The Stone-Weierstraß Theorem and Its Applications

Julian Schennach

March 17, 2024

1 Introduction

Working with general continuous real-valued functions can be quite complicated, as we may not necessarily know how those functions behave in certain regions. Applying them in ergodic theory, then, becomes problematic. Polynomials, however, are rather predictable and easy to calculate and manipulate. It would be desirable if we can reduce examining continuous functions to examining those convenient polynomial functions - and we can in fact do so by arbitrarily closely approximating functions by polynomials. More formally, the Weierstraß Approximation Theorem states the following:

Theorem 1 (Weierstraß Approximation Theorem)**.** Let $f : [a, b] \to \mathbb{R}$ be a continuous function. For any $\epsilon > 0$, there exists a polynomial function p such that, for all $x \in [a, b]$, we have $|f(x) - p(x)| < \epsilon$.

The Stone-Weierstraß Theorem is much more general than the above theorem, proving that not only polynomials, but any set of functions that "separate points" can approximate a real-valued function:

Definition 1. A set P *separates points* in X if, for any $x, y \in X$, there exists a function $p \in P$ such that $p(x) \neq p(y)$.

Theorem 2 (Stone-Weierstraß Theorem)**.** Whenever X is a compact Hausdorff space and P is a closed unital subalgebra of $C(X, \mathbb{R})$, the set P is dense in $C(X, \mathbb{R})$ if and only if P separates points in X.

The statement "P is dense in $C(X, \mathbb{R})$" is equivalent to "functions in P approximate any continuous function from X to $\mathbb{R}$ arbitrarily closely."

In this paper, we will be proving both theorems. Specifically, we will be using the classic constructive proof by Bernstein for the Weierstraß Approximation Theorem, and for the general Stone-Weierstraß Theorem, we will use a proof using lattices, sets of functions where, for any elements f, g in the lattice, $\min_{fg}(x) = \min(f(x), g(x))$ and $\max_{fg}(x) = \max(f(x), g(x))$ are also in the lattice.

Afterwards, we will be applying both theorems in order to not only determine whether we can check unique ergodicity by just using polynomials or other dense sets of functions, but also to introduce Fourier series and approximations of functions by *trigonometric* polynomials.

2 Weierstraß Approximation Theorem

We first prove the Weierstraß Approximation Theorem, which is the more intuitive and straightforward of the two theorems. Notice that this theorem states that polynomials converge *uniformly* to continuous functions, with the rate of convergence being the same for all $x \in [a, b]$. Without loss of generality, we will assume that $a = 0$ and $b = 1$. Indeed, we can always scale $f(x)$ for $x \in [a, b]$ down to a continuous function $f((b-a)x + a)$ over $[0, 1]$, and scale the approximating polynomial $p(x)$ for $x \in [0, 1]$ up to the polynomial $p(\frac{x-a}{b-a})$. Before we continue with the proof, which we adapt from Bernstein's original paper [1], we define the polynomials that approximate the continuous functions over $[0, 1]$, the Bernstein polynomials.

Definition 2. The *nth Bernstein polynomial* of a continuous function f (over $[0, 1]$) is expressed as:

$$B_n(f, x) = \sum_{k=0}^{n} f\left(\frac{k}{n}\right)\binom{n}{k} x^k (1-x)^{n-k}.$$

Note that Bernstein used the notation E_n, not B_n, for his polynomials, as he used probabilistic reasoning. In that context, the Bernstein polynomials represent an expected winnings from a game where the prize for an event, of probability x, occurring exactly k times in n attempts is $f\left(\frac{k}{n}\right)$. Indeed, it makes sense that the expected winnings would converge to $f(x)$ (representative of the expected price for an event occurring a fraction x of the time) when n goes to ∞.

Proof of Theorem 1. We want to prove that $\lim_{n\to\infty} B_n(f, x) = f(x)$ uniformly for every $x \in [0, 1]$. Since f is continuous, and, on the compact set $[0, 1]$, uniformly continuous, for any $\epsilon > 0$, there exists a $\delta_\epsilon > 0$ such that whenever $|x - y| < \delta_\epsilon$, $|f(x) - f(y)| < \epsilon$. Let $\|f\|_\infty$ be the "supremum norm" of f, i.e. the supremum of $|f(x)|$ for any $x \in [0, 1]$ (or, in general, any x in a set X). $\|f\|_\infty$ is finite because $f(x)$ must exist for all $x \in [0, 1]$ and because $[0, 1]$ is compact. Select some constant $c \in [0, 1]$. Either $|f(x) - f(c)| < \frac{\epsilon}{2}$ when $|x - c| < \delta_{\frac{\epsilon}{2}}$, or $|f(x) - f(c)| \leq |f(x)| + |f(c)| \leq 2\|f\|_\infty \leq 2\|f\|_\infty(\frac{x-c}{\delta_{\frac{\epsilon}{2}}})^2 + \frac{\epsilon}{2}$ when $|x - c| \geq \delta_{\frac{\epsilon}{2}}$. Thus, the inequality $|f(x) - f(c)| \leq 2\|f\|_\infty(\frac{x-c}{\delta_{\frac{\epsilon}{2}}})^2 + \frac{\epsilon}{2}$ holds true for all $x \in [0, 1]$.

Notice that:

$$\begin{aligned} B_n(f - f(c), x) &= \sum_{k=0}^{n} (f - f(c))\left(\frac{k}{n}\right)\binom{n}{k} x^k (1-x)^{n-k} \\ &= \sum_{k=0}^{n} f\left(\frac{k}{n}\right)\binom{n}{k} x^k (1-x)^{n-k} - \sum_{k=0}^{n} f(c)\binom{n}{k} x^k (1-x)^{n-k} \\ &= B_n(f, x) - f(c)\left(\sum_{k=0}^{n}\binom{n}{k} x^k (1-x)^{n-k}\right) \\ &= B_n(f, x) - f(c)(x + 1 - x)^n = B_n(f, x) - f(c). \end{aligned}$$

We can now derive that:

$$\begin{aligned}
|B_n(f,x) - f(c)| &= |B_n(f - f(c), x)| \\
&= |\sum_{k=0}^{n} \left(f\left(\frac{k}{n}\right) - f(c) \right) \binom{n}{k} x^k (1-x)^{n-k}| \\
&\leq \sum_{k=0}^{n} \left(2\|f\|_\infty \left(\frac{\frac{k}{n} - c}{\delta_{\frac{\epsilon}{2}}} \right)^2 + \frac{\epsilon}{2} \right) \binom{n}{k} x^k (1-x)^{n-k} \\
&= \frac{2\|f\|_\infty}{\delta_{\frac{\epsilon}{2}}^2} \sum_{k=0}^{n} \left(\frac{k}{n} - c \right)^2 \binom{n}{k} x^k (1-x)^{n-k} + \frac{\epsilon}{2}(x + 1 - x)^n \\
&= \frac{2\|f\|_\infty}{\delta_{\frac{\epsilon}{2}}^2} \sum_{k=0}^{n} \left(\frac{k}{n} - c \right)^2 \binom{n}{k} x^k (1-x)^{n-k} + \frac{\epsilon}{2}.
\end{aligned}$$

We wish to evaluate $\sum_{k=0}^{n} (\frac{k}{n} - c)^2 \binom{n}{k} x^k (1-x)^{n-k}$. We have:

$$\sum_{k=0}^{n} \left(\frac{k}{n} - c \right)^2 \binom{n}{k} x^k (1-x)^{n-k}$$

$$\begin{aligned}
&= \sum_{k=0}^{n} \left(\frac{k}{n} \right)^2 \binom{n}{k} x^k (1-x)^{n-k} - \sum_{k=0}^{n} \frac{2kc}{n} \binom{n}{k} x^k (1-x)^{n-k} + \sum_{k=0}^{n} c^2 \binom{n}{k} x^k (1-x)^{n-k} \\
&= \sum_{k=0}^{n} \frac{k}{n^2} \binom{n}{k} x^k (1-x)^{n-k} + \sum_{k=0}^{n} \frac{k(k-1)}{n^2} \binom{n}{k} x^k (1-x)^{n-k} - 2cx \sum_{k=1}^{n} \binom{n-1}{k-1} x^{k-1} (1-x)^{n-k} + c^2 \\
&= \frac{x}{n} \sum_{k=1}^{n} \binom{n-1}{k-1} x^{k-1} (1-x)^{n-k} + \frac{(n-1)x^2}{n} \sum_{k=2}^{n} \binom{n-2}{k-2} x^{k-2} (1-x)^{n-k} - 2cx + c^2 \\
&= \frac{x}{n} + \frac{(n-1)x^2}{n} - 2cx + c^2.
\end{aligned}$$

Thus:

$$\begin{aligned}
|B_n(f,x) - f(c)| &\leq \frac{2\|f\|_\infty}{\delta_{\frac{\epsilon}{2}}^2} \sum_{k=0}^{n} \left(\frac{k}{n} - c \right)^2 \binom{n}{k} x^k (1-x)^{n-k} + \frac{\epsilon}{2} \\
&= \frac{2\|f\|_\infty}{\delta_{\frac{\epsilon}{2}}^2} \left(\frac{x}{n} + \frac{(n-1)x^2}{n} - 2cx + c^2 \right) + \frac{\epsilon}{2}.
\end{aligned}$$

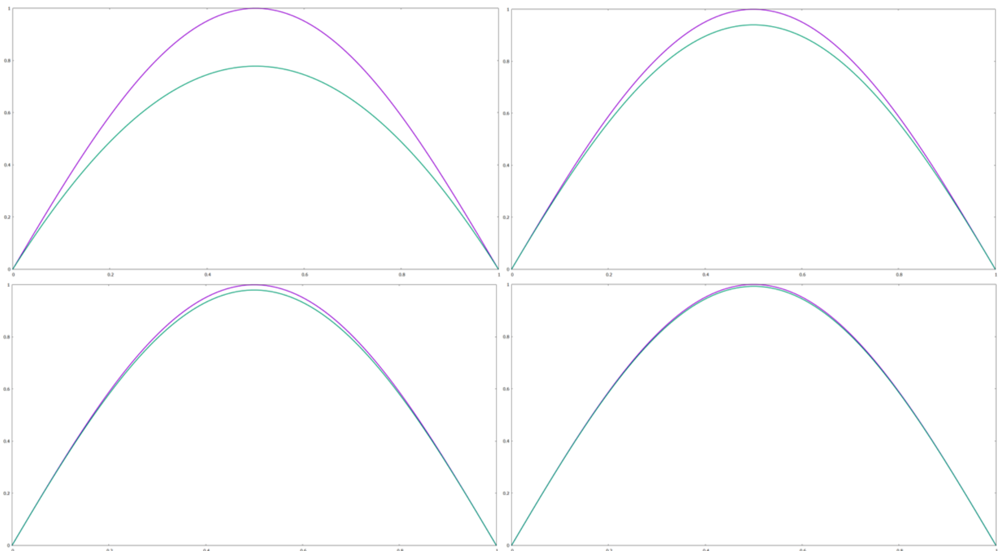
For $x = c$, we have $|B_n(f,c) - f(c)| \leq \frac{2\|f\|_\infty}{\delta_{\frac{\epsilon}{2}}^2}(\frac{c}{n} + \frac{(n-1)c^2}{n} - 2c^2 + c^2) + \frac{\epsilon}{2} = \frac{2\|f\|_\infty}{\delta_{\frac{\epsilon}{2}}^2}(\frac{c-c^2}{n}) + \frac{\epsilon}{2}$. We know that the maximum for $c - c^2 = c(1-c)$ occurs at $c = \frac{1}{2}$ and evaluates to $\frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$, so:

$$|B_n(f,c) - f(c)| \leq \frac{2\|f\|_\infty}{\delta_{\frac{\epsilon}{2}}^2} \left(\frac{c - c^2}{n} \right) + \frac{\epsilon}{2} \leq \frac{\|f\|_\infty}{2n\delta_{\frac{\epsilon}{2}}^2} + \frac{\epsilon}{2}.$$

For $n > \frac{\|f\|_\infty}{\epsilon\delta_{\frac{\epsilon}{2}}^2}$, $\frac{\|f\|_\infty}{2n\delta_{\frac{\epsilon}{2}}^2} < \frac{\epsilon}{2}$. So, for any ϵ, $|B_n(f,c) - f(c)| < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon$ for any sufficiently large n. In other words, we indeed have $\lim_{n\to\infty} B_n(f,x) = f(x)$ and the proof of the Weierstraß Approximation Theorem is complete.

□

Example 1. We can approximate $\sin(\pi x)$ on $[0, 1]$ by the respective Bernstein polynomials $B_n(f, x) = \sum_{k=0}^{n} \sin\left(\pi \frac{k}{n}\right) \binom{n}{k} x^k (1-x)^{n-k}$ for various values of n $(5, 20, 60, 170)$:



3 Definitions for the Stone-Weierstraß Theorem

Unlike the Weierstraß Approximation Theorem, the Stone-Wererstraß Theorem applies to functions that do not necessarily have the reals as their domain. Thus, we must first define some algebraic terms and structures.

Definition 3. A collection T of subsets of X is named a *topology* whenever

1. $\emptyset$ and X are in T,
2. any union of elements of T is in T, and
3. any finite intersection of elements of T is in T.

We then call (X, T) (or, shortly, X) a *topological space* and elements of T are called *open*, while their complements are *closed*.

The topology on $\mathbb{R}$ that we usually use is simply the one constructed by taking the unions or intersects of open intervals of $\mathbb{R}$. However, for the conditions on X in the Stone-Weierstraß Theorem, we must still define two more terms:

Definition 4. A topological space X is *compact* if, for any collection of open sets whose union covers X (i.e. for any open cover of X), we can select finitely many of those open sets that also cover X (i.e. there exists a finite subcover).

In the reals, it suffices that X is closed and bounded, and therefore, closed intervals $[a, b]$, which are compact, are used in the Weierstraß Approximation Theorem.

There are useful special properties of compact sets. For instance, closed subsets A of a compact set X are compact: Select any open cover of A. Include the open set $X \backslash A$ in the cover, forming a cover of X. Since X is compact, there exists a finite subcover. If $X \backslash A$ is not in the subcover, we have a finite subcover of the original open cover for A. If $X \backslash A$ is included, we can just remove it from the subcover to get a finite subcover for A. Hence, A is compact.

Definition 5. A topological space X is called *Hausdorff* if, for any distinct elements x, y of X, there exist disjoint open sets U_x, U_y that include x,y, respectively.

We continue by defining continuity (forgive the pun) not only for functions from the reals to reals, but from any topological space to another.

Definition 6. For a *continuous function* f from one topology to another, the preimage of an open set f is also open. The set $C(X, \mathbb{R})$ is defined as the set of all continuous functions from X to $\mathbb{R}$.

If f and g are continuous, so is $f \circ g$: $(f \circ g)^{-1}(U) = g^{-1}(f^{-1}(U))$, where U is open, and due to the continuity of f and g, $f^{-1}(U)$ and $g^{-1}(f^{-1}(U))$ are also open. In addition, for $f, g \in C(X, \mathbb{R})$, $f + g$ and fg are continuous.

Furthermore, if A is compact and f is continuous, $f(A)$ is compact. Form any open cover of $f(A)$. Since f is continuous, the preimages of each open set of the cover are open and cover A. Since A is compact, there exists a finite subcover of these preimages for A. The respective images are a finite subcover for $f(A)$. So $f(A)$ is compact.

Now we must explain what a "subalgebra P of $C(x, \mathbb{R})$ is. The definition of a subalgebra depends on the definition of an algebra.

Definition 7. An *algebra* is a set closed under addition, multiplication, and scalar multiplication, and we can also include a multiplicative identity to form a *unital algebra*. A *subalgebra* is simply a subset of the algebra that is also closed under the same operations, and a *unital subalgebra* must also include the multiplicative identity.

We let the unital algebra on $C(X, \mathbb{R})$ be normal addition and multiplication for functions, with the multiplicative identity $1(x) = 1$ - indeed, the function 1 is continuous, and, for $f, g \in C(X, \mathbb{R})$, we have $f + g \in C(X, \mathbb{R})$ and $fg \in C(X, \mathbb{R})$. Scalar multiplication derives from letting g be a constant (and therefore continuous) function. The unital subalgebra P must be closed under this addition and multiplication, and must include the function 1.

However, for the proof of the theorem itself, we need one final definition, the concept of "lattices":

Definition 8. A **lattice** L is a set of real-valued functions so that, for any $f, g \in L$, $\min_{fg}(x) = \min(f(x), g(x))$ and $\max_{fg}(x) = \max(f(x), g(x))$ are in L as well.

Note that there are also general lattices that include any type of element that can be ordered (i.e. we know which element is greater or smaller). However, this specific version suffices for the Stone-Weierstraß Theorem, as we are indeed only considering real-valued functions.

4 Proof of the Stone-Weierstraß Theorem

Having explained all definitions, we can continue with some preliminary lemmas. The closure of P (denoted $\overline{P}$) is simply defined as the set P unioned with all limit points of P (i.e. limits of elements of P).

Lemma 1. When X is a compact Hausdorff space and P is a closed unital subalgebra of $C(X, \mathbb{R})$, $\overline{P}$ is a lattice.

Note that $\|p\|_\infty < \infty$ for $p \in P$: Because P is a closed subset of the compact $C(X, \mathbb{R})$, P is compact as well. Then, for the open cover $(c - \epsilon, c + \epsilon)$ for every constant function c (and a constant function ϵ), there exists a finite subcover $(c_i - \epsilon, c_i + \epsilon)$, the union of which includes all p for $p \in P$. So p (and $|p|$) must be bounded, and $\|p\|_\infty$ is bounded too.

Proof of Lemma 1. We first prove that for any element p of P, $|p|$ is also in $\overline{P}$. We now use the Weierstraß Approximation Theorem: We know that $|y|$ is continuous for real y. So, for any $\epsilon > 0$ and $y \in [-\|p\|_\infty, \|p\|_\infty]$, there exists a polynomial function q' such that $||y| - q'(y)| < \frac{\epsilon}{2}$. Let $q(y) = q'(y) - q'(0)$ (i.e. the polynomial q' with constant term 0). We know that $||0| - q'(0)| = |q'(0)| < \frac{\epsilon}{2}$. Hence, $||y| - q(y)| = ||y| - q'(y) + q'(0)| \leq ||y| - q'(y)| + |q'(0)| < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon$ for any $y \in [-\|p\|_\infty, \|p\|_\infty]$. This implies that $||p(x)| - q(p(x))| < \epsilon$ for any $x \in X$, as $p(x)$ always lies between $-\|p\|_\infty$ and $\|p\|_\infty$. Since $q(p) = a_1 p + a_2 p^2 + a_3 p^3 + \cdots$ for real a_i, and P is closed under addition and multiplication, $q(p)$ is in P. Since $|p|$ can be approximated arbitrarily closely by functions in P (i.e. $|p|$ is a limit point of $\overline{P}$), and $\overline{P}$ is closed, $|p|$ is in $\overline{P}$.

Finally, for $p_1, p_2 \in P$, we know that $\max_{p_1 p_2}(x) = \frac{p_1 + p_2 + |p_1 - p_2|}{2}$ and $\min_{p_1 p_2}(x) = \frac{p_1 + p_2 - |p_1 - p_2|}{2}$: When $p_1(x) \geq p_2(x)$, we determime that $\frac{p_1(x) + p_2(x) + (p_1(x) - p_2(x))}{2} = p_1(x)$ and $\frac{p_1(x) + p_2(x) - (p_1(x) - p_2(x))}{2} = p_2(x)$, and when $p_1(x) < p_2(x)$, we have $\frac{p_1(x) + p_2(x) - (p_1(x) - p_2(x))}{2} = p_2(x)$ and $\frac{p_1(x) + p_2(x) + (p_1(x) - p_2(x))}{2} = p_1(x)$, as desired. Since P is closed under addition and multiplication, and $|p_1 - p_2|$ is in $\overline{P}$, $\max_{p_1 p_2}(x)$ and $\min_{p_1 p_2}(x)$ must be in $\overline{P}$. Thus, $\overline{P}$ is a lattice.

□

Lemma 2. If X is a compact Hausdorff space and P is a closed unital subalgebra of $C(X, \mathbb{R})$ that separates points in X, P strongly separates points X. In other words, for any $x_1, x_2 \in X$ and $a_1, a_2 \in \mathbb{R}$, there exists a function $p \in P$ such that $p(x_1) = a_1$ and $p(x_2) = a_2$.

Proof of Lemma 2. We use a constructive proof: We know that there exists some $q \in P$ such that $q(x_1) \neq q(x_2)$. Thus, $q(x_1) - q(x_2) \neq 0$ and we can define:

$$p(x) = a_1 \frac{q(x) - q(x_2)}{q(x_1) - q(x_2)} + a_2 \frac{q(x) - q(x_1)}{q(x_1) - q(x_2)}.$$

This function is in P, because P is closed under addition and multiplication. Note that $\frac{1}{q(x_1)-q(x_2)}$ is constant, so even if we cannot divide by functions in P, we *can* multiply functions by the constant $\frac{1}{q(x_1)-q(x_2)}$. We have $p(x_1) = a_1\frac{q(x_1)-q(x_2)}{q(x_1)-q(x_2)} + a_2\frac{q(x_1)-q(x_1)}{q(x_1)-q(x_2)} = a_1$ and $p(x_2) = a_1\frac{q(x_2)-q(x_2)}{q(x_1)-q(x_2)} + a_2\frac{q(x_2)-q(x_1)}{q(x_1)-q(x_2)} = a_2$, so this $p \in P$ strongly separates points, as desired.

□

Now we are prepared for the full proof of the Stone-Weierstraß Theorem, which, after having proven all of the lemmas above, will be rather straightforward. We wish to prove the following theorem:

Theorem 2 (Stone-Weierstraß Theorem)**.** Whenever X is a compact Hausdorff space and P is a closed unital subalgebra of $C(X, \mathbb{R})$, the set P is dense in $C(X, \mathbb{R})$ if and only if P separates points in X.

We first quickly prove that if P is dense in $C(X, \mathbb{R})$, P separates points. The reasoning utilizes many of the special properties of X and compact sets in general.

Proof for Dense ⇒ Separation of Points. Select $x, y \in X$. Select some $f \in C(X, \mathbb{R})$ that may not separate x and y. Since X is Hausdorff, there exists an open set U_x including x that does not include y. X is closed (as its complement in X, $\emptyset$, is open), so $X \backslash U_x$ is closed as well. Closed subsets of a compact set are compact as well, and continuous functions map compact sets to compact sets. So $X \backslash U_x$ is compact and f maps it to a compact set in $\mathbb{R}$. $\mathbb{R}$ itself is not compact (as it is not bounded), so some real is not mapped from $X \backslash U_x$ by f. Define a new function g such that g maps the elements of $X \backslash U_x$ to the same outputs as f, but maps all of U_x to some real r not in $f(X \backslash U_x)$. This g is continuous: Select any open set $V \in \mathbb{R}$. If it doesn't include r, its preimage must be in $X \backslash U_x$ and is open by the continuity of f. If V includes r, the intersection $V \cap g(X \backslash U_x)$ is open and its preimage is in $X \backslash U_x$ and is open, by the continuity of f. The preimage of the remaining elements in V can only be U_x, as only the element in V that can potentially be outputted by g is r. Thus, g is continuous. And since $g(x) = r$ (as $x \in U_x$), and $g(y)$ cannot equal r (we defined $r \notin g(X \backslash U_x)$, but $g(y) \in g(X \backslash U_x)$), we have a continuous function $g \in C(X, \mathbb{R})$ that separates the points x, y. Finally, since P is dense in $C(X, \mathbb{R})$, there exists some $p \in P$ such that $|g(x) - p(x)| < \frac{|g(x)-g(y)|}{2}$ and $|g(y) - p(y)| < \frac{|g(x)-g(y)|}{2}$, or, in other words, $p(x) \in (g(x) - \frac{|g(x)-g(y)|}{2}, g(x) + \frac{|g(x)-g(y)|}{2})$ and $p(y) \in (g(y) - \frac{|g(x)-g(y)|}{2}), g(x) + \frac{|g(x)-g(y)|}{2})$. So $p(x) \neq p(y)$ as well, and P separates any points x, y.

□

Now we proceed with the converse, that if P separates points, P is dense in $C(X, \mathbb{R})$. This is usually the more useful direction of the Stone-Weierstraß Theorem - proving that P separates points means just finding valid elements, while proving that P is dense requires significantly more work. The following proof, along with the lemmas above, are adapted from Dividson and Donsig [2].

Proof for Separation of Points ⇒ Dense. We wish to prove that for any function $f \in C(X, \mathbb{R})$ and any $\epsilon > 0$, there exists a function $p \in \overline{P}$ such that $|f - p| < \epsilon$. In other words,

$f(x) - \frac{\epsilon}{2} < p(x) < f(x) + \frac{\epsilon}{2}$ for any $x \in X$. We will first find a function $q \in P$ such that $q(x) < f(x) + \frac{\epsilon}{2}$, and then find a function $p \in P$ such that $f(x) - \frac{\epsilon}{2} < p(x) \leq q(x) < f(x) + \frac{\epsilon}{2}$.

By Lemma 2, we know that for any $x_1, x_2 \in X$ and $a, b \in \mathbb{R}$, there exists a function $r \in P$ such that $r(x_1) = a$ and $r(x_2) = b$. Therefore, there exists a function $p_{y,z} \in P$ such that $p_{y,z}(y) = f(y)$ and $p_{y,z}(z) = f(z)$. Define the sets U_z as the sets of elements x for which $p_{y,z}(x) < f(x) + \frac{\epsilon}{2}$. In other words, U_z is $p_{y,z}^{-1}((-\infty, f(x) + \frac{\epsilon}{2}))$, and since f is continuous and $(-\infty, f(x) + \frac{\epsilon}{2})$ is open, its preimage U_z is open too. Also, since $p_{y,z}(z) = f(z) < f(z) + \frac{\epsilon}{2}$, z is included in U_z. Since z can be any element of X, the collection of open sets $U_{z \in Z}$ is an open cover for X. X is compact, so there exists a finite subcover U_{z_i}. The respective functions $p_{y,z}$ are p_{y,z_i}.

Define p_y as the minimum of all finitely many p_{y,z_i}. Since $\overline{P}$ is a lattice by Lemma 1, this minimum is in $\overline{P}$. Furthermore, $p_y(x) = p_{y,z_j}(x) < f(x) + \frac{\epsilon}{2}$ for the minimizing p_{y,z_j}. Now define V_y as $p_y^{-1}((f(x) - \frac{\epsilon}{2}, \infty))$, which, as before, are all open. Again, $y \in V_y$ as $p_y(y) = p_{y,z_i}(y) = f(y) > f(y) - \frac{\epsilon}{2}$. So the V_y form an open cover for X, and there is a finite subcover V_{y_i} with respective p_{y_i}. Define p' to be the maximum of all finitely many p_{y_i}. We have $p'(x) = p_{y_j}(x) < f(x) + \frac{\epsilon}{2}$ and $p'(x) = p_{y_j}(x) > f(x) + \frac{\epsilon}{2}$ for the maximizing p_{y_j}. Since $\overline{P}$ is a lattice, $p' \in \overline{P}$. Elements in $\overline{P}$, by the definition of the closure of P, can be approximated arbitrarily closely by elements of P. So we can find a function p such that $|p'(x) - p(x)| < \frac{\epsilon}{2}$ and such that $|f(x) - p(x)| \leq |f(x) - p'(x)| + |p'(x) - p(x)| < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon$ for any $x \in X$. Thus, P is dense in $C(X, \mathbb{R})$.

□

Corollary 1. Notice that we proved that, for any ϵ, there is some $p \in P$ for which $|f - p| < \epsilon$. In other words, $|f(x) - p(x)| < \epsilon$ for any $x \in X$, which implies that any function $f \in C(X, \mathbb{R})$ is not only a limit, but also a uniform limit of functions $p \in P$. This is useful for proofs regarding ergodicity.

Corollary 2. The Weierstraß Approximation Theorem is technically a special case of the Stone-Weierstraß Theorem: The set of polynomials P is a closed unital subalgebra of $C([0, 1], \mathbb{R})$, as it is closed under addition and multiplication, and includes the multiplicative identity (1 is a polynomial). Furthermore, the polynomial x separates any pair of nonidentical reals. So P is dense in $C([0, 1], \mathbb{R})$.

However, as we used the Weierstraß Approximation Theorem in our proof above, this reasoning is circular. Of course, we could prove that $|x|$ can be approximated by polynomials separately, and then prove the approximation theorem, but that ultimately requires more effort.

5 Applications

The Stone-Weierstraß Theorem allows us to prove certain statements regarding uniform limits of functions more easily by only considering the approximating functions (like polynomials). Unfortunately, many special properties of functions, such as measure-preservation or ergodicity, do not have a condition using limits of functions. However, one highly applicable theorem, Oxtoby's Theorem, actually does use limits of functions in order to prove *unique* ergodicity. A transformation is uniquely ergodic if it is ergodic for only one measure, and this property is equivalent to a condition that does not use measures, as Oxtoby's Theorem states:

Theorem 3. (Oxtoby's Theorem)

Let X be a compact Hausdorff space, and T be a continuous transformation from X to itself. T is uniquely ergodic if and only if, for any real-valued continuous function f on X, there exists a constant c_f such that:

$$\lim_{n\to\infty}\left\|\frac{1}{n}\sum_{i=0}^{n-1} f\circ T^i - c_f\right\|_\infty = 0.$$

In other words, the uniform limit of $\frac{1}{n}\sum_{i=0}^{n-1} f(T^i(x))$ is c_f. Note that c_f can change for different f.

While we will not prove Oxtoby's in its entirety here, we do prove that we must only consider $p \in P$ that approximate any continuous, real-valued f.

Proof of Sufficiency of Using Oxtoby's on Functions in P. First, we assume that, for $p \in P$ and some respective constant c_p, the equation $\lim_{n\to\infty} \|\frac{1}{n}\sum_{i=0}^{n-1} p\circ T^i - c_p\|_\infty = 0$ holds. For any function f and real $\epsilon > 0$, there exists some $p_\epsilon \in P$ such that $\|f - p_\epsilon\|_\infty < \epsilon$. So $\|f\circ T^i - p_\epsilon \circ T^i\|_\infty < \epsilon$ for any $0 \le i < n$. By the Triangle Inequality, we have:

$$\|\sum_{i=0}^{n-1} f\circ T^i - p_\epsilon\circ T^i\|_\infty \le \sum_{i=0}^{n-1}\|f\circ T^i - p_\epsilon\circ T^i\|_\infty < n\epsilon.$$

Hence:

$$\|\frac{1}{n}\sum_{i=0}^{n-1} f\circ T^i - p_\epsilon\circ T^i\|_\infty = \frac{1}{n}\|\sum_{i=0}^{n-1} f\circ T^i - p_\epsilon\circ T^i\|_\infty < \epsilon.$$

Finally, let n' be a sufficiently large positive integer such that $\|\frac{1}{n'}\sum_{i=0}^{n'-1} p_\epsilon\circ T^i - c_{p_\epsilon}\|_\infty < \epsilon$ for the same ϵ as above. Then, we can use the Triangular Inequality again to ultimately get $\|\frac{1}{n'}\sum_{i=0}^{n'-1} f\circ T^i - c_{p_\epsilon}\|_\infty \le \|\frac{1}{n'}\sum_{i=0}^{n'-1} p_\epsilon\circ T^i - c_{p_\epsilon}\|_\infty + \|\frac{1}{n'}\sum_{i=0}^{n'-1} f\circ T^i - p_\epsilon\circ T^i\|_\infty < \epsilon + \frac{\epsilon}{n}$. Since $\epsilon > 0$ is arbitrary, and the c_{p_ϵ} converge to a constant c_f, we conclude that $\lim_{n\to\infty}\|\frac{1}{n}\sum_{i=0}^{n-1} f\circ T^i - c_f\|_\infty = 0$, as desired. □

Sometimes, we do not want to just use polynomials from the Weierstraß Approximation Theorem. Instead, we'd often want to use a function based on sines and cosines, so that the integral over $[0,1)$ could be 0 (which makes the constant c_f in Oxtoby's Theorem nice). Fortunately, we can also approximate functions arbitrarily closely by what are called "trigonometric polynomials":

Definition 9. A *trigonometric polynomial* of degree n is of the form $\sum_{k=0}^{n} a_k\cos(kx) + b_k\sin(kx)$ for real a_i, b_i. Equivalently, they are expressible as $\sum_{k=-n}^{n} c_k e^{2i\pi kx}$ for potentially imaginary c_i, since the polynomials are equal to $\sum_{k=0}^{n}(a_k\frac{e^{2i\pi kx}+e^{-2i\pi kx}}{2} + b_k\frac{e^{2i\pi kx}+e^{-2i\pi kx}}{2i})$ by Euler's formula.

Proof That Trigonometric Polynomials Are Dense. We prove that trigonometric polynomials approximate functions in $C_p([0, 2\pi], \mathbb{R})$, which consists of *periodic* continuous functions. By the Stone-Weierstraß Theorem, we must simply prove that the set of trigonometric polynomials is a closed unital subalgebra of $C_p([0, 2\pi], \mathbb{R})$ and the set separates points (except the pair 0 and 2π, as they are equivalent in $C_p([0, 2\pi], \mathbb{R})$).

First, this set is clearly closed under addition, as the sum of powers of $e^{2i\pi x}$ remain the sum of powers of $e^{2i\pi x}$. Furthermore, the product of sums of powers of $e^{2i\pi x}$ is also of the form $\sum_{k=-n}^{n} c_k e^{2i\pi kx}$, as the product of pairs of individual terms (which ultimately sum to the full product) are sums of powers of $e^{2i\pi x}$. Furthermore, by letting $c_0 = 1$ but $c_i = 0$ for all other i, we obtain the identity function 1. Thus, the set of trigonometric polynomials is a closed unital subalgebra.

Next, we prove that that set separates points. When $x, y \in [0, \pi)$ or $x, y \in [\pi, 2\pi]$, we can use the trigonometric polynomial $\cos(x)$, which has different outputs for x, y unless $x = y$. Otherwise, when $x \in [0, \pi]$ and $y \in (\pi, 2\pi)$, or $y \in [0, \pi]$ and $x \in (\pi, 2\pi)$, we use $\sin(x)$, which, again, has different outputs except when $x = y$.

□

These dense trigonometric polynomials serve as the foundation for Fourier series. The issue is, of course, that the Stone-Weierstraß Theorem does not give us the approximating trigonometric polynomials. It turns out that the Fourier series of a real-valued function f (i.e. the trigonometric polynomial that approxmates f) on $[0, 2\pi]$ is defined as $\sum_{k=0}^{n} \left((\frac{1}{2\pi} \int_0^{2\pi} f(x) \cos(kx)\, dx) \cos(kx) + (\frac{1}{2\pi} \int_0^{2\pi} f(x) \sin(kx)\, dx) \sin(kx) \right)$.

If the coefficients for the sine and cosine terms converge to 0, we can approximate f by a Fourier series with a finite number of terms. This is especially useful when working with computers, as the latter can only have a finite amount of memory. Hence, whenever a computer uses an arbitrary continuous function (such as RGB colors in images), it transforms the function into the simpler, finite Fourier series approximation. While this transformation may result in some reduction in quality (in the case of images), it also allows the computer to store the data and perform calculations on the approximated function more easily. Fourier series are, therefore, so integral to all modern computer systems.

6 Conclusion

The Weierstraß Approximation Theorem, as well as its generalization, the Stone Weierstraß Approximation Theorem, allow us to prove statements on continuous real-valued functions by only checking that those statements are satisfied by certain sets of functions that separate points. A common application of these theorems is checking unique ergodicity of transformations. While simple polynomials can approximate continuous functions on closed intervals, trigonometric polynomials are often a more useful type of approximating function, and they form an essential part of Fourier series. These series, in turn, have applications within computers and various other pieces of technology.

References

[1] Sergei Bernstein. Démonstration du théorème de weierstrass fondée sur le calcul des probabilités. *Kharkov Mathematical Society*, 13:1–2, 1912.

[2] Kenneth R. Davidson and Allan P. Donsig. *Real Analysis and Applications*. Springer, New York, New York, 2010.