hence, $\mathcal{N} \subset C_B(E)$ is countable, closed under multiplication, s.p., and s.s.p. on E, so $\mathfrak{W}_{g_0}[\mathcal{N}]$ s.s.p. and s.p. on $\mathscr{M}^+(E)$ by Proposition 6.3.5.

Further, $\mathcal{N} \subset C_U(E, \mathfrak{S}_{\mathcal{M}})$ since $C_U(E, \mathfrak{S}_{\mathcal{M}})$ is closed under multiplication and contains $\mathcal{M}$. The result then follows from
$\mathfrak{W}_{g_0}[\mathcal{N}] \subset \mathfrak{W}_{g_0}[C_U(E, \mathfrak{S}_{\mathcal{M}})]$. ■

Chapter 7

Universal Approximation Results

Finally it is time to put all of the tools we have developed to use. It is now that we will provide our universal approximation results for Tychonoff spaces and spaces of measures. The route we take is to provide a uniform dense result on compact Hausdorff spaces while making use of the Stone-Weierstrass Theorem. We then use our homeomorphism methods to compactify Tychonoff spaces, which leads to our first main result: universal approximation of uniformly continuous functions. Lastly, we apply the Tychonoff result to spaces of positive-finite measures.

7.1 Compact Hausdorff Spaces

Interestingly, just about all of the mathematical background presented so far is irrelevant for this particular section as we instead rely heavily on the Stone-Weierstrass Theorem to pull us directly to universal approximation on compact Hausdorff spaces. Although we say "the" Stone-Weierstrass Theorem, there are in reality many different varieties of Stone-Weierstrass-like theorems. What is common between them, however, is they are all about uniform dense collections of algebras of functions, which we define next.

Definition 7.1.1 (Algebra)**.** An algebra is a vector space that is also closed under multiplication. That is, A is an algebra if it satisfies

1. $x, y \in A$ implies $c_1 x + c_2 y \in A$ for every $c_1, c_2 \in \mathbb{R}$
2. $x, y \in A$ implies $xy \in A$

We now state a version of the Stone-Weierstrass Theorem adapted from Rudin 1991 (details in Appendix A).

Theorem 7.1.2 (Stone-Weierstrass)**.** *Let X be a compact Hausdorff space and let $C(X)$ be the set of real continuous functions on X equipped with the sup metric*[1]*. Suppose that:*

1. *A is a closed subalgebra of $C(X)$,*
2. *A separates points on X,*
3. *A vanishes nowhere on X (i.e., at every $p \in X$, $f(p) \neq 0$ for some $f \in A$).*

Then $A = C(X)$.

Therefore, an algebra of continuous functions on a compact space X is uniform dense in $C(X)$ if it separates points and vanishes nowhere. In fact, the algebra strongly separates points on X as we saw in Proposition 4.3.1 separate points implies strong separation of points in a compact Hausdorff setting.

Given a subset of continuous functions $\mathcal{M} \subset C(X)$ closed under addition, we can construct an algebra of continuous functions using the exponential function.

Lemma 7.1.3. *Suppose X is a topological space and $\mathcal{M} \subset C(X)$ is closed under addition. Then the following collection of functions*

$$\Lambda(\mathcal{M}) \doteq \left\{ p \mapsto \sum_{i=1}^{n} c_i e^{g_i(p)} \mid g_i \in \mathcal{M}; c_i \in \mathbb{R}; n \in \mathbb{N} \right\}, \tag{7.1}$$

is an algebra and $\Lambda(\mathcal{M}) \subset C(X)$.

[1] The sup metric was defined in Definition 3.2.7.

Proof. Scalar multiplication, addition, and composition of continuous functions is continuous, so clearly $\Lambda(\mathcal{M}) \subset C(X)$. Now let $\lambda \doteq \sum_{i=1}^{n} c_i e^{g_i}$, $\phi \doteq \sum_{j=1}^{m} d_j e^{h_j}$ and $a_1, a_2 \in \mathbb{R}$. Then,

$$a_1\lambda + a_2\phi = a_1 \sum_{i=1}^{n} c_i e^{g_i} + a_2 \sum_{j=1}^{m} d_j e^{h_j} \tag{7.2}$$

$$= \sum_{i=1}^{n} a_1 c_i e^{g_i} + \sum_{j=1}^{m} a_2 d_j e^{h_j} \tag{7.3}$$

$$= \sum_{(b,f)\in W} b e^{f} \in \Lambda(\mathcal{M}), \tag{7.4}$$

where $W = \{(a_1 c_i, g_i)\}_{i=1}^{n} \cup \{(a_2 d_j, h_j)\}_{j=1}^{m}$. Since $W \in \mathscr{R}_0[\mathbb{R} \times \mathcal{M}]$, we have $a_1\lambda + a_2\phi \in \Lambda(\mathcal{M})$. Also, we have

$$\lambda\phi = \left(\sum_{i=1}^{n} c_i e^{g_i}\right)\left(\sum_{j=1}^{m} d_j e^{h_j}\right) \tag{7.5}$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{m} c_i d_j e^{g_i + h_j} \tag{7.6}$$

$$= \sum_{(b,f)\in V} b e^{f}, \tag{7.7}$$

where $V = \{(c_i d_j, g_i + h_j) : i = 1, \ldots, n; j = 1, \ldots, m\}$. Since $V \in \mathscr{R}_0[\mathbb{R} \times \mathcal{M}]$ (as $\mathcal{M}$ is closed under addition), we have $\lambda\phi \in \Lambda(\mathcal{M})$. ■

Next, we use the Stone-Weierstrass Theorem to provide a supporting Lemma indicating when our algebra is uniform dense in $C(X)$.

Lemma 7.1.4. *Let X be a compact Hausdorff space, and let $\mathcal{M} \subset C(X)$ be closed under addition and separate points on X. Then $\Lambda(\mathcal{M})$ (defined by 7.1) is uniform dense in $C(X)$.*

Proof. We show $\Lambda(\mathcal{M})$ satisfies the conditions of the Stone-Weierstrass Theorem. $\mathcal{M}$ is closed under addition, so $\Lambda(\mathcal{M})$ is a subalgebra of $C(X)$. For any $g \in \mathcal{M}$, $e^g \in \Lambda(\mathcal{M})$ and $e^{g(p)} > 0$ for all $p \in X$, so $\Lambda(\mathcal{M})$ vanishes nowhere. $\mathcal{M}$ separates points, so for $p \neq q \in X$, there is a $g \in \mathcal{M}$ such that $g(p) \neq g(q)$, and hence $e^{g(p)} \neq e^{g(q)}$, so $\Lambda(\mathcal{M})$ separates points. ■

Now we provide our main result for this section, which is universal approximation on compact Hausdorff spaces. We build our class of functions by taking a collection that separate points on X and then input those too functions that are uniform dense on compact sets.

Theorem 7.1.5. *Let X be a compact Hausdorff space, and let $\mathcal{M} \subset C(X)$ separate points on X. Suppose that, for every $n \in \mathbb{N}$, $\mathcal{F}_n$ is uniform dense on compacts of $\mathbb{R}^n$. Then the following set is a uniform dense subset of $C(X)$:*

$$\mathfrak{H}(\mathcal{M}) \doteq \big\{p \mapsto f(g_1(p), \, \dots \, , g_n(p)) : n \in \mathbb{N}; \; f \in \mathcal{F}_n; \; \{g_i\}_{i=1}^n \in \mathscr{R}_0(\mathcal{M})\big\}. \tag{7.8}$$

Proof. Clearly, $\mathfrak{H}(\mathcal{M}) \subset C(X)$ as $f(g_1, \, \dots \, , g_n)$ is a composition of continuous functions, so what is left is to show $\mathfrak{H}(\mathcal{M})$ is uniform dense in $C(X)$.

We wish to employ Lemma 7.1.4 and the transitive property of dense sets; however, Lemma 7.1.4 assumes $\mathcal{M}$ is closed under addition (unlike Theorem 7.1.5). To circumvent this, we construct a set $\mathcal{M}^{(+)}$ which is closed under addition and show $\mathfrak{H}(\mathcal{M})$ is dense in $\Lambda(\mathcal{M}^{(+)})$ which is dense in $C(X)$.

Define $\mathcal{M}^{(+)}$ as follows:

$$\mathcal{M}^{(+)} \doteq \left\{ \sum_{g \in H} g \mid H \subset \mathcal{M}, \; H \text{ finite} \right\}. \tag{7.9}$$

Clearly, $\mathcal{M}^{(+)}$ is closed under addition and inherits the separating points property from $\mathcal{M}$ (since $\mathcal{M} \subset \mathcal{M}^{(+)}$), hence $\Lambda(\mathcal{M}^{(+)})$ is dense in $C(X)$ by Lemma 7.1.4.

We now show $\mathfrak{H}(\mathcal{M})$ is dense in $\Lambda(\mathcal{M}^{(+)})$. Each $\lambda \in \Lambda(\mathcal{M}^{(+)})$ takes on the following form:

$$\lambda = \sum_{i=1}^m c_i \exp\{h_i\} \qquad h_i \in \mathcal{M}^{(+)}, \; c_i \in \mathbb{R} \tag{7.10}$$

$$= \sum_{i=1}^m c_i \exp\left\{ \sum_{g \in H^{(i)}} g \right\} \qquad H^{(i)} \subset \mathcal{M}, \; |H^{(i)}| < \infty. \tag{7.11}$$

Letting $n = \left|\bigcup_{i=1}^m H^{(i)}\right|$, we can rewrite λ as a composition of continuous functions $\Gamma : X \to \mathbb{R}^n$, $\Phi : \mathbb{R}^n \to \mathbb{R}^m$, and $\Psi : \mathbb{R}^m \to \mathbb{R}$ defined as

$$\Gamma(x) \mapsto (g_i(x))_{i=1}^n \tag{7.12}$$

$$\Phi((y_i)_{i=1}^n) \mapsto \left(\sum_{i=1}^n y_i I_{[g_i \in H^{(j)}]} \right)_{j=1}^m \tag{7.13}$$

$$\Psi((z_i)_{i=1}^m) \mapsto \sum_{i=1}^m c_i e^{z_i}, \tag{7.14}$$

so $\lambda = \Psi \circ \Phi \circ \Gamma$.

X is compact and Γ is continuous, so $\Gamma(X) \subset \mathbb{R}^n$ is compact. Since $\Psi \circ \Phi \in C(\mathbb{R}^n)$, we have for each $\epsilon > 0$, there exists a function $f \in \mathcal{F}_n$ such that

$$\epsilon > \sup \{ |(\Psi \circ \Phi)(y) - f(y)| : y \in \Gamma(X) \} \tag{7.15}$$

$$= \sup \{ |(\Psi \circ \Phi \circ \Gamma)(x) - (f \circ \Gamma)(x)| : x \in X \} \tag{7.16}$$

$$= \sup \{ |\lambda(x) - (f \circ \Gamma)(x)| : x \in X \} . \tag{7.17}$$

By definition $f \circ \Gamma \in \mathfrak{H}(\mathcal{M})$, so we have shown that $\mathfrak{H}(\mathcal{M})$ is dense in $\Lambda(\mathcal{M}^{(+)})$ and $C(X)$. ■

7.2 Tychonoff Spaces

Now we are ready for our first main result: universal approximation on Tychonoff spaces. Our plan of action is to build a homeomorphism based on the strong separation of points background that we developed in Chapter 4. The homeomorphism can be extended to a compact Hausdorff space, which is the correct setting to make use of our universal approximation result from the previous section. The result demonstrates that we can approximate uniformly continuous functions from a unique uniformity which has an associated metric in the case where we have a countable collection of functions that strongly separate points.

Definition 7.2.1 (Topological Neural Network)**.** Suppose X is a topological space; $\mathcal{M} \subset C_B(X)$; and, for each $n \in \mathbb{N}$, $\mathcal{F}_n \subset C(\mathbb{R}^n)$. Then we let

$\mathfrak{N}(\mathcal{M}, \{\mathcal{F}_n\}_{n=1}^{\infty})$ denote the following collection of functions:

$$\bigcup_{n=1}^{\infty} \big\{p \mapsto f(g_1(p), \ldots, g_n(p)) : f \in \mathcal{F}_n;\ \{g_i\}_{i=1}^{n} \in \mathscr{R}_0(\mathcal{M})\big\}. \tag{7.18}$$

We call functions of the form given by (7.18) *topological neural networks* and $\mathfrak{N}(\mathcal{M}, \{\mathcal{F}_n\}_{n=1}^{\infty})$ is the collection of neural networks generated by $\mathcal{M}$ and $\{\mathcal{F}_n\}_{n=1}^{\infty}$.

Theorem 7.2.2. *Suppose X is a topological space; $\mathcal{M} \subset C_B(X)$ separate and strongly separate points on X; and, for each $n \in \mathbb{N}$, $\mathcal{F}_n$ is uniform dense on compacts of $\mathbb{R}^n$. Then $\mathfrak{N}(\mathcal{M}, \{\mathcal{F}_n\}_{n=1}^{\infty})$ is a uniform dense subset of $C_U(X, \mathfrak{S}_{\mathcal{M}})$[2]. Additionally, if $\mathcal{M}$ is countable with cardinality $N \in \mathbb{N} \cup \{\infty\}$, then $\mathfrak{S}_{\mathcal{M}}$ is equivalent to the metric uniformity generated by the following metric:*

$$d(x, y) \mapsto \sum_{i=1}^{N} 2^{-i} \left(|g_i(x) - g_i(y)| \wedge 1\right) \qquad \forall x, y \in X. \tag{7.19}$$

Proof. By Proposition 4.3.3, $\bigotimes \mathcal{M}$ extends to a homeomorphism $\widehat{h}_{\mathcal{M}} \colon S \to \mathfrak{cl}[\bigotimes \mathcal{M}(X)]$ where $\mathfrak{cl}[\cdot]$ denotes closure in $\mathbb{R}^{\mathcal{M}}$ and S is compact.

For each $g \in \mathcal{M}$, define $\widehat{g} \doteq \pi_g \circ \widehat{h}_{\mathcal{M}}$. Then $\widehat{g}$ is a continuous extension of g to the compact set S. Define $\widehat{\mathcal{M}} \doteq \{\widehat{g} : g \in \mathcal{M}\}$. We find that $\bigotimes \widehat{\mathcal{M}}$ is $\widehat{h}_{\mathcal{M}}$, so it is a homeomorphism and by Proposition 4.3.3 we see that $\widehat{\mathcal{M}}$ separates points (and strongly separates points) on S.

Therefore, $\mathfrak{N}(\widehat{\mathcal{M}}, \{\mathcal{F}_n\}_{n=1}^{\infty})$ is a uniform dense subset of $C(S)$ by Theorem 7.1.5, from which it then follows that $\mathfrak{N}(\mathcal{M}, \{\mathcal{F}_n\}_{n=1}^{\infty})$ is a uniform dense subset of $C(S)|_X = C_U(X, \mathfrak{S}_{\mathcal{M}})$ by Proposition 5.3.8.

Further, $\widehat{\mathcal{M}}$ has the same cardinality as $\mathcal{M}$, so when $\mathcal{M}$ is countable, it follows by Proposition 4.3.4 that S is metrized by the following metric:

$$\widehat{d}(x, y) \mapsto \sum_{i=1}^{N} 2^{-i} \left(|\widehat{g}_i(x) - \widehat{g}_i(y)| \wedge 1\right) \qquad \forall x, y \in S. \tag{7.20}$$

[2] $\mathfrak{S}_{\mathcal{M}}$ was defined in Notation 5.3.15.

Proposition 5.3.7 (1) implies that $\mathfrak{S}_{\mathcal{M}}(S)$ is exactly the metric uniformity $\mathscr{U}(S;\widehat{d})$ as it is unique. Since $\widehat{g}_i|_X = g_i$ for each i, the metric d is just

$$d(x,y) \mapsto \widehat{d}(x,y) \qquad\qquad \forall x,y \in X. \tag{7.21}$$

So by Proposition 5.2.19 the subspace uniformity on X inherited from S is the metric uniformity generated by d, which is to say that $\mathscr{U}_S(X) = \mathfrak{S}_{\mathcal{M}} = \mathscr{U}_d(X)$. ■

Remark 7.2.3. Even though we did not specifically say X is a Tychonoff space, Proposition 4.1.7 tells us it has to be as there is some collection of continuous functions that separate and strongly separate points on X.

The previous result showed that any particular uniformly continuous function may be approximated with a finite number of g's. So how can we be sure that we are capable of picking the correct ones for a given approximation task considering there may be an infinite number of g's to pick from? The next result shows that when $\mathcal{M}$ is countable, the way in which we choose the g's is irrelevant for function approximation (assuming we can pick a large enough number of them).

Theorem 7.2.4. *Suppose X is a topological space, and let $\mathcal{M} = \{g_i\}_{i=1}^N \subset C_B(X)$ be countable where $N \in \mathbb{N} \cup \{\infty\}$. Suppose that, for each $n \in \mathbb{N}$, $\mathcal{F}_n$ is uniform dense on compacts of $\mathbb{R}^n$. Then the following functions:*

$$\mathfrak{M}_n \doteq \left\{p \mapsto f(g_1(p), \ldots, g_n(p)) : f \in \mathcal{F}_n\right\} \tag{7.22}$$

have the property that $\mathfrak{M}_{n+1}$ is uniform dense in $\mathfrak{M}_n$.

Proof. Let $\mathcal{M}_n \doteq \{g_i\}_{i=1}^n$. Then, since each g_i is bounded, we see that for each $n \in \mathbb{N}$, $\bigotimes \mathcal{M}_n(X) \subset K_n$ for some compact $K_n \subset \mathbb{R}^n$. Now suppose $h \in \mathcal{F}_n|_{K_n}$, so $h\colon K_n \to \mathbb{R}$ is continuous. Observe the following function $h'\colon K_{n+1} \to \mathbb{R}$ defined as

$$h'(x) \mapsto h(\pi_1(x), \ldots, \pi_n(x)) \tag{7.23}$$

for each $x \in K_{n+1}$. Since h' is continuous on the compact $K_{n+1} \subset \mathbb{R}^{n+1}$, then given $\epsilon > 0$, there exists some function $f \in \mathcal{F}_{n+1}|_{K_{n+1}}$ such that

$$\epsilon > \sup\{|f(x) - h'(x)| : x \in K_{n+1}\} \tag{7.24}$$
$$> \sup\{|f(g_1(p), \dots, g_{n+1}(p)) - h'(g_1(p), \dots, g_{n+1}(p))| : p \in X\} \tag{7.25}$$
$$= \sup\{|f(g_1(p), \dots, g_{n+1}(p)) - h(g_1(p), \dots, g_n(p))| : p \in X\}, \tag{7.26}$$

implying $\mathfrak{M}_{n+1}$ is uniform dense in $\mathfrak{M}_n$. ■

So adding more g's only enriches the class of functions in which we can approximate.

7.3 Spaces of Measures

In this section we show how to apply Theorem 7.2.2 to spaces of probability and finite measures. We also provide a few examples of how one can use the coming result for universal approximation on spaces of measures.

Recall the following notation introduced in Proposition 6.3.5. Given a topological space E; bounded function g_0; and collection $\mathcal{M} \subset C_B(E)$, the following are functionals on $\mathscr{M}^+(E)$

$$\mathfrak{W}_{g_0}[\mathcal{M}] \doteq \left\{\mu \mapsto g^*\left(\frac{\mu}{\mu(E)}\right) : g \in \mathcal{M}, g \neq 1\right\} \cup \{\mu \mapsto g_0(\mu(E))\}. \tag{7.27}$$

Definition 7.3.1 (Distributional Neural Network)**.** Suppose E is a topological space; $g_0 \in C_B((0,\infty))$; for each $n \in \mathbb{N}$, $\mathcal{F}_n \subset C(\mathbb{R}^n)$; and $\mathcal{M} \subset C_B(E)$. Let $\mathfrak{D}_{g_0}(\mathcal{M}, \{\mathcal{F}_n\}_{n=1}^{\infty})$ denote the following collection of mappings:

$$\mu \mapsto f\left(g_0\left(\mu\left(E\right)\right), \int_E g_1 \frac{d\mu}{\mu(E)}, \dots, \int_E g_n \frac{d\mu}{\mu(E)}\right), \tag{7.28}$$

where $f \in \mathcal{F}_{n+1}$; $g_1, ..., g_n \in \mathcal{M}$; and $n \in \mathbb{N}$. We call functions of the form given by (7.28) *distributional neural networks* and $\mathfrak{D}_{g_0}(\mathcal{M}, \{\mathcal{F}_n\}_{n=1}^{\infty})$ is the collection of neural networks generated by $\mathcal{M}$ and $\{\mathcal{F}_n\}_{n=1}^{\infty}$.

Theorem 7.3.2. *Suppose E is a topological space; $g_0 \in C_B(E)$ s.p. and s.s.p. on $(0,\infty)$; $\mathcal{M} = \{g_i\}_{i=1}^N \subset C_B(E)$ s.p., s.s.p., is countable and closed under multiplication; and $\mathcal{F}_n \subset C(\mathbb{R}^n)$ is uniform dense on the compacts of $\mathbb{R}^n$ for each $n \in \mathbb{N}$. Then $\mathfrak{D}_{g_0}(\mathcal{M}, \{\mathcal{F}_n\}_{n=1}^\infty)$ is a uniform dense subset of $C_U(\mathscr{M}^+(E), \mathfrak{S}_{\mathfrak{W}_{g_0}[\mathcal{M}]})$. Additionally, $\mathfrak{S}_{\mathfrak{W}_{g_0}[\mathcal{M}]}$ is equivalent to the metric uniformity generated by the following metric:*

$$d(\mu,\nu) \mapsto \left(\begin{array}{c} |g_0(\mu(E)) - g_0(\nu(E))| \wedge 1 \\ + \sum_{i=2}^{N} 2^{-i} \left(\left| g_i^* \left(\frac{\mu}{\mu(E)} \right) - g_i^* \left(\frac{\nu}{\nu(E)} \right) \right| \wedge 1 \right) \end{array} \right) \tag{7.29}$$

for each $\mu, \nu \in \mathscr{M}^+(E)$.

Proof. By Proposition 6.3.5, $\mathfrak{W}_{g_0}[\mathcal{M}] \subset C_B(\mathscr{M}^+(E))$ is countable, s.s.p., and s.p. on $\mathscr{M}^+(E)$. So the result follows directly from Theorem 7.2.2. ∎

Next, we provide a Lemma that makes use of Proposition 4.4.1 that says if $\mathcal{M}$ s.s.p. or s.p. and $\mathcal{M}_0$ is uniform dense in $\mathcal{M}$, then $\mathcal{M}_0$ s.s.p. or s.p. also. We end up with a similar result for spaces of measures.

Lemma 7.3.3. *Let E be a metrizable topological space and assume $\mathcal{M}, \mathcal{M}_0 \subset C_B(E)$; $\mathcal{M}^*$ s.s.p. and s.p. on $\mathscr{P}(E)$; and $\mathcal{M}_0$ is uniform dense in $\mathcal{M}$. Then $\mathcal{M}_0^*$ s.s.p. and s.p. on $\mathscr{P}(E)$. Further, given $g_0 \in C_B((0,\infty))$ and $\mathcal{F}_n \subset C(\mathbb{R}^n)$ for each $n \in \mathbb{N}$, we find that $\mathfrak{D}_{g_0}(\mathcal{M}_0, \{\mathcal{F}_n\}_{n=1}^\infty)$ is uniform dense in $\mathfrak{D}_{g_0}(\mathcal{M}, \{\mathcal{F}_n\}_{n=1}^\infty)$.*

Proof. We have $\mathcal{M}_0$ is uniform dense in $\mathcal{M}$, so given $f \in \mathcal{M}$ and $\epsilon > 0$, there is a $g \in \mathcal{M}_0$ such that

$$\epsilon > \sup\{|f(p) - g(p)| : p \in E\}. \tag{7.30}$$

Therefore,

$$\left| \int_E f\,dP - \int_E g\,dP \right| \leq \int_E |f(p) - g(p)|\,dP \tag{7.31}$$

$$< \int_E \epsilon\,dP \tag{7.32}$$

$$= \epsilon \tag{7.33}$$

holds for each $P \in \mathscr{P}(E)$. which is to say that $\mathcal{M}_0^*$ is uniform dense in $\mathcal{M}^*$, so it s.s.p. and s.p. on $\mathscr{P}(E)$ too by Proposition 4.4.1. As such, we can pick some $g_0 \in C_B((0,\infty))$ that s.p. and s.s.p. and conclude that $\mathfrak{W}_{g_0}[\mathcal{M}_0]$ s.s.p. and s.p. on $\mathscr{M}^+(E)$.

Next, we would like to show $\mathfrak{D}_{g_0}(\mathcal{M}_0, \{\mathcal{F}_n\}_{n=1}^{\infty})$ is uniform dense in $\mathfrak{D}_{g_0}(\mathcal{M}, \{\mathcal{F}_n\}_{n=1}^{\infty})$. Let $\psi \in \mathfrak{D}_{g_0}(\mathcal{M}, \{\mathcal{F}_n\}_{n=1}^{\infty})$. Then for some $n \in \mathbb{N}$; $\left\{\widetilde{f}_i\right\}_{i=2}^n \subset \mathfrak{W}_{g_0}[\mathcal{M}]$; and $h \in C(\mathbb{R}^n)$, ψ has the following form:

$$\psi(\mu) \mapsto h\left(g_0(\mu(E)),\, \widetilde{f}_2(\mu), \ldots, \widetilde{f}_n(\mu)\right). \tag{7.34}$$

$$\tag{7.35}$$

Further, let $\mathcal{N} = \{\mu \mapsto g_0(\mu(E))\} \cup \left\{\widetilde{f}_i\right\}_{i=2}^n$ and realize that for some compact set $K \subset \mathbb{R}^n$ we have $\bigotimes \mathcal{N}(\mathscr{M}^+(E)) \subset K$. Hence, $h|_K$ is a continuous function on a compact set, so by Proposition 5.3.7, there is a unique uniformity compatible with the topology on K and $h|_K$ is uniformly continuous. As the topology is metrizable (inherited from $\mathbb{R}^n$), the unique uniformity is equivalent to the metric uniformity generated by any metric that generates the topology on K. Therefore, for any $\epsilon > 0$, there is a $\delta_\epsilon > 0$ such that

$$d(p,q) < \delta_\epsilon \quad \text{implies} \quad |h(p) - h(q)| < \epsilon \qquad\qquad p, q \in K, \tag{7.36}$$

where d is any metric that generates the topology on K; however, it will be of convenience to choose the following metric

$$d(p,q) \mapsto \sum_{i=1}^{n} |\pi_i(p) - \pi_i(q)| \tag{7.37}$$

which is a metric on $\mathbb{R}^n$ when n is finite (inspired by Proposition 3.3.8). Previously, we found that for each $\widetilde{f}_i$, there is a $\widetilde{g}_i \in \mathfrak{W}_{g_0}[\mathcal{M}_0]$ such that

$$\sup\left\{\left|\widetilde{f}_i(\mu) - \widetilde{g}_i(\mu)\right| : \mu \in \mathscr{M}^+(E)\right\} < \frac{\delta_\epsilon}{n}, \tag{7.38}$$

and letting $\mathcal{N}_0 = \{\mu \mapsto g_0(\mu(E))\} \cup \{\widetilde{g}_i\}_{i=2}^{n}$, we have

$$d\left(\bigotimes \mathcal{N}(\mu), \bigotimes \mathcal{N}_0(\mu)\right) = \sum_{i=2}^{n} |\widetilde{f}_i(\mu) - \widetilde{g}_i(\mu)| \tag{7.39}$$

$$< \delta_\epsilon \tag{7.40}$$

holds for all $\mu \in \mathscr{M}^+(E)$. Hence, $\mathfrak{D}_{g_0}(\mathcal{M}_0, \{\mathcal{F}_n\}_{n=1}^\infty)$ is uniform dense in $\mathfrak{D}_{g_0}(\mathcal{M}, \{\mathcal{F}_n\}_{n=1}^\infty)$. ■

What if we use topological neural networks to build distributional neural networks? That is what we do next. Take note that the "closed under addition constraint" has been removed from $\mathcal{M}$. We will talk more about why this may be of use to practitioners in the next chapter.

Theorem 7.3.4. *Suppose E is a topological space; $g_0 \in C_B(E)$ s.p. and s.s.p. on $(0, \infty)$; $\mathcal{M} = \{g_i\}_{i=1}^N$ is countable, s.p. and s.s.p. on E; and, for each $n \in \mathbb{N}$, $\mathcal{F}_n, \mathcal{H}_n \subset C(\mathbb{R}^n)$ are uniform dense on the compacts of $\mathbb{R}^n$. Then $\mathfrak{D}_{g_0}(\mathfrak{N}(\mathcal{M}, \{\mathcal{F}_n\}_{n=1}^\infty), \{\mathcal{H}_n\}_{n=1}^\infty)$ is a uniform dense subset of $C_U\big(\mathscr{M}^+(E), \mathfrak{S}_{\mathfrak{W}_{g_0}[C_U(E,\mathfrak{S}_\mathcal{M})]}\big)$.*

Proof. By Theorem 6.3.6, $\mathfrak{W}_{g_0}[C_U(E, \mathfrak{S}_\mathcal{M})]$ are bounded continuous functions that s.s.p. and s.p. on $\mathscr{M}^+(E)$, so it then follows by Theorem 7.3.2 that $\mathfrak{D}_{g_0}(C_U(E, \mathfrak{S}_\mathcal{M}), \{\mathcal{H}_n\}_{n=1}^\infty)$ is a uniform dense subset of $C_U\big(\mathscr{M}^+(E), \mathfrak{S}_{\mathfrak{W}_{g_0}[C_U(E,\mathfrak{S}_\mathcal{M})]}\big)$. Also, Theorem 7.2.2 says $\mathfrak{N}(\mathcal{M}, \{\mathcal{F}_n\}_{n=1}^\infty)$ is a uniform dense subset of $C_U(E, \mathfrak{S}_\mathcal{M})$. Hence, the result follows from Lemma 7.3.3(with $\mathcal{M}_0 = \mathfrak{N}(\mathcal{M}, \{\mathcal{F}_n\}_{n=1}^\infty)$ and $\mathcal{M} = C_U(E, \mathfrak{S}_\mathcal{M})$). ■

7.3.1 Examples

Example 7.3.5. Suppose $E = [0, 1]$, $x \in E$ and $g_i(x) = x^i$. As the identity function is a homeomorphism, we have that g_1 s.s.p. and s.p. on E. Define $\mathcal{M} = \{g_i : i \in \mathbb{N}\}$. Therefore, $\mathcal{M} \subset C_B(E)$ is countable, closed under multiplication, s.p and s.s.p on E. Let $g_0(x) = \arctan(x)$, which is 1-1 and s.s.p on $\mathbb{R}$ (as it is a homeomorphism) and we use the result from Hornik in Theorem 1.1.3 to select as our $\mathcal{F}_n$. Putting this all together, we have that functions of the following form

$$\sum_{j=1}^{n} \beta_j \sigma \left(a_j' \left(\arctan\left(\mu\left([0,1]\right)\right), \int x^1 \frac{d\mu}{\mu\left([0,1]\right)}, ..., \int x^k \frac{d\mu}{\mu\left([0,1]\right)} \right) - \theta_j \right), \tag{7.41}$$

where $k, n \in \mathbb{N}$, $a_j \in \mathbb{R}^{k+1}$, and $\beta_j, \theta_j \in \mathbb{R}$; are uniformly dense in $C_U(\mathscr{M}^+([0,1]), \mathfrak{S}_{\mathfrak{W}_{g_0}[\mathcal{M}]})$ by Theorem 7.3.2.

Example 7.3.6. Suppose $E = [0,1]^d$ and define $h_i \colon [0,1]^d \to \mathbb{R}$ as $h_i(x) \mapsto e^{\pi_i(x)}$. It then follows (via homeomorphisms) that the collection $H = \{h_i : i = 1, \ldots, d\}$ s.p. and s.s.p. on E. We could repeat the same technique as in Example 7.3.5 to get a collection that is countable and closed under multiplication, which would result in functions of the following form

$$x \mapsto e^{n_1\pi_1(x)+\ldots+n_d\pi_d(x)}; \tag{7.42}$$

however, we instead take inspiration from Ma et al. 2020 (presented in Chapter 1 of this document) and look at parameterized functions of the following form

$$x \mapsto e^{v'x} \tag{7.43}$$

where $v \in \mathbb{R}^d$ represent parameters and $'$ denotes transpose. Therefore, take $\mathcal{M} = \{x \mapsto e^{v'x} : v \in \mathbb{R}^d\}$ and we can build similar functions as in Example 7.3.5 to approximate the uniformly continuous functions on $\mathscr{M}^+([0,1]^d)$.

Chapter 8

Concluding Thoughts

Now that the theoretical work has been covered, we use our final chapter to bring forth some ideas inspired by our main results that may be of interest to practitioners and researchers. The first idea illustrates the importance of Theorem 7.3.4 for practitioners who would like to build distributional neural networks to approximate continuous functions of positive-finite measures. The second realization is more for researchers interested in deep sets, where we demonstrate their relatedness to distributional neural networks. Lastly, we briefly suggest some areas for new research based on topological neural networks.

8.1 For Practitioners

In this section, we discuss how Theorem 7.3.4 can aid the learning process by not requiring $\mathcal{M}$ to be closed under multiplication.

Suppose we wish to learn a function $\lambda \in C(E)$. Typically, the algorithm designer chooses a parameterized function λ_Ω, where $\Omega \in \mathbb{R}^N$ represents N parameters, and the designer wishes to find an optimal $\widetilde{\Omega}$ such that $\lambda_{\widetilde{\Omega}}$ is *closest* to λ than any other λ_Ω (typically guided by data). However, often the set of functions $\{\lambda_\Omega : \Omega \in \mathbb{R}^N\}$ is not uniform dense in $C(E)$, so there is typically some amount of approximation error in $\lambda_{\widetilde{\Omega}}$ that we are unable to remove.

Consider the case where $E = [0, 1]$ and the algorithm designer chooses to use a neural network of the kind found in Theorem 1.1.3. The designer would choose a particular $n \in \mathbb{N}$ and function σ, while $(\beta_j, a_j, \theta_j)_{j=1}^n \subset \mathbb{R}^{3n}$ would be parameters. The set of functions $\{\lambda_\Omega : \Omega \in \mathbb{R}^{3n}\}$, then becomes

$$\left\{ \sum_{j=1}^{n} \beta_j \sigma(a_j x - \theta_j); a_j, \beta_j, \theta_j \in \mathbb{R} \right\} \tag{8.1}$$

which is clearly not uniform dense in $C([0, 1])$ as it is unable to uniformly approximate (all) functions of the form

$$\left\{ \sum_{j=1}^{n+1} \beta_j \sigma(a_j x - \theta_j); a_j, \beta_j, \theta_j \in \mathbb{R} \right\}, \tag{8.2}$$

so there will likely be some unavoidable approximation error in $\lambda_{\widetilde{\Omega}}$ (note that there could be no approximation error in the case where $\lambda \in \{\lambda_\Omega : \Omega \in \mathbb{R}^{3n}\}$, though this is unlikely the case in most real world problems).

Where did this approximation error come from? It stems from the algorithm designer having to choose a particular n. Often n will be chosen as large as is practically feasible considering computational constraints; regardless, a decision must be made and will result in some amount of approximation error. The magnitude of the approximation error is typically unknown, but could be large enough to render $\lambda_{\widetilde{\Omega}}$ useless for a given application. In the particular case of choosing n, often algorithm designers will try several guesses from a carefully selected set (such as $n \in \{2^5, 2^6, 2^7, 2^8\}$), so they can observe when increasing n does not practically reduce the approximation error.

The act of increasing n to reduce approximation error serves as an example that may be used analogously across many other neural network architectures (such as the number of layers or type of activation function); however, sometimes these sorts of methods are difficult or impractical as we will soon see.

Recall Theorem 7.3.2 and the subsequent example in subsection 7.3.1, and suppose we wish to learn a function $\lambda \in C_U(\mathscr{M}^+([0, 1]))$. As the algorithm designer, we have to make several decisions when constructing our set $\{\lambda_\Omega : \Omega \in \mathbb{R}^N\}$. We need to choose a particular $n \in \mathbb{N}$ and activation function σ

like we did previously; however, we also need to choose a particular $k \in \mathbb{N}$ and functions $(g_1, \ldots, g_k) \subset \mathcal{M}$.

How do we go about picking $(g_1, \ldots, g_k)$? This is quite a difficult task as $\mathcal{M}$ is typically at least countably infinite due to it being closed under multiplication. In the particular case of the example in subsection 7.3.1, it may seem sensible to keep the powers small and choose $(x^1, x^2, \ldots, x^k)$; however, we could just as easily have chosen $(x^{100}, x^{200}, \ldots, x^{100k})$ or any other k functions from $\mathcal{M}$. What if perhaps we *need* x^{10^5} as one of our chosen functions? It is unclear how we would be able to practically diagnose the need for x^{10^5} using a technique similar to selecting n from a set like $\{2^5, 2^6, 2^7, 2^8\}$.

The problem is exacerbated as the dimmensionality of E increases. Suppose $D \in \mathbb{N}$, $E = [0,1]^D$, and $x \in [0,1]^D$. We can employ the same ideas from the example in subsection 7.3.1, except we choose

$$\mathcal{M} = \left\{ \prod_{i=1}^{D} x_i^{d_i} : d_i \in \mathbb{N} \cup \{0\} \right\} \tag{8.3}$$

which is closed under multiplication, s.p., and s.s.p. on $[0,1]^D$. How do we choose a good collection $(g_1, \ldots, g_k) \subset \mathcal{M}$? Again, it seems reasonable to try to keep the powers d_i small, but even by keeping $d_i \leq 2$ would require $k \geq 3^D$.

Interestingly, we can quite easily find a finite class of functions that s.p. and s.s.p. on $[0,1]^D$, but is not closed under multiplication. For example, we can simply use the projection functions $(\pi_1, \ldots, \pi_D)$. It is the need for being closed under multiplication that causes $\mathcal{M}$ to be so large. That is the main justification for Theorem 7.3.4. We will provide an example to see how this may be implemented.

8.1.1 Practical Example

Suppose $D \in \mathbb{N}$, $E = [0,1]^D$, and $\mathcal{M} = \{\pi_1, \ldots, \pi_D\}$. Therefore, $\mathcal{M} \subset C_B(E)$ is countable, s.p and s.s.p on E. Let $g_0(x) = \arctan(x)$, which is bounded, continuous, s.p., and s.s.p on $\mathbb{R}$.

For $\mathcal{F}_n$ and $\mathcal{H}_n$ we use the neural networks described by Hornik in Theorem 1.1.3. However, we use the multidimensional version of Hornik for the $\mathcal{F}_n$.

That is, neural networks from $\mathbb{R}^k$ to $\mathbb{R}^m$ are uniform dense on compacts of $\mathbb{R}^k$ (it is not hard to convince oneself that this is true). The multidimensional neural networks we use will have the following form:

$$f(x) \doteq \left[f_1(x), f_2(x), \ldots, f_m(x)\right]' \qquad ' \text{ is transpose, } x \in \mathbb{R}^k \tag{8.4}$$

$$f_i(x) \doteq \sum_{j=1}^{n} \beta_{i,j}\sigma(a_j' x - \theta_j) \qquad i = 1, \ldots, m \tag{8.5}$$

which allows us to represent the functions from Theorem 7.3.4 as

$$h\left(g_0(\mu(E)), \int_E f(g_1, \ldots, g_n)\, d\mu\right). \tag{8.6}$$

We can rewrite f in matrix notation as

$$f(x) \doteq \boldsymbol{B}\sigma(\boldsymbol{A}' x - \phi) \qquad ' \text{ is transpose, } x \in \mathbb{R}^k \tag{8.7}$$

$$\boldsymbol{A} = \begin{bmatrix} a_1 & \ldots & a_n \end{bmatrix} \in \mathbb{R}^{k \times n} \qquad \boldsymbol{B} = \begin{bmatrix} \beta_{1,1} & \ldots & \beta_{i,n} \\ \vdots & \ddots & \vdots \\ \beta_{m,1} & \ldots & \beta_{m,n} \end{bmatrix} \in \mathbb{R}^{m \times n} \tag{8.8}$$

$$\Theta = \begin{bmatrix} \theta_1 \\ \vdots \\ \theta_n \end{bmatrix} \in \mathbb{R}^n \tag{8.9}$$

where σ operates elementwise on vectors; that is, $\sigma\left(\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}\right) = \begin{bmatrix} \sigma(x_1) \\ \sigma(x_2) \\ \sigma(x_3) \end{bmatrix}$.

As the algorithm designer, we need to choose a particular $n_1, n_2, m \in \mathbb{N}$, and activation function σ. In order to construct our λ_Ω, we first define the following function $\psi_{\boldsymbol{A},\boldsymbol{B},\Theta} \colon \mathscr{M}^+(E) \to \mathbb{R}^{1+m}$ as

$$\psi_{\boldsymbol{A},\boldsymbol{B},\Theta}(\mu) \doteq \begin{bmatrix} \arctan\left(\mu\left([0,1]^D\right)\right) \\ \displaystyle\int_{[0,1]^D} \boldsymbol{B}\sigma\left(\boldsymbol{A}' \begin{bmatrix} \pi_1(x) \\ \vdots \\ \pi_D(x) \end{bmatrix} - \Theta\right) \frac{d\mu}{\mu\left([0,1]^D\right)} \end{bmatrix}, \tag{8.10}$$

which results in the following set of functions $\{\lambda_\Omega : \Omega \in \mathbb{R}^N\}$ given as

$$\left\{ \sum_{j=1}^{n_2} q_j \sigma \left(p_j' \, \psi_{\boldsymbol{A},\boldsymbol{B},\Theta}(\mu) - \phi_j \right) : \boldsymbol{A} \in \mathbb{R}^{D \times n_1}; \;\; \boldsymbol{B} \in \mathbb{R}^{m \times n_1} \right. \\ \left. \Theta \in \mathbb{R}^{n_1}; \;\; p_j \in \mathbb{R}^{1+m}; \;\; q_j, \phi_j \in \mathbb{R} \right\} \tag{8.11}$$

.

The main advantage of the functions listed in (8.11) is that we no longer require the algorithm designer to choose a small subset of functions $(g_1, \ldots, g_k)$ from a large class $\mathcal{M}$, but instead learn the best m functions (as determined by the data).

8.2 Deep Sets

Deep sets refer to neural network like functions whose inputs (or outputs) are sets. Some possible applications of deep sets include:

- online shopping where a customer may purchase multiple items in a single online order,
- sports analytics where the goal is to understand the effectiveness of different lineup combinations of players in team sports, and
- a computer player for card games where players are dealt a hand of cards.

In the following discussion, we will only focus on when the input is a set.

Let $[0,1]$ be the collection of all possible items that could be in a set, so $2^{[0,1]}$ represents the collection of all possible subsets of $[0,1]$. Then a *set function* is a real valued function with domain $2^{[0,1]}$. The works of Zaheer et al. 2017 and Wagstaff et al. 2019 both study the ability of neural networks to approximate set functions, for which their analysis does so through *permutation invariant functions*. A function $t\colon [0,1]^n \to \mathbb{R}$ is permutation invariant if, for any permutation p on n elements, it satisfies the following

$$t(x_1, \ldots, x_n) = t(x_{p(1)}, \ldots, x_{p(n)}), \tag{8.12}$$

with the intuition being that the order of objects in a set is irrelevant. Zaheer et al. 2017 have identified how to express permutation invariant functions in the following result.

Proposition 8.2.1. *A function* $t\colon [0,1]^n \to \mathbb{R}$ *is permutation invariant if and only if it can be represented as*

$$t(x_1, \dots, x_n) = \rho\left(\sum_{i=1}^{n} \phi(x_i)\right), \tag{8.13}$$

for some continuous functions $\phi\colon [0,1] \to \mathbb{R}^{n+1}$ *and* $\rho\colon \mathbb{R}^{n+1} \to \mathbb{R}$.

Proof. See Zaheer et al. 2017 Theorem 7. ■

The representation for the permutation invariant functions listed in Theorem 8.2.1 bear a striking resemblance to the neural networks described in (8.6). To see this further, let us reconsider the measure μ described in (1.14) from the particle filtering example except with the likelihood of each particle set to 1. For each set $\{x_i\}_{i=1}^n \subset \mathscr{R}_0([0,1])$, μ then becomes:

$$\mu(A) = \sum_{i=1}^{n} I_{x_i}(A), \tag{8.14}$$

for each measurable $A \in \mathscr{B}([0,1])$. Combining this representation with (8.6) and the fact that the identity function s.p. and s.s.p. on $[0,1]$ (as it is a homeomorphism) yields the following

$$h\left(g_0(\mu([0,1])), \int_{[0,1]} f(z)\, d\mu(z)\right) \tag{8.15}$$

$$=h\left(g_0(n), \sum_{i=1}^{n} f(x_i)\right) \tag{8.16}$$

$$=h\left(\psi\left(\sum_{i=1}^{n}(1, f(x_i))\right)\right) \tag{8.17}$$

where $g_0\colon \mathbb{R} \to \mathbb{R}$ is bounded, 1-1, and s.s.p. on $\mathbb{R}$; $f\colon [0,1] \to \mathbb{R}^n$ is a neural network like those in (8.5); and $\psi\colon \mathbb{R}^{n+1} \to \mathbb{R}^{n+1}$ is defined as

$$\psi(y) \mapsto \big(g_0(\pi_1(y)),\, \pi_2(y),\, \ldots,\, \pi_{n+1}(y)\big). \tag{8.18}$$

Clearly, (8.18) is of the same form as the functions in Theorem 8.2.1 where $\rho = h \circ \psi$ and $\phi = (1, f)$. So we have shown that deep sets can be thought of as a special case of Theorem 7.3.4.

In fact, it is perhaps more sensible to think of deep sets in terms of functions on spaces of positive-finite measures. Theorem 3.3 of Wagstaff et al. 2019 says that there exist set functions $t\colon 2^{[0,1]} \to \mathbb{R}$ which cannot be represented in the form of 8.13, which is not particularly surprising due to there being uncountably infinite subsets of $[0,1]$ (some of which may be non-measurable). However, the theory we have developed in this document is naturally able to handle the case of (measurable) infinite sets. Suppose $B \in \mathscr{B}([0,1])$ is a set we would like to input to a neural network. We can represent it as a measure defined as

$$\mu_B(A) \mapsto \ell(A \cap B), \tag{8.19}$$

where ℓ is the Lebesgue measure, and results in the following functions

$$h\left(g_0(\mu_B([0,1])), \int_{[0,1]} f(z)\, d\mu_B(z)\right) \tag{8.20}$$

$$=h\left(g_0(\ell(B)), \int_B f(z)\, d\ell(z)\right). \tag{8.21}$$

One must be careful with μ_B as it is incompatible with μ from (8.14) in the sense that if B is finite (or countable), then $\mu_B = \ell(B) = 0 \neq \mu$. We could try to make the finite and uncountably infinite cases more compatible by using the following measure:

$$\nu_B(A) \mapsto \begin{cases} \frac{1}{|B|} \sum_{x \in B} I_x(A), & B \text{ finite and non-empty} \\ \frac{\ell(A \cap B)}{\ell(B)}, & B \text{ uncountably infinite} \end{cases}, \tag{8.22}$$

where $|B|$ is the cardinality of B. It is clear that ν_B is a probability measure for each $B \in \mathscr{B}([0,1])$, so ν is more informative about the "spread" of the set rather than its size. Ultimately, the correct representation is likely to depend on the particular application at hand.

8.3 Conclusions and Future Research

We have come along way. From topologies, to homeomorphisms and point separation, to uniformities (which come with their own topologies), to measure spaces and even topologies on measures. It has been quite the journey. And we achieved what we sought out to do. Theorem 7.2.2 gave us universal approximation on (non-compact) Tychonoff spaces and we applied the result to find two Theorems for spaces of measures. The first one is Theorem 7.3.2, and the second is Theorem 7.3.4 which was discovered with practitioner concerns in mind. But, how can we extend our work further? What is next?

Let us look back at the previous section as inspiration for future research. Deep sets were developed purposefully to have neural networks with permutation invariant inputs, which is a useful property unto itself. Connecting deep sets to functions on positive-finite measures is interesting particularly because it was unexpected. This all came about by attempting to pass a positive-finite measure into a neural network, so it begs the question: what would happen if we were to choose some other mathematical object? Would we discover new types of neural networks with their own interesting properties? The main tool needed to embark on such an endeavor is provided in Theorem 7.2.2 for universal approximation on Tychonoff spaces. One can dream up their own mathematical object, define a topology on the collection of them, find a class of continuous functions which separate and strongly separate points, and then Theorem 7.2.2 gives us universal approximation. The possibilities are endless.

Bibliography

[ABM85] P Antosik, TK Boehme, and F Mohanadi. “Concerning Sequential Spaces”. In: (1985).

[EK86] Stewart N. Ethier and Thomas G. Kurtz. *Markov processes : characterization and convergence.* Wiley, 1986. ISBN: 9780470316658.

[Cyb89] George Cybenko. “Approximation by superpositions of a sigmoidal function”. In: *Mathematics of control, signals and systems* 2.4 (1989), pp. 303–314.

[Hor91] Kurt Hornik. “Approximation capabilities of multilayer feedforward networks”. In: *Neural networks* 4.2 (1991), pp. 251–257.

[Rud91] Walter Rudin. *Functional analysis*. 2nd ed. McGraw-Hill, 1991.

[Mun00] James R Munkres. *Topology*. 2nd ed. Prentice Hall, 2000.

[Dud02] R. M. Dudley. *Real analysis and probability.* Cambridge studies in advanced mathematics: 74. Cambridge University Press, 2002. ISBN: 9780511755347.

[Wil04] Stephen Willard. *General Topology*. Dover Publications, 2004.

[BK10] Douglas Blount and Michael Kouritzin. “On convergence determining and separating classes of functions”. In: *Stochastic processes and their applications* 120.10 (2010), pp. 1898–1907.

[Ver10] Stijn Vermeeren. “Sequences and nets in topology”. In: *arXiv preprint arXiv:1006.4472* (2010).

[Kou16] Michael Kouritzin. “On tightness of probability measures on Skorokhod spaces”. In: *Transactions of the American Mathematical Society* 368.8 (2016), pp. 5675–5700.

[Zah+17] Manzil Zaheer et al. “Deep sets”. In: *Advances in neural information processing systems* 30 (2017).

[Wag+19] Edward Wagstaff et al. “On the limitations of representing functions on sets”. In: *International Conference on Machine Learning*. PMLR. 2019, pp. 6487–6494.

[DK20] Chi Dong and Michael Kouritzin. "Replication and Its Application to Weak Convergence". In: *arXiv preprint arXiv:2011.00484* (2020).

[KL20] Patrick Kidger and Terry Lyons. "Universal approximation with deep narrow networks". In: *Conference on learning theory*. PMLR. 2020, pp. 2306–2327.

[Ma+20] Xiao Ma et al. "Discriminative particle filter reinforcement learning for complex partial observations". In: *arXiv preprint arXiv:2002.09884* (2020).

Appendix A

Stone-Weierstrass for Real Functions

A.1 Background

The Stone-Weierstrass Theorem provides conditions for an algebra of functions to be uniformly dense in the set of continuous functions. However, there are many different versions that either rely on different assumptions or are for different domains or codomains. In this appendix, we demonstrate the Stone-Weierstrass version from Rudin 1991 for complex functions implies the version used in section 7.

First, we introduce some notation and terminology used to address complex functions. A complex number is expressed as $z = a + ib$ where $i \doteq \sqrt{-1}$ is the imaginary unit and $a, b \in \mathbb{R}$ are the real and imaginary components of z, respectively. Letting $\mathbb{C}$ be the set of complex numbers, we use $Re\colon \mathbb{C} \to \mathbb{R}$ and $Im\colon \mathbb{C} \to \mathbb{R}$ to denote the functions which extract the real or imaginary components of its input ($Re(z) \mapsto a$ and $Im(z) \mapsto b$). The complex conjugate of a complex number z is denoted $\bar{z} \doteq Re(z) - iIm(z)$. The barred notation is used for the conjugate of complex valued functions as well.

We now state the Stone-Weierstrass Theorem from section 5.7 of Rudin 1991.

Theorem A.1.1. *Let X be a compact Hausdorff space and let $C(X, \mathbb{C})$ be the* *set of complex continuous functions on X equipped with the uniform norm. Suppose that:*

1. *A is a closed subalgebra of $C(X, \mathbb{C})$*
2. *$f \in A$ implies $\bar{f} \in A$*
3. *A separates points on X*
4. *At every $p \in X$, $f(x) \neq 0$ for some $f \in A$*

Then $A = C(X, \mathbb{C})$.

A.2 Proof of Theorem 7.1.2

We are ready to justify the real-valued version as stated in Theorem 7.1.2.

Proof. First observe the following:

$$Re\big(C(X, \mathbb{C})\big) \doteq \{Re(f) \mid f \in C(X, \mathbb{C})\} = C(X, \mathbb{R}). \tag{1.1}$$

Now let B be a closed subalgebra of $C(X, \mathbb{R})$ that satisfies conditions 3 and 4 of Theorem A.1.1. Define the following:

$$A \doteq \big\{f + ig : f, g \in B\big\}. \tag{1.2}$$

A meets the conditions of Theorem A.1.1, therefore $A = C(X, \mathbb{C})$ and by (1.1) we have $B = Re(A) = C(X, \mathbb{R})$. ■