APPROXIMATING CONTINUOUS FUNCTIONS

SIDDHAARTH KOTHAPALLI

Abstract. Foundational approximation theorems such as the Weierstrass approximation theorem, the generalization of the Weierstrass approximation theorem to any finite number of dimensions, and the Trigonometric approximation theorem have applications ranging from theoretical physics to computer science. Moreover, all of these theorems are special cases of the Stone-Weierstrass theorem. In this paper, we shall explore the developments in mathematical analysis leading up to this theorem, provide a proof of the theorem, and examine a few consequences of it. We shall also see how this theorem generalizes to complex-valued functions.

1. Introduction

When we ask a computer to evaluate a complicated function at a given value of x, the computer usually provides us with a decimal approximation of the exact value almost instantaneously. Have you ever wondered how this decimal value is determined so quickly? Part of the reason is the immense computational power of the computer. However, the underlying mathematical methods of polynomial approximation play a bigger role. When the computer is asked to evaluate a complicated function at a given value, what it actually does is locally approximate that function with a polynomial, maintaining all but a minimal loss of accuracy. Furthermore, we can ask the computer to calculate the output values of a function over an entire interval. In this case, the computer resorts to global approximation methods, which differ quite a bit from local approximation methods. In most global polynomial approximation methods, the function being approximated is required to be continuous over the entire interval on which it is approximated.

Now, let us take a step back and ask the general question. How do we know that this is always possible? That is, *for an arbitrary continuous function, how do we know that we can always find a polynomial that is arbitrarily close to our function over an entire designated interval?* The answer to this question lies in the Weierstrass approximation theorem. The Weierstrass approximation theorem states that for any continuous function f and any closed and bounded interval in the domain of f, we can always find a polynomial p that approximates f over the entire interval at once, which means that the error differences introduced by this approximation across various values of x are all bounded by a single positive number. The lower this bound, the higher the accuracy of the approximation of f. Since every finite open interval (a, b) is contained in the finite closed interval $[a, b]$, this theorem effectively guaranties that even the most complicated continuous functions can be approximated with polynomials as closely as desired over any finite interval in the function's domain. This

Date: July 14, 2026.

makes the Weierstrass approximation theorem a very powerful result in mathematical analysis. The fact that the function f need not necessarily be differentiable but only continuous makes it a much more general result than other polynomial approximation theorems.

The Weierstrass approximation theorem was first formulated by Karl Weierstrass in 1885, and subsequently generalized by Marshall H. Stone in 1937 to the Stone-Weierstrass theorem. The originality of Stone's generalization comes from the fact that rather than trying to expand the Weierstrass approximation theorem analytically like other mathematicians, he attempted to expand it topologically. Stone's generalization states that if X is a compact Hausdorff space, and if A is a subset of the set of all continuous functions defined on X with the following properties:

(1) A is closed under addition, multiplication, and scalar multiplication
(2) For every $x, y \in X$, there exists a function $f \in A$ such that $f(x) \neq f(y)$
(3) A contains a nonzero constant function

then, A is dense in the set of all continuous functions defined on X with respect to the supremum norm. Our intention in stating this theorem here is to give the reader an idea about the main result of this paper. We defer the details of this statement, including the necessary definitions, to later sections.

The remainder of this paper is structured as follows. Section 2 covers basic topology, including definitions, commonly known properties, some set theory, and a few standard results in point-set topology. Section 3 elaborates on functions defined on Euclidean spaces and functions defined on more generic spaces. It also establishes several fundamental relationships between the concepts introduced in section 2. Section 4 provides two contrasting proofs of the Weierstrass approximation theorem. Section 5 introduces and gives a proof of the Stone-Weierstrass theorem. Section 6 examines the applications of the Stone-Weierstrass theorem and concludes by introducing and proving the complex version of the Stone–Weierstrass theorem.

2. Topological background

We really ought to begin by defining the most general class of topological spaces with which we will be working, namely Hausdorff spaces; but instead, we begin by covering elementary concepts related to metric spaces because the ideas used in the proof of the Stone-Weierstrass theorem for Hausdorff spaces are nearly identical to those used in the proof of the Stone-Weierstrass theorem for metric spaces. Accordingly, we postpone our discussion of Hausdorff spaces until right before we commence the proof of the Stone-Weierstrass theorem. The reader may notice that not all the results proven in this section are restricted to metric spaces. This is intentional, and its purpose will become clear in section 6. We start with definitions.

Definition 2.1. A metric space X is a set with a function d known as the metric satisfying the following metric axioms

(1) $d(x, y) \geq 0$ and $d(x, y) = 0 \iff x = y$
(2) $d(x, y) = d(y, x)$

(3) $d(x,y) \leq d(x,z) + d(z,y)$ for $x, y, z \in X$.

Definition 2.2. Let $a \in X$. An open ball of a in metric space X is the set of all points b such that $d(a,b) < r$ where $r \geq 0$ is the radius of the open ball.

Definition 2.3. A set A in metric space X is open if for any $x \in A$, there exists a open ball N of x such that $N \subseteq A$.

Definition 2.4. A set A in metric space X is closed if its complement in X is open.

Note that the intersection of any finite collection of open sets is open, and the union of a finite or infinite collection of open sets is open. Similarly, for closed sets, the union of any finite collection of closed sets is closed, and the intersection of any finite or infinite collection of closed sets is closed. The latter statement about closed sets is implied by the former statement about open sets.

We write A^c for the complement of a set A in a metric space X, and whenever we refer to the function d, we mean the distance function of the metric space in which we operate.

Definition 2.5. A limit point a of set A in metric space X is a point such that every open set containing a also contains a $b \neq a$ with $b \in A$.

Definition 2.6. a is a limit point of set A in topological space X if every open set containing a also contains $b \neq a$ with $b \in A$.

The following proposition gives a new definition for a closed set.

Proposition 2.7. *Every limit point of a closed set A lies inside the set A.*

Proof. Let X be a topological space, and let A be a closed subset of X. Suppose that there exists a limit point $x \in X$ of $A \subseteq X$ outside of A. Then, $x \in X \backslash A = A'$ where A' is an open set by definition. By the definition of an open set, there exists an open ball N of x such that $N \subseteq A'$. Hence, $N \nsubseteq A$. This violates the definition of a limit point, as every open ball is an open set. ∎

Next, we look at special classes of metric spaces, namely compact spaces and bounded spaces. A cornerstone result in point-set topology, known as the Heine-Borel theorem, establishes a relationship between boundedness, closedness, and compactness of a Euclidean space.

Definition 2.8. A subset A of metric space X is bounded if there exists a M such that for any $x, y \in A$, $d(x,y) \leq M$.

Boundedness essentially guaranties that sequences formed in the space do not have values shooting off to infinity. Next, we turn our attention to compactness, an indispensable topological hypothesis of the Stone-Weierstrass theorem. The most common definition of compactness is in terms of open covers.

Definition 2.9. An open cover of a set A in topological space X is a collection of open sets in X such that A is a subset of the union of that collection.

Definition 2.10. A subset E of topological space X is compact if for every open cover of E, there exists a finite sub-cover of E.

Intuitively speaking, compactness ensures that the topological space can be treated as a finite closed space. This is what leads us to the next proposition.

Proposition 2.11. *Every compact subset of a metric space is closed and bounded.*

Proof. Let E be a compact subset of metric space X. Let $p \in E^c$. For each $q \in E$, let V_q and W_q be open balls of p and q with radius $r < \frac{1}{2}d(p,q)$. The union of all W_q's is an open covering of E, so since E is compact, there exists a finite collection of points $q_1, q_2, q_3, \dots$ such that

$$E \subseteq F = W_{q_1} \cup W_{q_2} \cup \dots W_{q_n}$$

of E. Let $F' = V_{q_1} \cap V_{q_2} \cap \dots V_{q_n}$. Then, F' is an open ball of p such that $F' \cap F = \emptyset$ as all W_q and V_q are disjoint. Thus, $F' \subseteq E^c$. By definition, E^c is open, and so by definition E is closed. Consider $\epsilon > 0$. For each $p \in E$, let N_p be the open ball of p with radius ϵ. Then, as the N_p's form an open cover of E, there exists a finite collection of points $p_1, p_2, p_3, \dots$ such that

$$E \subseteq N = N_{p_1} \cup N_{p_2} \cup \dots N_{p_n}.$$

Because each N_{p_i} is bounded, N is bounded, so E is bounded. ∎

The converse of proposition 2.11 generally does not hold, but an important sub-case of the converse is the Heine-Borel theorem, which is only applicable to $\mathbb{R}^m$. The following lemma will help us prove the Heine-Borel theorem.

Lemma 2.12. *A closed subset of a compact space is compact.*

Proof. Let A be a closed subset of the compact space E. Let V be an open cover of A. Adjoining A^c to V, we obtain an open cover $W = A^c \cup V$ of E. As E is compact, there is a finite sub-cover of W, say W', that covers E. This finite sub-cover W' also covers A. Let $V' = W' \backslash A^c$. Then, V' is also an open cover of A that is specifically a finite sub-cover of V. ∎

We can now prove the Heine-Borel theorem. For the proof below, we will use the fact that every Cauchy sequence in $\mathbb{R}^m$ is convergent. The proof of this fact can be found in [Rud76, Chapter 3].

Theorem 2.13 (Heine-Borel)**.** *A subset of $\mathbb{R}^m$ is closed and bounded if and only if it is compact.*

Proof. Suppose E is a compact subset of X. Then E is closed and bounded by Proposition 2.11. Conversely, suppose that E is closed and bounded. Then, E lies in a n-box

$$T_0 = [-a, a]^n$$

where $a > 0$. We shall show that T_0 is compact through proof by contradiction.

Assume that T_0 is not compact. Then, there is an open cover C of T_0 that admits no finite sub-cover. Divide each side of T_0 in half, partitioning T_0 into 2^n closed sub-boxes, each with one-half the diameter of T_0. If every sub-box had a finite sub-cover of C, their union would be a finite sub-cover of T_0. Thus, at least one sub-box has no finite sub-cover. Let this sub-box be T_1. Continuing in the same manner, we get a decreasing sequence of nested n-boxes.

$$T_0 \supseteq T_1 \supseteq T_2 \supseteq \cdots \supseteq T_k \supseteq \cdots$$

none of which have a finite sub-cover of C. The side length of each T_k is $\frac{2a}{2^k}$. We see that

$$\frac{2a}{2^k} \to 0 \text{ as } k \to \infty.$$

Define a sequence $x = x_1, x_2, \ldots$ such that each x_k is in T_k. This sequence is Cauchy because, given an $\epsilon > 0$, we can find a T_N with a side length less than ϵ that guaranties that if $n, m > N$, then $|x_n - x_m| < \epsilon$ as $x_n, x_m \in T_N$. Hence, there exists a $L = \lim_{i \to \infty} x_i$ as $\mathbb{R}^m$ is complete. Consider an arbitrary $m \in \mathbb{N}$. For this m, $x_k \in T_m$ for all $k \geq m$. Since each T_k is closed, $L \in T_m$. Hence, $L \in \bigcap_{k=0}^{\infty} T_k$ as m was arbitrary. Because C is an open cover of T_0, there exists a set $U \in C$ such that $L \in U$. Since U is open, there exists $r > 0$ such that $B \subseteq U$ where B is the open ball centered at L with radius r. As the diameters of the boxes T_k converge to 0 and each box contains L, there exists a N such that

$$T_N \subseteq B.$$

Hence,

$$T_N \subseteq U.$$

Therefore, the single set U covers T_N. As this contradicts the construction of T_N, T_0 must be compact. Hence, By lemma 2.12, E is compact.

■

Definition 2.14. The closure of a set A in topological space X is the intersection of all the closed sets containing A

We write $\bar{A}$ for the closure of a set A. A trivial property of closure is that for any subset A of a space X,

$$\bar{\bar{A}} = \bar{A}$$

because here we are essentially finding the smallest closed set containing a closed set. Closure is also closely related to density, which we will discuss in the next section.

3. FUNCTIONS

Definition 3.1. Let sets A and B belong to metric space X. Then, A is dense in B if for every $b \in B$ and every $\epsilon > 0$, there is a $a \in A$ such that $d(a, b) < \epsilon$.

An example of this is the set of all rationals $\mathbb{Q}$ being dense in the set of all reals $\mathbb{R}$.

$$A \text{ is dense in } B \iff \bar{A} = B \tag{3.1}$$

This is because every $b \in B$ is a limit point of A, since the open ball centered at an arbitrary $b \in B$ with a radius of $\epsilon > 0$ contains a $a \in A$ for any $\epsilon > 0$, by the definition of density. The new condition highlighted in 3.1 is a better condition for density because it does not invoke the metric. Before going further, we recall what it means for a function to be continuous. After all, our purpose here is to explore how well we can approximate continuous functions with various different kinds of sets. A function f mapping the metric space M to the metric space N is said to be continuous if and only if, for every $\epsilon > 0$ and every $x \in M$, there exists a $\delta > 0$ such that

$$d(x, y) < \delta \Rightarrow d(f(x), f(y)) < \epsilon$$

for $y \in M$. This is known as the $\epsilon - \delta$ condition for continuity. This condition is only valid for continuous functions defined on metric spaces, and as we want to approximate continuous functions defined on non-metric spaces as well as those defined on metric spaces, we need a new condition for continuity. This new condition uses pre-images along with the open set. For a function $f : M \to N$, the pre-image of a set $V \subseteq N$ is

$$f^{\text{pre}}(V) = \{p \in M : f(p) \in V\}.$$

The following is the general topological definition of continuity.

Definition 3.2. A function $f : M \to N$ is continuous if the pre-image of each open subset of N is a open subset of M.

We now show that this condition for continuity is equivalent to the previous condition for continuity when M and N are metric spaces. This is what theorem 0.17 is about.

Theorem 3.3. *Let $f : M \to N$. The following statements are equivalent.*

(1) $\forall \epsilon$ *and* $\forall x \in M$, $\exists \delta$ *such that*

$$y \in M \text{ and } |x - y| < \delta \Rightarrow |f(x) - f(y)| < \epsilon$$

when M and N are equipped with a metric.

(2) *The pre-image for f of an open set in N is an open set in M.*

Proof. Let M and N be metric spaces, and let f be a function mapping M to N. Suppose that f satisfies the $\epsilon - \delta$ condition for continuity. Let $U \subseteq N$ be open. Consider an arbitrary point $x \in f^{\text{pre}}(U)$. Then, $f(x) \in U$. Since U is open, there is an $\epsilon > 0$ such that $B_N \subseteq U$, where B_N is an open ball centered at $f(x)$ with radius ϵ. By the $\epsilon - \delta$ condition for continuity at x, there exists a $\delta > 0$ such that if $d_M(x, y) < \delta$, then $d_N(f(x), f(y)) < \epsilon$. Let B_M be an open ball centered at x with radius δ. Then,

$$y \in B_M \Rightarrow f(y) \in B_N.$$

So,

$$B_M \subseteq f^{\text{pre}}(U).$$

Since x was arbitrary, every point $x \in f^{\text{pre}}(U)$ has an open ball contained in $f^{\text{pre}}(U)$ centered at x. Thus, $f^{\text{pre}}(U)$ is open. Hence, f satisfies the general topological condition for continuity. Suppose that f satisfies the general topological condition for continuity. Then, the pre-image of every open set in N is open in M. Consider an $x \in M$ and an $\epsilon > 0$.

Consider the open ball U centered at $f(x)$ with radius ϵ. Since U is open and f is continuous, $f^{\text{pre}}(U)$ is open. Because $f(x) \in U$, $x \in f^{\text{pre}}(U)$. Since $f^{\text{pre}}(U)$ is open, there exists a $\delta > 0$ such that

$$B_M \subseteq f^{\text{pre}}(U)$$

, where B_M is an open ball centered at x with radius δ. If $d(x,y) < \delta$, then $y \in f^{\text{pre}}(U)$, which implies that $f(y) \in U$. Since U is equal to the ball centered at $f(x)$ with radius ϵ,

$$d(x,y) < \delta \Rightarrow d(f(x), f(y)) < \epsilon.$$

As $x \in M$ and $\epsilon > 0$ were arbitrary, there exists such a δ for any $x \in X$ and $\epsilon > 0$. Thus, f satisfies the $\epsilon - \delta$ condition for continuity. Hence, we can conclude that the ϵ, δ definition of continuity is equivalent to the general topological definition of continuity. ■

Another strong result in point-set topology, called the continuous image theorem, combines continuity with compactness.

Theorem 3.4 (Continuous image theorem)**.** *Let $f : X \to Y$ be a continuous function. If $A \subseteq X$ is compact, then $f(A) \subseteq Y$ is also compact.*

Proof. Let $f : X \to Y$ be a continuous function and let A be a compact subset of X. Let E be a subset of Y such that $\{U_\alpha : \alpha \in E\}$ is an open cover of $f(A)$. Then, each U_α is an open set in Y such that the union of all U_α for $\alpha \in E$ is a super-set of $f(A)$. Since f is continuous, theorem 3.3 guaranties that $f^{\text{pre}}(U_\alpha)$ is an open set in X for each $\alpha \in E$. We will show that $\{f^{\text{pre}}(U_\alpha) : \alpha \in E\}$ is an open cover of A. We do this by considering an arbitrary $x \in A$. It is clear that $f(x) \in f(A)$, and since

$$f(x) \in f(A) \subseteq \bigcup_{\alpha \in E} U_\alpha,$$

$f(x) \in U_\beta$ for some $\beta \in E$. By the definition of pre-image, x is contained in the open set $f^{\text{pre}}(U_\beta)$. As x is an arbitrary point in A, every point in A is contained in some $f^{\text{pre}}(U_\alpha)$ for $\alpha \in E$. Thus, $A \subseteq \bigcup_{\alpha \in E} f^{\text{pre}}(U_\alpha)$. The compactness of A asserts that there exist finitely many points $\alpha_1, \alpha_2, \ldots, \alpha_n \in E$ such that

$$A \subseteq \bigcup_{i=1}^{n} f^{\text{pre}}(U_{\alpha_i}).$$

It is left for us to show that the collection of open sets $U_{\alpha_1}, U_{\alpha_2}, \ldots, U_{\alpha_n}$ forms a finite open cover of $f(A)$ to conclude that $f(A)$ is compact. We show this by considering an arbitrary $y \in f(A)$. It is clear that $y = f(x)$ for some $x \in A$. Since $A \subseteq \bigcup_{i=1}^{n} f^{\text{pre}}(U_{\alpha_i})$, there exists a i, $1 \leq i \leq n$ such that $x \in f^{\text{pre}}(U_{\alpha_i})$. Hence, $y = f(x) \in U_{\alpha_i}$. Since the y in $f(A)$ we considered was arbitrary,

$$f(A) \subseteq \bigcup_{i=1}^{n} U_{\alpha_i}.$$

Therefore, $f(A)$ is compact. ■

An alternate shorter proof of the Continuous image theorem can be found in [Pug15, Chapter 2]. A much stronger condition than continuity is uniform continuity. We define it as follows.

Definition 3.5. Let f be a function mapping metric space M to metric space N. f is uniformly continuous if for every $\epsilon > 0$, there exists a $\delta > 0$ such that

$$d(x,y) < \delta \text{ with } x, y \in M \Rightarrow d(f(x), f(y)) < \epsilon$$

Note that the above definition only works for functions defined on metric spaces. This will not be a problem for us because we will only be using uniform continuity when discussing the Weierstrass approximation theorem. Continuity on a compact space implies uniform continuity. This is what brings us to the second theorem. The following theorem is known as the Heine-Cantor theorem.

Theorem 3.6 (Heine-Cantor)**.** *Every continuous function f defined on a compact subset of $\mathbb{R}^m$ is uniformly continuous.*

Proof. Let E be a compact subset of $\mathbb{R}^m$. Let f be a continuous function defined on E. Suppose that f is not uniformly continuous. Then, there exists an $\epsilon > 0$ such that for any $\delta > 0$, there exist points $p, q \in E$ such that

$$d(p, q) < \delta \text{ but } d(f(p), f(q)) \geq \epsilon.$$

Let $\delta_n = \frac{1}{n}$. For each δ_n, let p_n and q_n be points in E such that $d(p_n, q_n) < \delta_n$ and $d(f(p_n), f(q_n)) \geq \epsilon$. Then, $q_1, q_2, q_3, \ldots$ and $p_1, p_2, p_3, \ldots$ are sequences in E. By proposition 2.11, E is bounded, so $q_1, q_2, q_3, \ldots$ and $p_1, p_2, p_3, \ldots$ are bounded in $\mathbb{R}$. By the Bolzano-Weierstrass theorem, there exist subsequences $p_{s_1}, p_{s_2}, p_{s_3}, \ldots$ and $q_{s_1}, q_{s_2}, q_{s_3}, \ldots$ of $p_1, p_2, p_3, \ldots$ and $q_1, q_2, q_3, \ldots$ that converge to some $p \in E$ and some $q \in E$. Since

$$d(p_n, q_n) \to 0 \text{ as } n \to \infty,$$

$\lim_{n\to\infty} p_{s_n} = \lim_{n\to\infty} q_{s_n} \Rightarrow p = q$. As f is continuous,

$$\lim_{n\to\infty} f(p_{s_n}) = f(p) = f(q) = \lim_{n\to\infty} f(q_{s_n}).$$

Hence, there exists a $N_1 > 0$ such that if $n > N_1$, then

$$|f(p_{s_n}) - f(p)| < \frac{\epsilon}{2},$$

and there exists a N_2 such that if $n > N_2$, then

$$|f(p) - f(q_{s_n})| < \frac{\epsilon}{2}.$$

If $n > N = \max(N_1, N_2)$, then

$$|f(p_{s_n}) - f(q_{s_n})| \leq |f(p_{s_n}) - f(p)| + |f(p) - f(q_{s_n})| < \epsilon,$$

contrary to our assumption that $|f(p_n) - f(q_n)| \geq \epsilon$ for all n. Hence, f is uniformly continuous. ∎

A different proof of the Heine-Cantor theorem that relies on sequential compactness can be found in [Pug15, Chapter 2].

The Weierstrass approximation theorem sheds light on the importance of this theorem. The Weierstrass approximation theorem states the following:

Theorem 3.7. *Let $f : [a, b] \to \mathbb{R}$ be a continuous function with $a, b \in \mathbb{R}$. Then, f can be uniformly approximated by polynomials as closely as desired.*

One of the conditions of the Weierstrass approximation theorem, as can be seen above, is that the continuous function f is defined on a closed and bounded interval of $\mathbb{R}$. The Heine-Borel theorem (theorem 2.13) implies that this interval is compact. By theorem 3.6, commonly known as the Heine-Cantor theorem, any continuous function defined on that interval is uniformly continuous, and hence f must be uniformly continuous. So, we are basically making f uniformly continuous without explicitly saying so. This property will be greatly exploited in the proof of the Weierstrass approximation theorem. For us to exploit this property, it would be convenient for us to state it formally, which leads us to the following corollary.

Note that the above statement of the Weierstrass approximation theorem uses a uniform approximation that is tied to the notion of uniform convergence. We will look at uniform convergence shortly.

Corollary 3.8. *Any continuous function defined on a closed and bounded interval is uniformly continuous.*

Proof. Follows from the Heine-Borel theorem and the Heine-Cantor theorem. ■

Finally, we look at uniform convergence and its relation to completeness and the supremum norm.

Definition 3.9. A sequence of functions $f_1, f_2, f_3, \dots$ converges uniformly on $E \subseteq \mathbb{R}$ to a function f if for each $\epsilon > 0$, there exists an $N > 0$ such that

$$n \geq N \Rightarrow |f_n(x) - f(x)| < \epsilon$$

for all $x \in E$. The uniform convergence of a sequence of functions $f_1, f_2, f_3, \dots$ to a function f is written as $f_n \rightrightarrows f$. If f is defined on E, each f_n is defined on E, and $f_n \rightrightarrows f$ as $n \to \infty$, we say that the set of $f_n's$ uniformly approximates f on the domain E. That is, uniform convergence implies uniform approximation, and vice versa.

Definition 3.10. A function f can be uniformly approximated by elements of a set B over set E if there exists a sequence of functions $f_1, f_2, \dots \in B$ such that they uniformly converge on E to f.

We will often prove uniform convergence in statements of theorems that require us to show uniform approximation. Consequently, if A is dense in B, then every element of B can be uniformly approximated by elements of A. We also incorporate closure in this. If f can be uniformly approximated by elements of a set B, then there exist $f_1, f_2, \dots \in B$ such that $f_n \rightrightarrows f$ as $n \to \infty$, so f is a limit point of B, so $f \in \bar{B}$.

Uniform convergence is most relevant in **function space**, a set with functions as its elements. We write C^0 for the class of all continuous bounded functions, and we write $C^0(X, \mathbb{R})$ for the class of all continuous bounded real-valued functions defined on X. We define the **supremum norm** of a function $f \in C^0(X, \mathbb{R})$ as $||f|| = \sup_{x \in X} |f(x)|$. The supremum norm defines a metric. Hence, the supremum norm satisfies the metric axioms discussed at the beginning of the second section. It is important to note that

$$n \to \infty \Rightarrow ||f_n - f|| \to 0$$

is equivalent to $f_n \rightrightarrows f$.

Definition 3.11. A topological space X is complete if every Cauchy sequence in X converges to an element in X.

Theorem 3.12. *For any bounded set X, $C^0(X, \mathbb{R})$ is complete with respect to the supremum norm.*

Proof. Let X be a closed and bounded set. Let $f_1, f_2, f_3, \ldots$ be a sequence of functions in $C^0(X, \mathbb{R})$ that uniformly converges to a function f. If we manage to show that f is continuous, then we can conclude that $C^0(X, \mathbb{R})$ is complete, since every convergent sequence is Cauchy in any set, and $f \in X$ because X contains all its limit points according to proposition 2.7. Let an $\epsilon > 0$ be given. Consider the supremum norm on $C^0(X, \mathbb{R})$. Since $f_n \rightrightarrows f$, there is an $N > 0$ such that if $n > N$, then

$$||f_n(x) - f(x)|| < \frac{\epsilon}{3}. \tag{3.2}$$

Since f_{N+1} is continuous throughout X, there is $\delta > 0$ such that

$$|x - y| < \delta \Rightarrow |f_{N+1}(x) - f_{N+1}(y)| < \frac{\epsilon}{3}$$

for $x, y \in X$. Hence, if $|x - y| < \delta$ for $x, y \in X$, then

$$\begin{aligned}|f(x) - f(y)| &\leq |f(x) - f_{N+1}(x)| + |f_{N+1}(x) - f_{N+1}(y)| + |f_{N+1}(y) - f_{N+1}(y)| \\ &\overset{(a)}{\leq} \frac{\epsilon}{3} + \frac{\epsilon}{3} + \frac{\epsilon}{3} \\ &= \epsilon\end{aligned}$$

where (a) is due to equation 3.2. Hence, $C^0(X, \mathbb{R})$ is complete. ∎

4. Approximation by Polynomials

The reason we want to approximate continuous functions defined on an interval by polynomials, specifically out of all the various classes of functions, is that polynomials are undoubtedly nicer than many other continuous functions. Polynomials are smooth, meaning that they are infinitely differentiable and that every derivative of a polynomial is continuous. We shall now look at the central result in elementary approximation theory, which is not so surprising, the most important case of the Stone-Weierstrass theorem. We will now prove the Weierstrass approximation theorem.

Theorem 4.1 (Weierstrass approximation theorem)**.** *The set of all polynomials is dense in $C^0([a, b], \mathbb{R})$ for $a, b \in \mathbb{R}$.*

The statement of theorem 3.7 is equivalent to the statement of theorem 4.1, since here density implies that for each $f \in C^0$ and each $\epsilon > 0$, there is a polynomial function $p(x)$ such that for all $x \in [a, b]$, $|f(x) - p(x)| < \epsilon$.

We will provide two proofs of the Weierstrass approximation theorem, both of which are adapted from [Pug15]. The first proof, which shall be provided in detail, is by far the most elementary proof of the Weierstrass approximation theorem and, accordingly, the most beautiful. It invokes absolutely zero calculus machinery, instead relying on Bernstein polynomials to approximate $f \in C^0([a,b],\mathbb{R})$. The second proof is dry and terse in comparison with the first, yet still indispensable due to its use of the convolution method. We will give only a rough sketch of the second proof. In both proofs, we assume that $[a,b]=[0,1]$. The generalization of this theorem from the case of $[a,b]=[0,1]$ to the case where $[a,b]$ is an arbitrary closed interval is rather straightforward and is shown in the appendix.

4.1. **Bernstein's proof.** This proof is due to the work of Sergei Noviciate Bernstein. It uses two lemmas. For each $n \in \mathbb{N}$, consider the sum

$$p_n(x) = \sum_{k=0}^{n} \binom{n}{k} c_k x^k (1-x)^{n-k}. \tag{4.1}$$

where $c_k = f\left(\frac{k}{n}\right)$. It is obvious that p_n is a polynomial, but more specifically, p_n is the n^{th} Bernstein polynomial of f. Let $r_k(x) = \binom{n}{k} x^k (1-x)^{n-k}$. Our two lemmas will be about the properties of r_k.

Lemma 4.2. $\sum_{k=0}^{n} r_k(x) = 1$

Proof. By the binomial theorem,

$$\sum_{k=0}^{n} r_k(x) = \sum_{k=0}^{n} \binom{n}{k} x^k (1-x)^{n-k} = (x+(1-x))^n = 1.$$

■

Lemma 4.3. $\sum_{k=0}^{n} (k-nx)^2 r_k(x) = nx(1-x)$

Proof. Substituting $r_k(x) = \binom{n}{k} x^k (1-x)^{n-k}$ and expanding the square, we get that

$$\sum_{k=0}^{n} (k-nx)^2 r_k(x) = \sum_{k=0}^{n} k^2 r_k(x) - 2nx \sum_{k=0}^{n} k r_k(x) + n^2x^2 \sum_{k=0}^{n} r_k(x).$$

where the three sums on the right follow from the binomial theorem. By lemma 4.2,

$$\sum_{k=0}^{n} r_k(x) = 1 \tag{4.2}$$

Consider the binomial identity

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}.$$

Setting $y = 1-x$, then differentiating the above equation with respect to x and multiplying by x, we obtain

$$\sum_{k=0}^{n} k r_k(x) = nx. \tag{4.3}$$

Shifting our gaze to the second sum , we observe that $k^2 = k(k-1)+k$.

Once again, we differentiate $(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}$, but this time we do it twice, and then we multiply by x^2 and again set $y = 1-x$, which gives us

$$\sum_{k=0}^{n} k(k-1) r_k(x) = n(n-1)x^2.$$

Using $k^2 = k(k-1)+k$, we achieve the following identity.

$$\sum_{k=0}^{n} k^2 r_k(x) = n(n-1)x^2 + nx. \tag{4.4}$$

Substituting equations 4.2, 4.3, and 4.4 into our original equation leads us to

$$\sum_{k=0}^{n} (k-nx)^2 r_k(x) = \big(n(n-1)x^2 + nx\big) - 2n^2x^2 + n^2x^2 = -nx^2 + nx = nx(1-x).$$

Hence, proved. ■

Finally, we are in a position to prove the Weierstrass approximation theorem for the case of $[a,b] = [0,1]$.

Proof of the Weierstrass approximation theorem. Let $f : [0,1] \to \mathbb{R}$ be a continuous function. We will show that

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

where p_n for any $n \in \mathbb{N}$ is defined as in equation 4.1. We first write both p_n and f in terms of r_k. It is easy to see that $p_n = \sum_{k=0}^{n} c_k r_k(x)$. Since $\sum_{k=0}^{n} r_k(x) = 1$ by the lemma 4.2, $f(x) = \sum_{k=0}^{n} f(x) r_k(x)$. We divide the sum $p_n(x) - f(x) = \sum_{k=0}^{n} (c_k - f(x))\, r_k(x)$ into terms for which $\frac{k}{n}$ is close to x and terms for which $\frac{k}{n}$ is far from x. That is, for a chosen $\epsilon > 0$, we find a $\delta > 0$ such that

$$|t-s| < \delta \Rightarrow |f(t) - f(s)| < \frac{\epsilon}{2} \tag{4.5}$$

for all $t, s \in [0,1]$. It is always possible to find such a $\delta > 0$ because f is uniformly continuous on $[0,1]$ by corollary 3.8. Figure 1 illustrates the Bernstein polynomial approximation of $f(x) = 2^x e^{-x^2} + 3^x x^5 - x^4$. In the figure, the function being approximated is in black, the tenth Bernstein polynomial of that function is in blue, and the fifteenth Bernstein polynomial of that function is in purple.

We now put this δ to good use by letting $K_1 = \left\{[n] : \left|\frac{k}{n} - x\right| < \delta\right\}$ and $K_2 = [n] \backslash K_1$. By $[n]$ we mean $\{0, 1, 2, \ldots, n\}$. This tells us the following:

$$\begin{aligned} |p_n(x) - f(x)| &\leq \sum_{k=0}^{n} |c_k - f(x)| r_k(x) \\ &= \sum_{k\in K_1} |c_k - f(x)| r_k(x) + \sum_{k \in K_2} |c_k - f(x)| r_k(x) \end{aligned} \tag{4.6}$$

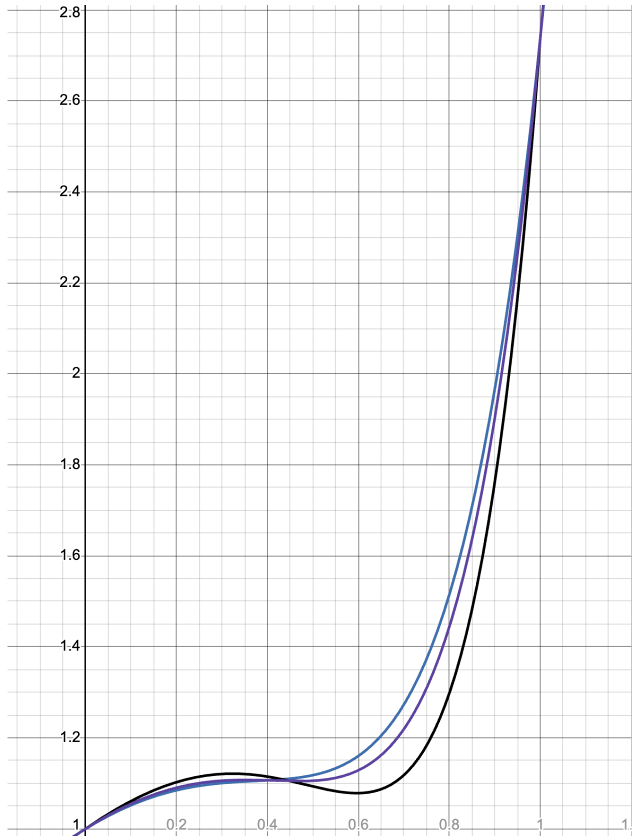

Figure 1. Approximating $2^x e^{-x^2} + 3^x x^5 - x^4$

In the above equation, all the terms where $\frac{k}{n}$ is close to x are in the first sum, and all the terms where $\frac{k}{n}$ is far from x are in the second sum, allowing us to deal with them separately. Using the property highlighted in equation 4.5, we see that

$$\sum_{k \in K_1} |c_k - f(x)| r_k(x) < \frac{\epsilon}{2} \sum_{k \in K_1} r_k(x)$$

because $c_k = f\left(\frac{k}{n}\right)$ and for every $k \in K_1$, $|\frac{k}{n} - x| < \delta$. By lemma 4.2 and by the definition of K_1,

$$\sum_{k \in K_1} |c_k - f(x)| r_k(x) < \frac{\epsilon}{2} \sum_{k \in K_1} r_k(x) \leq \frac{\epsilon}{2} \sum_{k=0}^{n} r_k(x) = \frac{\epsilon}{2} \tag{4.7}$$

as all $r_k(x)'s$ are nonnegative since the domain of p_n for any n is restricted to $[0, 1]$. We have successfully shown that the first sum in equation 4.6 is bounded by $\frac{\epsilon}{2}$. This leaves us with the second sum in equation 4.6. For the second sum, we use lemma 4.3 to write

$$nx(1-x) = \sum_{k=0}^{n} (k - nx)^2 r_k(x) \geq \sum_{k \in K_2} (k - nx)^2_k(x) \geq \sum_{k \in K_2} (n\delta)^2 r_k(x)$$

because $|\frac{k}{n} - x| \geq \delta$ for every $k \in K_2$. Notice that $\max(x(1-x)) = \frac{1}{4}$ for $x \in [0, 1]$. This is because, by the AM-GM inequality, $\sqrt{x(1-x)} \leq \frac{x+1-x}{2} \Rightarrow x(1-x) \leq \frac{1}{4}$ whenever

$x(1-x) \geq 0$, which occurs only when $x \in [0,1]$. Thus,

$$\sum_{k \in K_2} r_k(x) \leq \frac{nx(1-x)}{(n\delta)^2} \leq \frac{1}{4n\delta^2}.$$

By the Extreme value theorem [RS23, Chapter 31], f attains its maximum and minimum on $[a,b]$, say m_1 and m_2, as $[a,b]$ is a closed and bounded interval. Let $M = \max(|m_1|, |m_2|)$. Then,

$$\sum_{k \in K_2} |c_k - f(x)| \leq 2M.$$

This implies that there exists a large enough n such that

$$\sum_{k \in K_2} |c_k - f(x)| r_k(x) \leq \frac{M}{2n\delta^2} \leq \frac{\epsilon}{2} \tag{4.8}$$

Substituting equations 4.8 and 4.7 into 4.6, we obtain $|p_n(x) - f(x)| < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon$, which completes the proof. ■

Bernstein's proof is surprising because, at first glance, it is not clear why the Bernstein polynomials of f should approach f. It is only later, when we examine Bernstein polynomials, which are heavily influenced by binomial coefficients, that we see their underlying nature. The simple idea of splitting the main sum into far away terms and close terms, and exploiting each ones individual properties is what makes this proof so intuitive. It's elegance comes from it's combinatorial origins.

4.2. **Proof using the convolution method.** This proof uses Riemann integrals. Please note that here we only provide a sketch of the proof. For a detailed proof using the convolution method, see [Pug15, Chapter 4] or [Rud76, Chapter 7].

Proof of the Weierstrass approximation theorem. Let $f : [0,1] \to \mathbb{R}$ be continuous. We claim that there is no loss of generality in assuming that $f(0) = f(1) = 0$. The justification for our claim lies in the construction of a function g such that

$$g(x) = f(x) - f(0) - x[f(1) - f(0)].$$

From the above construction, we can easily deduce that $g \in C^0$ and that $g(0) = g(1) = 0$. $g \in C^0$ because continuity is preserved under the four elementary operations of arithmetic. Hence, if the statement of this theorem holds for g, then it also holds for f because $f(0) + x[f(1) - f(0)]$ is a linear polynomial.

Extend the domain of f to all of $\mathbb{R}$ by setting $f(x) = 0$ whenever $x \notin [0,1]$. This way, we can say that f is continuous on all of $\mathbb{R}$. Consider the function

$$\beta_n = b_n \left(1 - t^2\right)^n$$

for all $t \in [-1,1]$. Here, the constant b_n satisfies $\int_{-1}^{1} \beta_n \, dt = \int_{-1}^{1} b_n(1-t^2)^n = 1$. β_n is obviously a polynomial. Figure 3 offers a deeper look at the nature of β_n, particularly its bump. Let $P_n = \int_{-1}^{1} f(x+t)\beta_n \, dt$. By construction, P_n is the weighted average of the values of f using the weight function β_n.

In the complete proof, two things are shown. The first being that each P_n is indeed a polynomial, and the second being that

$$n \to \infty \Rightarrow P_n(x) \rightrightarrows f(x)$$

for all $x \in [0, 1]$.

It is shown that each P_n is a polynomial by writing $P_n(x)$ as $\int_0^1 f(u)\beta_n(x-u)\,du$ through a change of variables, and then inspecting its polynomial nature. Using integral manipulations, we achieve the following result: there exists a constant c such that $b_n \leq c\sqrt{n}$ for all n. We use this to conclude that $\beta_n(t) \leq c\sqrt{n}(1-\delta^2)^n \to 0$ as $n \to \infty$. It follows that $\beta_n \rightrightarrows 0$ as $n \to \infty$. Uniform continuity of f, which is implied by corollary 3.8, along with the fact that $\beta_n \rightrightarrows 0$ as $n \to \infty$ implies that $n \to \infty \Rightarrow P_n \rightrightarrows f$ as b_n was chosen such that $\int_{-1}^1 b_n(1-t^2)^n = 1$. This completes the rough sketch. ■

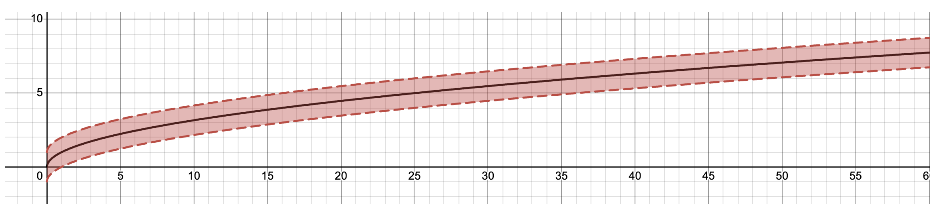

Figure 2. ϵ-tube

Graphically, what we tried to do in both proofs is construct a function that is within the ϵ-tube of our given function. In the first figure below, the epsilon tube of the function $\sqrt{x}$, whose graph is in black, is the red area around it.

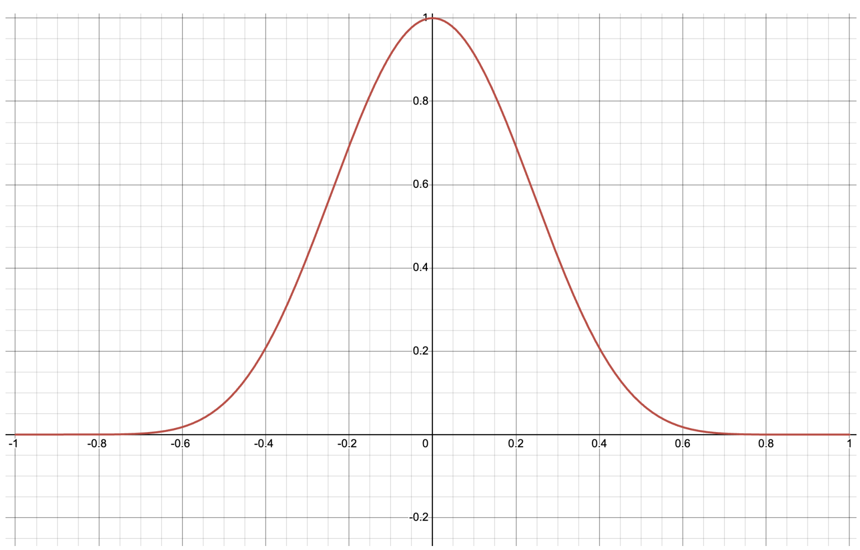

Figure 3. A rough sketch of β_n

5. Stone-Weierstrass Theorem (Real version)

For several years, it was thought that the Weierstrass approximation theorem could be generalized no further. The main reason for this was that very few classes of functions were as nice and as easy to compute as polynomials, so it was assumed that polynomials were unique in having the property that any continuous function f defined on a closed interval could be uniformly approximated as closely as desired by polynomial functions. We now know that this is not true. A counterexample is the trigonometric approximation theorem. This theorem asserts that the set of all trigonometric polynomials is dense in the set of all continuous 2π-periodic functions. We will give a proof of the trigonometric approximation theorem after giving a proof of the real Stone-Weierstrass theorem. To see why the property implied by the Weierstrass approximation theorem is not unique to the set of all polynomials and its super-sets, we pose a seemingly trivial question.

What are the defining features of the polynomials?

There are many answers to this question, but we are looking for an answer relevant to the Weierstrass approximation theorem. One could say that the smoothness of the polynomials is a defining feature, but the smoothness of the polynomials was not used in Bernstein's proof, nor is it used in any other proof of the Weierstrass approximation theorem.

Something that was used in both of the proofs we discussed, perhaps not directly, was that polynomials are closed under addition, subtraction, multiplication, and scalar multiplication. In fact, every polynomial can be built from two elementary functions, namely $g(x) = x$ and $f(x) = 1$, through addition, multiplication, and scalar multiplication. This is indeed a property of polynomials that is relevant here. We call a set with a similar property a function algebra. More formally, a function space A in the set X is a **function algebra** if it is closed under addition, multiplication, and scalar multiplication, and if it contains a nonzero constant function. In other words, if $f, g \in A$, then $fg \in A$, $f + g \in A$, and $df, dg \in A$ for any constant d, and there is a $f \in A$ such that $f(x) = c$ for all x in the domain of f.

Another property of the set of all polynomials is that for any x, there is always a polynomial p such that $p(x) \neq 0$. Otherwise, if the function $f \in C^0([a, b], \mathbb{R})$ we are approximating is nonzero at x, and if all the functions in our set are zero at x, then the approximation fails miserably. We say that a function space A vanishes nowhere if it has a similar property. Proceeding more formally, we say that a function algebra A **vanishes at a point** x if $f(x) = 0$ for all $f \in A$.

Every function algebra vanishes nowhere since it contains a nonzero constant function by definition. This is one method to force our set A to vanish nowhere. An alternative approach is to incorporate this condition into the statement of the Stone-Weierstrass theorem, and accordingly define function algebras to be merely closed under addition, multiplication, and scalar multiplication. We will not do this as the proof that assumes our set A contains a nonzero constant function is cleaner and better aligned with the central theme of this paper. An example of this alternative approach can be found in [Pug15, Chapter 4].

A nontrivial, vital property that a function algebra must have to be dense in $C^0(X, \mathbb{R})$ is that it must separate points. This is a discovery that was first made by Marshall Stone. We say that a function algebra A **separates points** if for each pair of distinct points $p_1, p_2 \in X$, there exists a $f \in A$ such that $f(p_1) \neq f(p_2)$. In a topological context, a topological space X is said to separate points if for each pair of distinct points $p_1, p_2 \in X$, there exist disjoint open sets B_1, B_2 such that $p_1 \in B_1$ and $p_2 \in B_2$. We say that a space that separates points topologically is **Hausdorff**. A Hausdorff space is implied to have unique limit points.

We now turn our attention to the finite interval $[a, b]$. A fact we strongly used in Bernstein's proof, specifically to assert that f is uniformly continuous, is that the finite interval $[a, b]$ is compact. We generalize the finite interval $[a, b]$ to an arbitrary compact Hausdorff space in the following statement of the real Stone-Weierstrass theorem. We could instead generalize the finite interval $[a, b]$ to an arbitrary compact metric space, but interestingly, the Stone-Weierstrass theorem for metric spaces is no easier to prove, and the proof itself never uses any kind of metric. Accordingly, we have pure topological definitions for compactness and continuity. All the results we will employ in our proof of the Stone–Weierstrass theorem have already been established earlier in this paper, and in fact have been proved at least for Hausdorff spaces.

Theorem 5.1 (Real Stone-Weierstrass theorem [Sim83]). *Let X be a compact Hausdorff space. Let A be a function algebra in $C^0(X, \mathbb{R})$ that separates points. Then, A is dense in $C^0(X, \mathbb{R})$ with respect to the supremum norm.*

We shall require two lemmas along with the Weierstrass approximation theorem and a few minor results. Lemma 5.2 and 5.4 are proven for all arbitrary compact Hausdorff spaces as we move away from the metric space concepts which were presented merely as a stepping stone to the true abstract nature of the Stone-Weierstrass theorem. These lemmas also hold for metric spaces as every metric space separates points topologically, and accordingly is Hausdorff. To see why this is true, assume that X is a metric space, and consider the open balls centered at distinct points $p_1, p_2 \in X$ with radii less than $\frac{d(p_1,p_2)}{2}$.

One could ask whether we are allowed to use open sets when operating without a metric. The answer is that, in the context of general topology, open sets are taken to be primitive, and everything else follows from them. Moreover, we define a **topological space** rigorously as a pair $(\mathcal{X}, \mathcal{T})$ where $\mathcal{X}$ is a set and $\mathcal{T}$ is a collection of subsets of $\mathcal{X}$ which are known as open sets satisfying

(1) $\emptyset, \mathcal{X} \in \mathcal{T}$
(2) arbitrary unions of members of T belong to T
(3) finite intersections of members of T belong to T.

A **closed set** is still defined as the complement of an open set, but now in a broader setting. Since every metric space induces a topology, namely the topology whose open sets are open balls defined by the metric, every metric space is naturally a topological space. From now on, we shall be operating in the general setting of topology, and accordingly, we interpret concepts such as continuity, closure, and compactness with respect to the open sets of the topology rather than simply the open sets induced by the metric.

Lemma 5.2. *If the function algebra A separates points, then for each pair of distinct points $p_1, p_2 \in X$ and each pair of real constants c_1,c_2, there exists a $f \in A$ such that $f(p_1) = c_1$ and $f(p_2) = c_2$.*

Proof. Let points p_1 and p_2 in X be given. Let real constants c_1 and c_2 also be given. Suppose that A is a function algebra that separates points. Then, by the separating points property of A, there exists a $h \in A$ such that

$$h(p_1) \neq h(p_2).$$

Construct a function g using $h(p_1)$ and $h(p_2)$ such that

$$g(x) = \frac{h(x) - h(p_2)}{h(p_1) - h(p_2)}.$$

Since A is a function algebra, $g \in A$ as g was constructed from h through adding real constants, adding $h(x)$, and multiplication by a real scalar. A function algebra such as A is closed under the addition of a constant because scalar multiplication on a nonzero constant function $f(x) = c$ can yield any constant function, ensuring that every function algebra contains all constant functions. Construct f such that

$$f(x) = c_2 + (c_1 - c_2)g(x).$$

Then, $f \in A$ as f was constructed from g through the addition of a constant and multiplication by a real scalar. Notice that $f(p_1) = c_1$ as $g(p_1) = 1$, and $f(p_2) = c_2$ as $g(p_2) = 0$. Hence, we are done. ■

The following lemma will be used to prove our second lemma. It's purpose is to illustrate the fundamental property of any function algebra.

Lemma 5.3. *If A is a function algebra in $C^0(X, \mathbb{R})$, then $\bar{A}$ is also a function algebra in $C^0(X, \mathbb{R})$.*

Proof. Let X be a compact Hausdorff space. To prove that $\bar{A}$ is a function algebra, we need to show that it is closed under addition, multiplication, and scalar multiplication, and that it contains a function f such that $f(x) = c$ for all x. First, we note that $\bar{A}$ contains all of its limit points by proposition 2.7, so every convergent sequence of functions in $A \subseteq \bar{A}$ converges to a function in $\bar{A}$. Let $f, g \in \bar{A}$. Then, there exist sequences $g_1, g_2, g_3, \ldots$ and $f_1, f_2, f_3, \ldots$ in A that converge to f and g. As $C^0(X, \mathbb{R})$ is complete with respect to the supremum norm by theorem 3.12,

$$n \to \infty \Rightarrow ||f_n - f|| \to 0 \text{ and } ||g_n - g|| \to 0. \tag{5.1}$$

Because A is a function algebra, $f_n + g_n \in A$ for all $n \in \mathbb{N}$. Thus, we see that

$$||(f_n + g_n) - (f + g)|| \leq ||f_n - f| + ||g_n - g|| \to 0$$

as $n \to \infty$, so $f_1 + g_1, f_2 + g_2, \ldots$ converges to $f + g$, so $f + g \in \bar{A}$. Consider an arbitrary constant c. Then, $cf_n \in A$ for all $n \in \mathbb{N}$, and we observe that

$$||cf_n - cf| \leq |c|||f_n - f|| \to 0$$

as $n \to \infty$ due to 5.1, so $cf \in \bar{A}$. $f_n g_n \in A$ for all $n \in \mathbb{N}$. We observe that

$$
\begin{aligned}
||f_n g_n - fg|| =& ||f_n g_n - f_n g + f_n g - fg|| \\
\leq& ||f_n|| \cdot ||g_n - g|| + ||f_n - f|| \cdot ||g||.
\end{aligned}
\tag{5.2}
$$

As $n \to \infty \Rightarrow f_n \rightrightarrows f$, $||f_n|| \leq ||f|| + ||f_n - f||$. This says that the sequence of $||f_n||'s$ is bounded, hence the terms in the right hand side of equation 5.2 tend to 0 as n tends to infinity. Thus, $fg \in \bar{A}$. A nonzero constant function g is in A, so g is in $\bar{A}$. Hence, $\bar{A}$ is a function algebra. ■

We have arrived at the second lemma needed to prove the real Stone-Weierstrass theorem.

Lemma 5.4. *Let A be a function algebra in $C^0(X, \mathbb{R})$ where X is a compact Hausdorff space. Then, $f \in \bar{A} \Rightarrow |f| \in \bar{A}$.*

Proof. Let X be a compact Hausdorff space. Let $A \subseteq C^0(X, \mathbb{R})$ be a function algebra. Consider an arbitrary $f \in \bar{A}$. Since the domain of f, namely X, is compact and f is continuous, the image of X on $\mathbb{R}$ is compact by the continuous image theorem. Therefore, the range of $f(x)$ is closed and bounded by the Heine-Borel theorem, so $f(X) \subseteq [a, b]$ for some $a, b \in \mathbb{R}$. Consider the function $h(y) = |y|$ defined solely on $[a, b]$. As $h(y)$ is continuous on $[a, b]$, we can apply the Weierstrass approximation theorem. By the Weierstrass approximation theorem, there exists a polynomial $p(x)$ such that

$$||p(y) - h(y)|| < \epsilon. \tag{5.3}$$

Since $p(y)$ is a polynomial, it is of the form $a_1 y + a_2 y^2 + \cdots + a_n y^n$ for constants a_i with $1 \leq i \leq n$. Let

$$g(x) = (p \circ f)(x) = a_1 f^2(x) + a_2 f^2(x) + \cdots + a_n f^n(x).$$

Since $f \in \bar{A}$ and $\bar{A}$ is function algebra by lemma 5.3, which accordingly contains all constant functions, $g \in \bar{A}$. Because $f(x) \in [a, b]$ for all $x \in X$, every $f(x)$ for $x \in X$ is equal to a $y \in [a, b]$. As p uniformly approximates h over $[a, b]$ due to equation 5.3 and the definition of the supremum norm, and $|f(x)|$ is equal to a $h(y)$ whenever $x \in X$,

$$x \in X \Rightarrow |g(x) - |f(x)|| = |p(y) - h(y)| < \epsilon.$$

Hence, g uniformly approximates $|f|$ over X, so $f \in \bar{\bar{A}} = \bar{A}$. ■

We are now ready to prove the real Stone-Weierstrass theorem.

Proof of the real Stone-Weierstrass theorem. Let X be a compact Hausdorff space, and let A be a function algebra in $C^0(X, \mathbb{R})$ that separates points. Consider an arbitrary $F \in C^0(X, \mathbb{R})$ and an arbitrary $\epsilon > 0$. To show that A is dense in $C^0(X, \mathbb{R})$, it suffices for us to find a $G \in \bar{A}$ such that

$$F(x) - \epsilon < G(x) < F(x) + \epsilon$$

for all $x \in X$. The reason for this is that finding such a $G \in \bar{A}$ implies that $\bar{A}$ is dense in $C^0(X, \mathbb{R})$ as the f we considered was arbitrary, so $\bar{\bar{A}} = C^0(X, \mathbb{R}) = \bar{A}$ by equivalence 3.1, but this implies that A is dense in $C^0(X, \mathbb{R})$ by equivalence 3.1.

Useful operations that we would like to perform are taking the maximum and minimum of a finite number of functions. But do we know that these operations are closed in $\bar{A}$? These operations are indeed closed in $\bar{A}$. We can see this by writing the minimum and maximum functions using absolute values.

$$\max(f,g) = \frac{f+g+|f-g|}{2}$$
$$\min(f,g) = \frac{f+g-|f-g|}{2}$$

This, along with lemma 5.4, allows us to conclude that if $f, g \in \bar{A}$, then $\max(f,g)$ and $\min(f,g)$ are in $\bar{A}$. Using this argument repeatedly shows that the maximum and minimum of any finite number of functions in $\bar{A}$ is in $\bar{A}$. Let p, q be distinct points in X. Since A is a function algebra and A separates points, lemma 5.2 says that we can find a function $H_{pq} \in A$ such that

$$H_{pq}(p) = F(p) \text{ and } H_{pq}(q) = F(q).$$

For the time being, fix p and allow q to vary. Consider an arbitrary $q \in X$ such that $q \neq p$. Let $L(x) = H_{pq}(x) - F(x) + \epsilon$. Then, $L(x)$ is a continuous function of x because $F, H_{pq} \in C^0(X, \mathbb{R})$. Since $L(x)$ is positive at $x = q$, $q \in L^{\mathrm{pre}}(0,\infty)$. The topological condition for continuity, which was established in theorem 3.3, asserts that $L^{\mathrm{pre}}(0,\infty)$ is open as $(0,\infty)$ is open. Let U_q be equal to $L^{\mathrm{pre}}(0,\infty)$. Then, U_q is an open set containing this q such that whenever $x \in U_q$,

$$0 < H_{pq}(x) - F(x) + \epsilon \Rightarrow F(x) - \epsilon < H_{pq}(x).$$

Since the q we considered was arbitrary, we know that for each $q \in X$, there exists an open set U_q containing that q such that

$$x \in U_q \Rightarrow F(x) - \epsilon < H_{pq}(x) \tag{5.4}$$

Due to the fact that $q \in U_q$ always holds, the collection of all $U_q's$ forms an open cover of X. As X is compact, there exist finitely many points $q_1, q_2, \ldots, q_n$ in X such that the finite collection of open sets $U_{q_1}, U_{q_2}, \ldots$ covers X. Let

$$G_p(x) = \max(H_{pq_1}(x), H_{pq_2}(x), \ldots, H_{pq_n}(x).$$

Then, $G_p \in \bar{A}$ because $H_{pq_n} \in \bar{A}$ for every n. This implies two things. The first being that $G_p(p) = F(p)$ since $H_{pq_n}(p) = F(p)$ for every n, and the second being that $F(x) - \epsilon < G_p(x)$ for all $x \in X$. We now let p also vary. Consider an arbitrary $p \in X$. Let $D(x) = G_p(x) - F(x) - \epsilon$. Then, $D(x)$ is continuous because $G_p, F \in C^0(X, \mathbb{R})$. As $D(x)$ is negative at $x = p$ since $G_p(p) = F(p)$, $p \in D^{\mathrm{pre}}(-\infty, 0)$. It is also true that $D^{\mathrm{pre}}(-\infty, 0)$ is open by the topological condition for continuity. Let V_p be equal to $D^{\mathrm{pre}}(-\infty, 0)$. Then, V_p is an open set containing p such that whenever $x \in V_p$,

$$G_p(x) < F(x) + \epsilon.$$

Because the p we considered was arbitrary, there exists such a V_p for each $p \in X$. Due to the fact that $p \in V_p$, the collection of all $V_p's$ forms an open cover of X. Compactness of X

implies that there exist finitely many points $p_1, p_2, \ldots, p_n$ in X such that the finite collection $V_{p_1}, V_{p_2}, \ldots, V_{p_n}$ covers X. Let

$$G(x) = \min(G_{p_1}, G_{p_2}, \ldots, G_{p_n}).$$

Then, $G \in \bar{A}$ as $G_{p_n} \in \bar{A}$ for every n. Hence,

$$F(x) - \epsilon < G(x) < F(x) + \epsilon$$

for all $x \in X$ due to the following two reasons:

$$x \in X \Rightarrow F(x) - \epsilon < G_p(x) \text{ and } x \in V_p \Rightarrow G_p(x) < F(x) + \epsilon.$$

This completes the proof of the real Stone-Weierstrass theorem. ■

6. Lattice version and Complex version

A stylistically different proof of the real Stone-Weierstrass theorem uses lattices. A **lattice** is a set that is closed under the maximum and minimum functions, that is, if L is a lattice, then

$$f, g \in L \;\Rightarrow\; \max(f,g) \in L \; and \min(f,g) \in L.$$

The proof using lattices does not quite prove the Stone-Weierstrass theorem; instead, it proves a weaker result specific to lattices and depends on some common properties of function algebras that separate points to conclude the result of the real Stone-Weierstrass theorem. This weaker result is presented below. It is called the lattice version of the Stone-Weierstrass theorem.

Theorem 6.1. *If X is a compact Hausdorff space with at least two points and L is a lattice in $C^0(X, \mathbb{R})$ with the following property: for any two distinct elements x and y of X and any two real numbers a and b, there exists a function $f \in L$ such that $f(x) = a$ and $f(y) = b$, then L is dense in $C(X, R)$.*

A proof of the lattice version of the real Stone-Weierstrass theorem can be found in [Sim83, Chapter 7]. Theorem 6.1, along with lemma 5.2 and 5.4, directly proves the real Stone-Weierstrass theorem. The real Stone-Weierstrass theorem enables us to generalize the Weierstrass approximation theorem to any finite number of dimensions.

Corollary 6.2. *If X is a closed and bounded subset of $\mathbb{R}^n$, the set of all polynomials $p(x_1, x_2, \ldots, x_n)$ in n variables defined on X is dense in $C^0(X, \mathbb{R})$.*

Proof. Let $X \subseteq \mathbb{R}^n$ be closed and bounded. The Heine-Borel theorem ensures that X is compact. Consider two distinct arbitrary polynomials $p(x_1, x_2, \ldots, x_n)$ and $q(x_1, x_2, \ldots, x_n)$ in n variables. We can write them as finite sums.

$$p(x_1, x_2, \ldots, x_n) = \sum_{i_1, \ldots, i_n = 0}^{n} x_1^{i_1} x_2^{i_2} \ldots x_n^{i_n} c_{i_1 \ldots i_n}$$

$$q(x_1, x_2, \ldots, x_n) = \sum_{i_1, \ldots, i_n = 0}^{n} x_1^{i_1} x_2^{i_2} \ldots x_n^{i_n} d_{i_1 \ldots i_n}$$

where the $c_{i_1,\ldots,i_n}$ and $d_{i_1,\ldots,i_n}$ are real constants. Hence, we see that $p(x_1, x_2, \ldots, x_n) + q(x_1, x_2, \ldots, x_n)$, $p(x_1, x_2, \ldots, x_n)q(x_1, x_2, \ldots, x_n)$, and $rq(x_1, x_2, \ldots, x_n)$ for some constant r are all finite sums with n independent variables. Furthermore, $f(x_1, x_2, \ldots, x_n) = 1$ is a polynomial in $\mathbb{R}^n$. Thus, the set of all polynomials $p(x_1, x_2, \ldots, x_n)$ in n variables forms a function algebra. Hence, by the Stone-Weierstrass theorem, the set of all polynomials in n variables is dense in $C^0(X, \mathbb{R})$. ■

Another famous consequence of the real Stone-Weierstrass theorem is one of the origins of the Fourier series. This consequence is the Trigonometric approximation theorem, for which we shall provide a proof below.

Theorem 6.3 (Trigonometric approximation theorem [DD])**.** *The set of all trigonometric polynomials is dense in the set of all continuous 2π-periodic real-valued functions.*

Proof. We will prove the trigonometric approximation theorem via the real Stone-Weierstrass theorem. The space of all 2π-periodic functions is the unit circle, which is compact, so it suffices for us to show that the set of all trigonometric polynomials forms a function algebra and separates points. We show the former first. Consider the trigonometric polynomial function $f(x) = \sin^2(x)+\cos^2(x)$. f is the constant function of 1. Recall that a trigonometric polynomial is of the form

$$\sum_{k=0}^{n} a_k \cos(kx) + \sum_{k=0}^{n} b_k \sin(kx) + a_0$$

where a_k's, the b_k's, and a_0 are real constants. Hence, the set of all trigonometric polynomials is closed under addition and scalar multiplication. Due to the following fairly standard trigonometric identities,

$$\begin{aligned}
\cos(mx)\cos(nx) &= \frac{1}{2}\left[\cos((m+n)x) + \cos((m-n)x)\right], \\
\sin(mx)\sin(nx) &= \frac{1}{2}\left[\cos((m-n)x) - \cos((m+n)x)\right], \\
\sin(mx)\cos(nx) &= \frac{1}{2}\left[\sin((m+n)x) + \sin((m-n)x)\right]
\end{aligned}$$

the set of all trigonometric polynomials is closed under multiplication. Hence, the set of all trigonometric polynomials forms a function algebra.

Let x and y be points in the unit circle. Then, the coordinates of point x are $(\cos(x), \sin(x))$, and the coordinates of point y are $(\cos(y), \sin(y))$. So, if $x \neq y$, then either $\sin(x) \neq \sin(y)$ or $\cos(x) \neq \cos(y)$. Thus, the set of all trigonometric polynomials separates points.

Applying the real Stone-Weierstrass theorem, we conclude that the set of all trigonometric polynomials is dense in the set of all continuous 2π-periodic functions, which completes the proof. ■

Next, we will look at the generalization of the Stone-Weierstrass theorem to complex numbers. Until now, we have dealt exclusively with real-valued functions. We will now look at approximations on the complex plane and try to generalize the Stone-Weierstrass theorem to complex-valued functions.

Let $C^0(X,\mathbb{C})$ denote the set of all continuous, bounded complex-valued functions defined on X. Unfortunately, the hypotheses of the real Stone-Weierstrass theorem will not suffice for the complex Stone-Weierstrass theorem. The complex Stone-Weierstrass theorem requires an additional condition that states the conjugate of a complex-valued function in the function algebra must remain in the function algebra. The conjugate of a complex-valued function f, denoted $\bar{f}$, is the function obtained by turning every i into $-i$ in the expression of f. Consequently, the conjugate of any real-valued function is itself. Every complex-valued function f can be split into an imaginary part denoted I_f and a real part denoted R_f, where

$$R_f = \frac{f+\bar{f}}{2} \text{ and } I_f = \frac{f-\bar{f}}{2i}.$$

Accordingly, $R_f(x), I_f(x) \in \mathbb{R}$ for all x in the respective domains of I_f and R_f. Since every complex-valued function can be written as $R_f(x) + I_f(x)i$,

$$f(a) \neq f(b) \Rightarrow R_f(a) \neq R_f(b) \text{ or } I_f(a) \neq I_f(b). \tag{6.1}$$

The above property will help us transition from functions with a range of $\mathbb{R}$ to functions with a range of $\mathbb{C}$. We are now ready to prove the complex Stone-Weierstrass theorem.

Theorem 6.4 (Complex Stone-Weierstrass theorem [Sim83])**.** *Let X be a compact Hausdorff space. Let A be a function algebra that separates points and contains the conjugate of each of its functions. Then, A is dense in $C^0(X,\mathbb{C})$ with respect to the supremum norm.*

The proof of the complex Stone-Weierstrass theorem relies on the real Stone-Weierstrass theorem.

Proof. Let X be a compact Hausdorff space, and let A be a function algebra that separates points and contains the conjugate of each of its functions. Let B be the set of all real-valued functions in A. Then, B is closed under addition, multiplication, and scalar multiplication by the nature of real-valued functions. By the definition of R_f and I_f, if $f \in A$, then $R_f, I_f \in A$ since A is a function algebra that contains each of its conjugates. Because $R_f, I_f \in C^0(X,\mathbb{R})$ for all $f \in C^0(X,\mathbb{C})$,

$$f \in A \Rightarrow R_f, I_f \in B. \tag{6.2}$$

A is a function algebra, so it contains a nonzero constant function $g(x) = a + bi$ where $a, b \in \mathbb{R}$. Since A contains the conjugate of each of its functions, the nonzero constant function $\bar{g}(x) = a - bi$ belongs to A, so

$$g(x)\bar{g}(x) = (a-bi)(a+bi) = a^2 + b^2$$

is a real-valued nonzero constant function in A, and thus $g\bar{g}$ is in B. Hence, B is a function algebra.

It is left for us to show that B separates points to conclude by the real Stone-Weierstrass theorem that B is dense in $C^0(X,\mathbb{R})$. Let a and b be distinct points in X. Since A separates points, there exists a $f \in A$ such that $f(x) \neq f(y)$. By the property highlighted in implication 6.1, either $R_f(x)$ or $I_f(x)$ takes on different values at a and b. Implication 6.2 along with the fact that the a and b we considered were arbitrary allows us to conclude that B

separates points. Therefore, B is dense in $C^0(X, \mathbb{R})$ by the real Stone-Weierstrass theorem. So, $\bar{B} = C^0(X, \mathbb{R})$. Consider an arbitrary function $f \in C^0(X, \mathbb{C})$. We will show that f must be in $\bar{A}$. It is clear that $R_f, I_f \in C^0(X, \mathbb{R}) = \bar{B}$, so since $f = R_f + I_f i$ and $\bar{A}$ is a function algebra by lemma 5.3, $f \in \bar{A}$. Hence, $\bar{A} = C^0(X, \mathbb{C})$, so A is dense in $C^0(X, \mathbb{C})$. ■

Acknowledgments

The author would like to thank Jay Paek and Linn Zaw for helpful conversations. He would also like to thank Simon-Rubinstein Salzedo for suggesting this topic.

References

[DD] Kenneth R Davidson and Allan P Donsig. Real analysis and applications: Theory in practice supplementary material.

[Pug15] Charles C. Pugh. *Real mathematical analysis*. Undergraduate Texts Math. Cham: Springer, 2nd ed. edition, 2015.

[RS23] Simon Rubinstein-Salzedo. *Transition to proofs*. World Scientific, 2023.

[Rud76] Walter Rudin. *Principles of mathematical analysis. 3rd ed.* 1976.

[Sim83] George F. Simmons. *Introduction to topology and modern analysis. Repr. of the 1963 orig. publ. by McGraw-Hill, New York*. 1983.

Appendix A. Extension of the Weierstrass Approximation Theorem to Arbitrary Closed and Bounded Intervals

Suppose that the Weierstrass approximation theorem holds for the interval $[0, 1]$. The following argument shows that it holds for all arbitrary closed and bounded intervals. Let $f \in C([a, b])$ for $a, b \in \mathbb{R}$ and let $\varepsilon > 0$. Let

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

where $x \in [0, 1]$. Since f is continuous, g must be continuous as $a + x(b - a)$ is a linear polynomial. By the Weierstrass approximation theorem on $[0, 1]$, there exists a polynomial q such that

$$|g(x) - q(x)| < \varepsilon$$

for every $x \in [0, 1]$. Define

$$p(t) = q\left(\frac{t - a}{b - a}\right)$$

where $t \in [a, b]$. Since $\frac{t-a}{b-a}$ is a polynomial function of t with degree one, p is a polynomial. Besides, for every $t \in [a, b]$,

$$\begin{aligned} |f(t) - p(t)| &= \left| f(t) - q\left(\frac{t - a}{b - a}\right) \right| \\ &= \left| g\left(\frac{t - a}{b - a}\right) - q\left(\frac{t - a}{b - a}\right) \right| < \varepsilon, \end{aligned}$$

since $\frac{t-a}{b-a} \in [0, 1]$. Therefore, p uniformly approximates f over the interval $[a, b]$.