THE STONE-WEIERSTRASS THEOREM AND ITS APPLICATIONS TO L^2 SPACES

PHILIP GADDY

Abstract. Throughout the course of this paper, we will first prove the Stone-Weierstrass Theroem, after providing some initial definitions. Afterwards, we will introduce the concept of an L^2 space and, using the Stone-Weierstrass theorem, prove that $L^2[0, 1]$ is separable.

Contents

1. Introduction

Most calculus students are familiar with the concept of a Taylor Polynomial and some of the associated results. While these polynomials can be made arbitrarily close to certain functions on a closed interval, they require that the functions be analytic (highly differentiable) which is a relatively small subclass of functions. This raises the question of whether this condition is actually necessary for a function to be approximated on a closed interval, to an arbitrary degree, by a polynomial. It turns out, as Karl Weierstrass proved with the Weierstrass Approximation Theorem, it is only necessary for f to be continuous on the closed interval in order for such polynomials to exist. This, however, assumes that f maps from $\mathbb{R}$ into $\mathbb{R}$. Marshall Stone, with the Stone-Weierstrass theorem, generalized the result to any continuous function that maps the elements of a compact Hausdorff space into $\mathbb{R}$. This result also generalizes the approximating functions from polynomials to the members of a subalgebra of the continuous functions that map X to $\mathbb{R}$. The remainder of this paper will be dedicated to exploring this theorem in more detail, as well as some applications of it.

2. Stone-Weierstrass Theorem

Before we get to the actual statement of the theorem, let's begin by defining a few terms necessary to state and prove this theorem. Now, we're interested in considering continuous on compact Hausdorff spaces.

Date: September 1, 2016.

Definition 2.1. A **Hausdorff space**, X, is a topological space such that for any two distinct elements, x_1 and x_2, there exist two open sets, U_1 and U_2, such that $x_1 \in U_1$, $x_2 \in U_2$, and $U_1 \cap U_2 = \varnothing$

$C(X, \mathbb{R})$ will be used to denote the set of all continuous $f : X \to \mathbb{R}$, where X is a compact Hausdorff space. Now, for any $f, g \in C(X, \mathbb{R})$, $x \in X$, and $c \in \mathbb{R}$, let $f + g$ be defined by $(f + g)(x) = f(x) + g(x)$ and cf by $(cf)(x) = c \cdot f(x)$. Since both $f + g$ and cf are continuous if f, g are continuous, $C(X, \mathbb{R})$ forms a vector space over $\mathbb{R}$. In fact, $C(X, \mathbb{R})$ is actually a normed space, with norm defined as follows

Definition 2.2. The **uniform norm** of a function, f, denoted by $\|f\|_u$ is given by $\|f\|_u = \sup_{x \in X} |f(x)|$.

Since X is compact and f is continuous, this supremum is always guaranteed to exist.

Furthermore, if, for all $f, g \in C(X, \mathbb{R})$ and $x \in X$, we define fg by $(fg)(x) = f(x)g(x)$, it follows that $fg \in C(X, \mathbb{R})$. Thus, we see that $C(X, \mathbb{R})$ actually forms an algebra over $\mathbb{R}$.

Definition 2.3. An **algebra** is a vector space over $\mathbb{R}$, $\mathcal{A}$, equipped with a metric and an associative bilinear product, $m : \mathcal{A} \times \mathcal{A} \to \mathcal{A}$. $\mathcal{A}$ will be referred to as unital if the operation m has an identity element. A **subalgebra** is a subset of an algebra which is also an algebra under the same operations.

The following definitions define a few terms related to algebras and sets of functions

Definition 2.4. A set of functions, S, is said to **separate points** if, for all x, y where $x \neq y$, there exists some $f \in S$ such that $f(x) \neq f(y)$.

Definition 2.5. An algebra, $\mathcal{A} \subset C(X, \mathbb{R})$, is said to be a **lattice** if, for all $f, g \in \mathcal{A}$, $\max\{f, g\}, \min\{f, g\} \in \mathcal{A}$.

Now that all preliminary definitions are out of the way, let's proceed with the statement of the Stone-Weierstrass theorem.

Theorem 2.6 (Stone-Weierstrass Theorem)**.** *Suppose $\mathcal{A}$ is a subalgebra of $C(X, \mathbb{R})$ that separates points, where X is a compact Hausdorff space. If there exists $x_0 \in X$ such that $f(x_0) = 0$ for all $f \in \mathcal{A}$, then $\mathcal{A}$ is dense in $\{f \in C(X, \mathbb{R})\} \mid f(x_0) = 0\}$. Otherwise, $\mathcal{A}$ is dense in $C(X, \mathbb{R})$.*

Before proving this theorem, we're going to slightly restate it to make it easier to prove. Since a set A is dense in another set, B, if and only if $\overline{A} = B$, where $\overline{A}$ is the closure of A, it will suffice to consider, and classify, the closed subalgebras of $C(X, \mathbb{R})$. The following is a restatement of the above theorem which takes this into account.

Theorem 2.7 (Stone-Weierstrass Theorem (Restatement))**.** *If $\mathcal{A}$ is a closed subalgebra of $C(X, \mathbb{R})$ that separates points, then either $\mathcal{A} = C(X, \mathbb{R})$ or $\mathcal{A} = \{f \in C(X, \mathbb{R})\} \mid f(x_0) = 0\}$ for some $x_0 \in X$.*

In order to prove this theorem, let's first start by considering a simpler case- one in which X consists of only two points, x_1 and x_2. Since any function $f : X \to \mathbb{R}$ is described entirely by the image of x_1 and x_2, each function can be represented by the ordered pair $(f(x_1), f(x_2))$. Thus, it will suffice to consider the closed subalgebras of $\mathbb{R}^2$.

Lemma 2.8. *Consider* $\mathbb{R}^2$ *as an algebra under coordinate addition and multiplication. The only subalgebras for* $\mathbb{R}^2$ *are* $\mathbb{R}^2$, $\{(0,0)\}$, $\{(x,0) \mid x \in \mathbb{R}\}$, $\{(0,x) \mid x \in \mathbb{R}\}$, *and* $\{(x,x) \mid x \in \mathbb{R}\}$.

Proof. Since each one of these sets is closed under coordinatewise addition and multiplication, they each form a subalgebra of $\mathbb{R}^2$. To see that these are the only ones, consider a point $(a,b) \in \mathcal{A}$. If $\mathcal{A}$ contains a point such that $a \neq b \neq 0$, then (a,b) and (a^2,b^2) are linearly independent. As a result, $\mathcal{A} = \mathbb{R}^2$. Now, the cases $a = b \neq 0$, $a \neq 0 = b$, or $a = 0 \neq b$ generate the other three nonzero subalgebras mentioned above. Finally, the only case remaining is if the only point happens when $a = b = 0$, which corresponds to the set $\{(0,0)\}$. Thus, the subalgebras mentioned above are the only possibilities. □

Now, we've essentially proven the theorem when X consists of 2 elements. It turns out that, with a few additional lemmas, proving the theorem in this case proves it in general. Before exploring this in more detail, we will first prove these additional lemmas, which focus largely on classifying functions within the subalgebra as well as determining whether a certain function is contained within the algebra. The first of these is a special case of the Weierstrass Approximation theorem, applied to the absolute value function.

Lemma 2.9. *For any* $\epsilon > 0$ *there is a polynomial* P *on* $\mathbb{R}$ *such that* $P(0) = 0$ *and* $||x| - P(x)| < \epsilon$ *for* $x \in (-1,1)$.

Proof. Let's start by considering the Maclaurin series for $f(t) = (1-t)^{\frac{1}{2}}$, given by $1 - \sum_{k=1}^{\infty} a_k t^k$, for constants a_k. Computing several derivatives, we see that $f^{(k+1)}(t) = \frac{(2k-1)f^{(k)}(t)}{2}$ for $k \geq 1$. Therefore,

$$a_{k+1} = \frac{f^{(k+1)}(0)}{(k+1)!} = \frac{(2k-1)f^{(k)}(0)}{2(k+1)k!} = \frac{(2k-1)a_k}{2(k+1)}$$

Therefore, we have that

$$\lim_{k\to\infty} \left| \frac{a_{k+1}t^{k+1}}{a_k t^k} \right| = \lim_{k\to\infty} \frac{2k-1}{2k+2}|t| = |t|$$

Thus, applying the ratio test, we see that the above series converges for $t \in (-1,1)$.

Now, let's show that this Maclaurin series actually equals $f(t)$. To see this, note that, according to Taylor's theorem, we know that the remainder for any Maclaurin polynomial of degree n must be given by $R_n(t) = \frac{f^{(n+1)(c)}}{(n+1)!}t^{k+1}$ for some $c \in (0,1)$. Now, since $t \in (-1,1)$ and $f^{(n+1)}(c)$ achieves it's maximum for $c = 0$, we see that this term must be less than a_{n+1}. Now, since the series above converges, we have that $\lim_{n\to\infty} R_n(t) = 0$, as required.

To see that this series also converges for $t = 1$, we can apply the monotone convergence theorem to the counting measure on the natural numbers to conclude that

$$\sum_{k=1}^{\infty} a_k = \lim_{t\to 1} \sum_{k=1}^{\infty} a_k = 1 - \lim_{t\to 1}(1-t)^{\frac{1}{2}} = 1$$

Therefore, we have that the Maclaurin series for $f(t)$ converges to $f(t)$ for $t \in [-1,1]$.

This means that for every $\varepsilon > 0$ there exists a polynomial, $Q(t)$, such that $|f(t) - Q(t)| < \frac{1}{2}\varepsilon$. Substituting $t = 1 - x^2$, we see that

$$|f(1 - x^2) - Q(1 - x^2)| = ||x| - R(x)| < \frac{1}{2}\varepsilon$$

where $R(x)$ is the polynomial given by $Q(1 - x^2)$. Finally, let $P(x) = R(x) - R(0)$. Then, we have that

$$||x| - P(x)| < ||x| - R(x)| + |R(0)| < \varepsilon$$

where the last step follows from plugging $x = 0$ into the above inequality. □

Now that we have that the absolute value function can be approximated sufficiently closely, we can use this to show that, for any f contained in the closed subalgebra, $\mathcal{A}$, $|f| \in \mathcal{A}$. As a result, $\mathcal{A}$ must be a lattice.

Lemma 2.10. *If $\mathcal{A}$ is a closed subalgebra of $C(X, \mathbb{R})$, then $|f| \in \mathcal{A}$ whenever $f \in \mathcal{A}$ and $\mathcal{A}$ is a lattice.*

Proof. If $f = 0$, then $|f| = 0$, and therefore, $|f| \in \mathcal{A}$. Now, consider $f \neq 0$. Let $h : X \to [-1, 1]$ be given by $h = \frac{f}{\|f\|_u}$. Therefore, by lemma 2.9, for every $\epsilon > 0$, there exists a polynomial P such that $\||h| - P \circ h\|_u < \epsilon$. Since $h \in \mathcal{A}$ and P has no constant term, $P \circ h \in \mathcal{A}$. Now, since we have constructed a sequence whose limit if $|h|$ and $\mathcal{A}$ is closed, it follows that $|h| \in \mathcal{A}$. Thus, $|f| = \|f\|_u |h| \in \mathcal{A}$, as required.

To see that $\mathcal{A}$ is a lattice, note that, by definition,

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

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

Therefore, by the first part of this lemma, we have that $\max\{f, g\}, \min\{f, g\} \in \mathcal{A}$. □

While the first of these results does provide some information on the contents of any potential subalgebra, the second condition gives an important restriction on the structure of any closed subalgebra, which can in turn be used to provide a condition for the inclusion of a function f within this algebra.

Lemma 2.11. *Suppose that $\mathcal{A}$ is a closed lattice in $C(X, \mathbb{R})$ and $f \in C(X, \mathbb{R})$. If for every $x, y \in X$ there exists $g_{xy} \in A$ such that $g_{xy}(x) = f(x)$ and $g_{xy}(y) = f(y)$, then $f \in \mathcal{A}$.*

Proof. Let $\epsilon > 0$ be given. For all $x, y \in X$, define $U_{xy} = \{z \in X \mid f(z) < g_{xy}(z) + \epsilon\}$ and $V_{xy} = \{z \in X \mid f(z) > g_{xy}(z) - \epsilon\}$ and note that $x, y \in U_{xy}$ and $x, y \in V_{xy}$. Fix $y \in X$. Since, for all x, $x \in U_{xy}$, the set $\{U_{xy} \mid x \in X\}$ forms an open cover of X. Since X is compact, there exists a finite subcover, $\{U_{x_i y} \mid 1 \leq i \leq n\}$. Let $g_y = \max\{g_{x_1 y}, \ldots, g_{x_n y}\}$. Now, we have that $f < g_y + \epsilon$ over X and $f > g_y - \epsilon$ on $V_y = \cap_{i=1}^{n} V_{x_i y}$. Since, for all y, $y \in V_y$, the set $\{V_y \mid y \in X\}$ is an open cover for X. Therefore, because X is compact, there exists a finite subcover, $\{V_{y_i} \mid 1 \leq i \leq k\}$. Let $g = \min\{g_{y_1}, \ldots, g_{y_k}\}$. From this we see that $\|f - g\|_u < \epsilon$. Since $\mathcal{A}$ is a lattice, it follows that $g \in \mathcal{A}$. Finally, since $\mathcal{A}$ is closed, we have that $f \in \mathcal{A}$. □

Now that we have conditions for classifying and identifying the functions that must be in any closed subalgebra of $C(X,\mathbb{R})$, we can prove the Stone-Weierstrass theorem. This will be done by appealing to the much simpler case proved in lemma 2.8, and then, using lemmas 2.10 and 2.11, expanding this to any compact Hausdorff space X.

Proof: Stone-Weierstrass theorem. Let $A_{xy} = \{(f(x), f(y)) \mid f \in \mathcal{A}\}$. Now, since $\mathcal{A}$ is a subalgebra of $C(X,\mathbb{R})$, A_{xy} is a subalgebra of $\mathbb{R}^2$. Therefore, by lemma 2.8 A_{xy} is either $\mathbb{R}^2$, $\{(0,0)\}$, $\{(x,0) \mid x \in \mathbb{R}\}$, $\{(0,x) \mid x \in \mathbb{R}\}$, or $\{(x,x) \mid x \in \mathbb{R}\}$. Now, since $\mathcal{A}$ separates points, A_{xy} cannot be $\{(0,0)\}$ or $\{(x,x) \mid x \in \mathbb{R}\}$. If $A_{xy} = \mathbb{R}^2$, then it follows from lemma 2.10 and lemma 2.11 that $\mathcal{A} = C(X,\mathbb{R})$. Finally, if A_{xy} is $\{(x,0) \mid x \in \mathbb{R}\}$ or $\{(0,x) \mid x \in \mathbb{R}\}$, then there exists some x_0 ($y = x_0$ or $x = x_0$, respectively) such that $f(x_0) = 0$ for all $f \in \mathcal{A}$. Furthermore, from lemma 2.10 and lemma 2.11 , we have that $\mathcal{A} = \{f \in C(X,\mathbb{R})\} \mid f(x_0) = 0\}$. Finally, note that if $\mathcal{A}$ contains a constant function, then there does not exist an x_0 such that $f(x_0) = 0$ for all $f \in \mathcal{A}$. Thus, $\mathcal{A} = C(X,R)$. □

3. L^2 Spaces

To start, let's define $L^2[0,1]$ and discuss some of its properties. This set consists of all functions $f : [0,1] \to \mathbb{R}$ which are square, Lebesgue integrable ($\int f^2 d\mu$ is finite). Now, if $f, g \in L^2[0,1]$ and $c \in \mathbb{R}$, then $f + g \in L^2[0,1]$ (this follows from $\int (f+g)^2 d\mu = \int f^2 d\mu + \int g^2 d\mu + 2\int fg d\mu$ and all these terms are finite) and $cf \in L^2[0,1]$. Thus, $L^2[0,1]$ forms a vector space over $\mathbb{R}$. Additionally, we can define a norm over $L^2[0,1]$, where, for $f \in L^2[0,1]$, $\|f\|_{L^2} = (\int_0^1 f^2 d\mu)^{\frac{1}{2}}$. This space is actually a Hillbert space, although the proof that $L^2[0,1]$ is complete is omitted here.

Now, we wish to show that $L^2[0,1]$ is separable.

Definition 3.1. A space is **separable** if it has a dense, countable subset.

We wish to show that this subset is the set of all polynomials, $p : [0,1] \to \mathbb{R}$ (this set will be denoted $\mathbb{P}[0,1]$). In order to do this, we will first prove that $\mathbb{P}[0,1]$ is dense in $C([0,1],\mathbb{R})$. Then, we will show that $C([0,1],\mathbb{R})$ is dense in $L^2[0,1]$. From these two results, and an additional lemma, we will conclude that $\mathbb{P}[0,1]$ is dense in $L^2[0,1]$.

Lemma 3.2 (Weierstrass Approximation Theorem)**.** *The set of real-valued polynomials, $\mathbb{P}$, is dense in $C(\mathbb{R},\mathbb{R})$.*

Proof. If $f, g \in \mathbb{P}$, then $f + g \in \mathbb{P}$ and $f \cdot g \in \mathbb{P}$. Therefore, $\mathbb{P}$ forms a subalgebra of $C(\mathbb{R},\mathbb{R})$. Additionally, since $\mathbb{P}$ contains every constant, there does not exist an $x_0 \in \mathbb{R}$ such that $f(x_0) = 0$ for all $f \in \mathbb{P}$. Thus, by the Stone-Weierstrass theorem, $\mathbb{P}$ is dense in $C(\mathbb{R},\mathbb{R})$. □

Now that we've proven the above result, we will next show that $C([0,1],\mathbb{R})$ is dense in $L^2([0,1])$. In order to show this, we will prove the equivalent statement that any function $f \in L^2([0,1])$ can be approximated by a continuous function, i.e. for every $\varepsilon > 0$, there exists a continuous function g such that $\|f - g\|_{L^2} = 0$. To show that this must be the case, we will start by proving this for indicator functions. From this, we can derive the result for simple functions, from which follows the same result for nonnegative functions. The final result follows from this.

Now, before proving this, we will first state a theorem that will be necessary to prove this

Theorem 3.3 (Lebesgue Dominated Convergence Theorem)**.** *Let $\{f_n\}$ be a sequence of measurable functions such that this sequence converges pointwise to some f and $|f_n| \leq g$ for all n and an integrable function g. Then, f is integrable and $\int f d\mu = \lim_{n\to\infty} \int f_n d\mu$*

Lemma 3.4. *The set $C([0,1],\mathbb{R})$ is dense in $L^2[0,1]$.*

Proof. It suffices to show that, for all $f \in L^2[0,1]$, there exists a sequence of functions, $\{g_n\}$, such that $\lim_{n\to\infty}\|f - g_n\|_{L^2} = 0$. In this case, we say that f is approximated by $\{g_n\}$.

First, let A be a closed subset of $[0,1]$ and K_A be it's indicator function. Now, define $t(x) = \inf_{y\in A}\{|x-y|\}$ and $g_n(x) = \frac{1}{1+nt(x)}$. For all n, we have that $|g_n|$ is continuous and, for all $x \in [0,1]$, $|g_n(x)| \leq 1$. In particular, for all $x \in A$, $g_n(x) = 1$. Furthermore, for $x \in B = [0,1] \setminus A$, $\lim_{n\to\infty} g_n(x) = 0$. Thus,

$$\lim_{n\to\infty} \|g_n(x) - K_A(x)\|_{L^2} = \lim_{n\to\infty} \left(\int_B g_n(x)^2 dx\right)^{\frac{1}{2}} = \left(\int_B \lim_{n\to\infty} g_n(x)^2 dx\right)^{\frac{1}{2}} = 0$$

where the last step follows from the Lebesgue dominated convergence theorem. Therefore, the indicator function for any closed subset of $[0,1]$ can be approximated by a sequence of continuous functions. Since simple functions are a finite linear combination of such indicator functions and continuity is preserved under finite linear combinations, any simple function can be approximated by a sequence of continuous functions.

Now, suppose that $f \in L^2[0,1]$ is nonnegative. Since f has these properties, there exists a sequence of nonnegative simple functions, $\{s_n\}$, such that $\lim_{n\to\infty} s_n = f$. Since each s_n, and f, is nonnegative, we have that $(f - s_n)^2 \leq f^2$. Therefore, by the Lebesgue dominated convergence theorem, it follows that $\|f - s_n\|_{L^2} = 0$. Thus, since every simple function can be approximated by a continuous function, every nonnegative function can be approximated by a sequence of continuous functions.

Finally, for any $f \in L^2[0,1]$, we can say that $f = f^+ - f^-$, where

$$f^+(x) = \begin{cases} f(x) & f(x) > 0 \\ 0 & \text{otherwise} \end{cases}$$

$$f^-(x) = \begin{cases} -f(x) & f(x) < 0 \\ 0 & \text{otherwise} \end{cases}$$

Since both f^+ and f^- are nonnegative functions, they can each be approximated by a sequence of continuous functions. Thus, f can be approximated by a sequence of continuous functions. Therefore, $C([0,1],\mathbb{R})$ is dense in $L^2([0,1])$, as required. □

Now, for the final lemma, we will prove that if a metric space M is dense in a metric space N, and N is dense in a metric space O, where M, N, and O all have the same metric. Then, M is dense in O.

Lemma 3.5. *Let M, N, and O be metric spaces equipped with the same metric, d. If M is dense in N and N is dense in O, then M is dense in O.*

Proof. Let $\epsilon > 0$ be given and choose $o \in O$. Since N is dense in O, there exists some $n \in N$ such that $d(o,n) < \frac{\epsilon}{2}$. Furthermore, since M is dense in N, there exists some $m \in M$ such that $d(m,n) < \frac{\epsilon}{2}$. Using the triangle inequality, we have that $d(m,o) \leq d(m,n) + d(n,o) < \epsilon$. Therefore, M is dense in O. □

Now, using the above lemma, we know that $\mathbb{P}[0,1]$ is dense in $L^2[0,1]$ under the uniform norm. Thus, it only remains to show that the L^2 norm is less than or equal to the uniform norm.

Theorem 3.6. $L^2[0,1]$ *is separable.*

Proof. By lemma 3.4 we have that $C([0,1],\mathbb{R})$ is dense in $L^2[0,1]$ under the L^2 norm. Furthermore, we have that the polynomials over $[0,1]$, $\mathbb{P}[0,1]$, are dense in $C([0,1],\mathbb{R})$ under the supremum norm. Now, note that

$$\left(\int_0^1 f^2 dx\right)^{\frac{1}{2}} \leq \left(\int_0^1 \|f\|_u^2 dx\right)^{\frac{1}{2}} = (\|f\|_u^2)^{\frac{1}{2}} = \|f\|_u$$

Thus, $\mathbb{P}[0,1]$ is dense in $C([0,1],\mathbb{R})$ under the L^2 norm. Therefore, by lemma 3.5 $\mathbb{P}[0,1]$ is dense in $L^2[0,1]$ under the L^2 norm. So, since $\mathbb{P}[0,1]$ is countable, $L^2[0,1]$ is separable. □

Now we have shown that $L^2[0,1]$ is a separable Hillbert space. Consequently, it has a countable orthogonal basis, which, among other things, allows us to define fourier coefficients for functions in this space. This can prove exceedingly useful in the study of these functions.

Acknowledgments. I would like to thank my mentors Ben O'Connor and Randy VanWhy for helping me to choose this topic as well as to understand the material involved. Additionally, I would also like to thank Peter May for organizing this summer's research program.

References

[1] Gerald B. Folland. Real Analysis: Modern Techniques and Their Applications Wiley-Interscience Texts
[2] Walter Rudin. Principles of Mathematical Analysis. McGraw-Hill, Inc.