STONE-WEIERSTRASS THEOREM

ALEX THOLEN

1. Definitions

The Weierstrass approximation theorem states that every single continuous function on the closed interval $[a, b]$ can be uniformly approximated as closely as desired by a polynomial. The Stone-Weierstrass theorem is an improvement made by Stone. It extends $[a, b]$ to an arbitrary compact Hausdorff space, and instead of polynomial elements more general subsets of $C(X)$ are considered.

Definition 1.1. A family χ of arbitrary functions on a domain X is said to be a separating family for the domain X if, whenever x and y are distinct points of X, there is some function f in χ with distinct values $f(x), f(y)$ at these points.

Definition 1.2. An open cover of a space is a list of open sets where any point in the space is in at least one of those open sets. For example, $[(0, \frac{1}{2}), (\frac{1}{4}, \frac{3}{4}), (\frac{5}{8}, \frac{7}{8}), (\frac{13}{16}, \frac{15}{16}) \cdots]$ is an open cover of $(0, 1)$.

Definition 1.3. A compact space is a space that whenever you have an open cover of the space there is a finite length open cover consisting of only sets in the bigger open cover. For example, any space with finitely many points is a compact space.

Definition 1.4. The linear lattice operations are addition of two lattices, multiplication of two lattices, and multiplication of a lattice and a real number.

Definition 1.5. The two lattice operations are the lattice maximum $\cup$ and the lattice minimum $\cap$. They denote the maximum and minimum pointwise. For example, the minimum and maximum of two points in the plane: $(3, 6) \cup (4, 2) = (4, 6)$ and $(3, 6) \cap (4, 2) = (3, 2)$.

Definition 1.6. $U_i(\chi)$ is the set of all functions reachable using i lattice operations of the terms in χ. $U(\chi)$ is the set of all functions reachable using lattice operations of the terms in χ.

Definition 1.7. The set S^* of S is formed from all the points obtainable using $\cup, \cap$ and limits using points in S. For example, with the set

$$S = \{(1, 1), (1, \frac{1}{2}), (1, \frac{1}{4}), (1, \frac{1}{8}), \cdots, (\frac{1}{2}, 1), (\frac{1}{4}, 1), (\frac{1}{8}, 1), \cdots\}$$

then S^* is $\{(\frac{1}{2^n}, \frac{1}{2^m}), (0, \frac{1}{2^m}), (\frac{1}{2^n}, 0), (0, 0)\}$for all n, m.

Definition 1.8. The set S^- of S is formed by taking all points x where any open ball centered at x must contain a point in S.

Definition 1.9. The sequence a_k is defined as $a_1 = \frac{1}{2}$, $a_k = \frac{1}{2} \sum_{m,n \geq 1}^{m+n=k} a_m a_n$.

Date: June 17, 2019.

2. Theorems

Let's begin by just looking at what is necessary to be in the set of functions reachable using lattice operations.

Theorem 2.1. *Let X be a compact space, χ the family of all continuous (necessarily bounded) real functions on X, χ_0 an arbitrary subfamily of χ, and $U(\chi_0)$ the family of all functions (necessarily continuous) generated from χ_0 by the lattice operations and uniform passage to the limit. Then, a necessary and sufficient condition for a function f in χ to be in $U(\chi_0)$ is that for all $x, y \in X, \epsilon > 0$, there exists a function f_{xy} obtained by applying the lattice operations alone to χ_0 and such that $|f(x) - f_{xy}(x)| < \epsilon, |f(y) - f_{xy}(y)| < \epsilon$. Any function that is arbitrarily well approximated at two points can be arbitrarily well uniformly approximated on all of X.*

With this, the next two corollaries follow:

Corollary 2.2. *Firstly, if χ_0 has the property that for all $x, y \in X, x \neq y$ and for all $\alpha, \beta \in \mathbb{R}$ there exists a function f_0 in χ_0 for which $f_0(x) = \alpha$ and $f_0(y) = \beta$, then $U(\chi_0) = \chi$, i.e., any continuous function on X can be uniformly approximated by lattice polynomials in functions belonging to the prescribed family χ_0. This means that we can apply Theorem 2.1 to obtain the whole family of functions.*

Corollary 2.3. *Secondly, if a continuous real function f on a compact space X is the limit of a monotonic sequence f_n of continuous functions, then the sequence converges uniformly to f.*

If $\chi_0 = U(\chi_0)$, we can say make a stronger statement about the contents of $U(\chi_0)$.

Theorem 2.4. *Let X be a compact space, χ the family of continuous real functions on X, and χ_0 a subfamily of χ which is closed under the lattice operations and uniform passage to the limit. Then, χ_0 is completely characterized by the group of pairs $\chi_0(x, y)^* = \chi_0(x, y)^-$.*

There are stronger statements about $U(\chi_0)$ to be made, even without knowing $\chi_0 = U(\chi_0)$. For example, the following theorem is one.

Theorem 2.5. *Let X be a compact space, χ the family of all continuous (necessarily bounded) real functions on X, χ_0 an arbitrary subfamily of χ, and $U(\chi_0)$ the family of all functions (necessarily continuous) generated from χ_0 by the linear lattice operations and uniform passage to the limit. Then, a necessary and sufficient condition for a function f in χ to be in $U(\chi_0)$ is that f satisfies every linear relation of the form $\alpha g(x) = \beta g(y), \alpha\beta \geq 0$, which is satisfied by all functions in χ_0. If χ_0 is a closed linear sublattice of χ - that is, if $\chi_0 = U(\chi_0)$, then χ_0 is characterized by the system of all the linear relations of this form which are satisfied by every function belonging to it. The linear relations associated with an arbitrary pair of points x, y in X must be equivalent to one of the following distinct types:*

1. $g(x) = g(y) = 0$
2. $g(x) = 0$ *or* $g(y) = 0$
3. $g(x) = g(y)$
4. $g(x) = \lambda g(y)$ *or* $g(y) = \lambda g(x)$ *for a unique value* λ, $0 < \lambda < 1$.

Knowing this, we can show the following four corollaries.

Corollary 2.6. *In order for $U(\chi_0)$ to contain a non-vanishing constant function it is necessary and sufficient that the only linear relations of the form $\alpha g(x) = \beta g(y), \alpha\beta > 0$ satisfied by every function in χ_0 be those reducible to the form $g(x) = g(y)$.*

Corollary 2.7. *In order to have $U(\chi_0) = \chi$ it is sufficient that the functions in χ_0 satisfy no linear relation of the form (1) - (4) of Theorem 2.5.*

Corollary 2.8. *If X is compact and if χ_0 is a separating family for X and contains a non-vanishing constant function, then $U(\chi_0) = \chi$.*

Corollary 2.9. *If χ_0 is a separating family, then so is χ. If χ is a separating family and $U(\chi_0) = \chi$, then χ_0 is also a separating family.*

Note that in general, the family χ of all continuous functions isn't necessarily separating. However, it is known that if X is a compact Hausdorff space then χ is a separating family. Now that we know how the completion of a set of functions work, let's manipulate the functions themselves to prove some theorems.

Theorem 2.10. *If ϵ is any positive number and $\alpha \leq \gamma \leq \beta$ any real interval, then there exists a polynomial $p(\gamma)$ in the real variable γ with $p(0) = 0$ such that $||\gamma| - p(\gamma)| < \epsilon$ for $\alpha \leq \epsilon \leq \beta$.*

Theorem 2.11. *The sequence a_k is defined as $a_1 = \frac{1}{2}$, $a_k = \frac{1}{2}\sum_{m,n\geq 1}^{m+n=k} a_m a_n$. Let $\sigma_n = \sum_{k=1}^{k=n} a_k$. Then $\sigma_n < 1$ for all n.*

Theorem 2.12. *Let $\sigma(x) = \sum_{k=1}^{\infty} a_k x^k$. Then $\sigma(x) = 1 - \sqrt{1-x}$ when $x < 1$.*

And lets put that all together.

Theorem 2.13. *Let X be a compact space, χ the family of all continuous real functions on X, χ_0 an arbitrary subfamily of χ, and $U(\chi_0)$ the family of all functions (necessarily continuous) generated from χ_0 by the linear ring operations and uniform passage to the limit. Then a necessary and sufficient condition for a function f in χ to be in $U(\chi_0)$ is that f satisfy every linear relation of the form $g(x) = 0$ or $g(x) = g(y)$ which is satisfied by all function in χ_0. If χ_0 is a closed linear subring of χ - that is $\chi_0 = U(\chi_0)$ - then χ_0 is characterized by the system of all the linear relations of this kind which are satisfied by every function belonging to it. In other words, χ_0 is characterized by the partition of X into mutually disjoint closed subsets on each of which every function in χ_0 is constant and by the specification of that one, if any, of these subsets on which every function in χ_0 vanishes.*

3. Proofs

Proof of Theorem 2.1. Necessity is trivial, as if it is in $U(\chi_0)$ then setting f_{xy} to f works. Sufficiency is the hard part. Starting with f_{xy} in $U_1(\chi_0)$ let's make an approximation for f. Let $G_y : (z|f(z) - f_{xy}(z) < \epsilon)$ where x is fixed. By hypothesis x and y are in G_y, so the union of all G_y is the entire space X. Due to the fact that X is compact, there are such points $y_1, y_2, \cdots, y_n$ such that $G_{y_1} \cup G_{y_2} \cup \cdots \cup G_{y_n} = X$. Setting $g_x = \max(f_{xy_1}, f_{xy_2}, \cdots, f_{xy_n})$, we can see that for any z in X we have $z \in G_{y_k}$ for a suitable choice of k and hence $g_x(z) \geq f_{xy_k}(z) > f(z) - \epsilon$. On the other hand the fact that $f_{xy}(x) < f(x) + \epsilon$ implies that $g_x(x) < f(x) + \epsilon$. We now basically repeat for g. $H_x = (z|g_x(z) < f(z) + \epsilon)$. Evidently x is in H_x, and so the union of all H_x is X. Since X is compact, the union of $H_{x_1}, \cdots, H_{x_m}$ is still X. Set $h = \min(g_{x_1}, \cdots, g_{x_m})$. We can see that any $z \in X$ there exists k such that $z \in H_{x_k}$., and so $h(z) \le g_{x_k}(z) < f(z) + \epsilon$. Since $g_x(z) > f(z) - \epsilon$ that means that $h(z) > f(z) - \epsilon$. Combining the two gives us that $|f(z) - h(z)| < \epsilon$. Since we only used the lattice operations, we are good. ∎

Proof of Corollaries 2.2 and 2.3. We take χ_0 to be the totality of functions occurring in the sequence f_n. Then $U_1(\chi_0) = \chi_0$ since monotonicity implies that $f_m \cup f_n$ coincides with one of the two functions f_m and f_n while $f_m \cap f_n$ coincides with the other. The assumption that $\lim_{n\to\infty} f_n(x) = f(X)$ for every x now shows that the condition of 2.1 is satisfied. So, f is in $U(\chi_0)$; and f is therefore the uniform limit of functions occurring in χ_0. Since $|f(x) - f_n(x)|$ decreases as n increases and since $|f(x) - f_N(x)| < \epsilon$ for all x and a suitable choice of N, we can see that $|f(x) - f_n(x)| > \epsilon$ for all $n \ge N$, as was to be proved. ∎

Proof of Theorem 2.4. Our hypothesis that $\chi_0 = U(\chi_0)$ shows that $\chi_0(x, y)$ has $\chi_0(x, y)^*$ as its closure, due to Theorem 2.1. Let us suppose that that there is some other $\mathfrak{G} = U(\mathfrak{G}) \subset \chi$, and that $\chi_0(x, y)^* = \mathfrak{G}(x, y)^*$. Then the conditions for f in χ to belong to χ_0 are identical to those for it to belong to $\mathfrak{G}$. Hence χ_0 and $\mathfrak{G}$ are the same. ∎

Proof of Theorem 2.5. Since $\mathfrak{G}_0 = U(\chi_0)$ is closed under the lattice operations and uniform passage to the limit, Theorem 2 can be applied to $\mathfrak{G}_0$. However, the fact that $\mathfrak{G}_0$ is also closed under the linear operations can be used to produce effective simplifications. Due to that, the set $\mathfrak{G}_0(x, y)$ where x and y are arbitrary points in X must be the entire plane, a straight line passing through the origin, or the one point set consisting of the origin alone. If $\mathfrak{G}_0(x, y)$ is the origin alone, then it falls under case (1). If $\mathfrak{G}_0$ is a line through the origin, then we have case (2) if it coincides with one of the coordinate axis, case (3) if it is at 45°, and case (4) otherwise. When $\mathfrak{G}_0$ is the whole plane, on the other hand, there is no such linear relation. We know that f in χ belongs to $\mathfrak{G}_0$ if and only if $(f(x), f(y)) \in \mathfrak{G}_0(x, y)$. Since $\chi_0 \subseteq \mathfrak{G}_0$, the conditions imposed on $\mathfrak{G}_0$ are also satisfied by the functions in χ_0. This also works in the other direction, as sums, multiplications by constants, absolute values, and uniform limits keep it to be true. That means that we can just use the same α, β for $\alpha g(x) = \beta g(y)$. ∎

Proof of Corollary 2.6. Let's look at the four cases. $g(x) = 0$ and $g(y) = 0$ doesn't work as that is a vanishing function. $g(x) = 0$ and $g(y)$ unrestricted doesn't work because they must be equal. $g(x) = g(y)$ is true. And finally $g(x) = \lambda g(y)$ doesn't work because λ must be 1. ∎

Proof of Corollary 2.7. If there are no linear relations, then according to Theorem 2.5 everything is in $U(\chi_0)$. Henceforth $U(\chi_0) = \chi$. ∎

Proof of Corollary 2.8. Since χ_0 contains a non-vanishing constant, the only condition possible is that $g(x) = g(y)$, due to Corollary 2.6. Since χ_0 is a separating family, this can't happen. Henceforth due to Corollary 2.7 this is true. ∎

Proof of Corollary 2.9. The first statement is trivial: If there exists one in χ_0, there exists one in χ. Since χ_0 is subject to no linear relations of the form $g(x) = g(y)$ other than the ones that are satisfied in $U(\chi_0) = \chi$, the second part is also true. ∎

Proof of Theorem 2.10. Unless $\gamma = 0$ is inside the given interval (α, β), we can obviously take $p(\gamma) = \pm\gamma()$. Thus there is no loss of generality in confining our attention to intervals of the form $(-\gamma, \gamma)$. We can also just look at $(-1, 1)$ as one can just scale it up. Just take the partial sum for $\sqrt{1 - (1 - \gamma^2)}$, which by Theorem 2.12 is $1 - \sigma(1 - \gamma^2)$. ∎

Proof of Theorem 2.11. Let's prove this by induction. $\sigma_1 = \frac{1}{2}$. The inductive case is as follows: $\sigma_{n+1} = a_1 + \sum_{k=2}^{n+1} a_k = \frac{1}{2} + \frac{1}{2}\sum_{k=2}^{n+1}\sum_{i,j\geq 1}^{i+j=k} a_i a_j \leq \frac{1}{2} + \frac{1}{2}\sum_{i,j=1}^{n} a_i a_j \leq \frac{1}{2}(1+\sigma_n^2)$. Since $\sigma_n < 1$ by the inductive hypothosis, we get $\sigma_{n+1} < 1$. ∎

Proof of Theorem 2.12. To prove this we show that $\sigma(x)(2-\sigma(x)) = x$. This proves it as $(1-\sqrt{1-x})(2-(1-\sqrt{1-x}) = (1-\sqrt{1-x})(1+\sqrt{1-x}) = 1-(1-x) = x$.Write this using partial sums of the power series and we get

$$\left(\sum_{i=1}^{n} a_i x^i\right)\left(2 - \sum_{j=1}^{n} a_j x^j\right) = 2\sum_{k=1}^{n} a_k x^k - \sum_{i,j=1}^{n} a_i a_j x^{i+j}$$

$$= 2\sum_{k=1}^{n} a_k x^k - 2\sum_{k=2}^{n} a_k x^k - \sum_{1\leq i,j\leq n}^{i+j\geq n+1} a_i a_j x^{i+j} = x - \sum_{1\leq i,j\leq n}^{i+j\geq n+1} a_i a_j x^{i+j}$$

We can estimate the final term as follows:

$$\sum_{1\leq i,j\leq n}^{i+j\geq n+1} a_i a_j x^{i+j} \leq \sum_{1\leq i,j\leq n}^{i+j\geq n+1} a_i a_j \leq \sum_{k=n+1}^{\infty}\sum_{i,j\geq 1}^{i+j=k} a_i a_j \leq 2\sum_{k=n+1}^{\infty} a_k$$

This tends to zero, and so $\sigma(x)(2-\sigma(x)) = x$. That means that $\sigma(x) = 1 \pm \sqrt{1-x}$. Since $\sigma(x) < 1$, we know that $\sigma(x) = 1 - \sqrt{1-x}$. ∎

Proof of Theorem 2.13. We know that if f is in $U(\chi_0)$ then $|f|$ is also in $U(\chi_0)$ due to Theorem 2.12. Since X is compact, we know that f is bounded, and so $\alpha \leq f(x) \leq \beta$ for all x, we can find a polynomial $p_\epsilon(x)$ such that $||x| - p_\epsilon(x)| < \epsilon$ for $\alpha \leq x \leq \beta$ while $p_\epsilon(0) = 0$. It is clear than $p_\epsilon(f)$ is in $U(\chi_0)$ and that $||f(x) - p_\epsilon(f(x))| > \epsilon$ for all x in X. So, $|f|$ is the uniform limit of functions in $U(\chi_0)$, and so $|f|$ is in $U(\chi_0)$. We know that the characterization given in Theorem 2.5 applies to $U(\chi_0)$. If every function in $U(\chi_0)$ were to satisfy a linear relation of the form $g(x) = \lambda g(y)$, we would find for every$f$ in $U(\chi_0)$ that, f^2 being also in $U(\chi_0)$, the relations $f(x) = \lambda f(y)$, $f^2(x) = \lambda f^2(y)$, $\lambda^2 f^2(y) = \lambda f^2(y)$ would hold; and we can conclude that $f(y) = 0$ for every f in $U(\chi_0)$ or that $\lambda = 0, 1$. So we know that f is in $U(\chi_0)$ if and only if it satisfies all the linear relations $g(x) = 0$ or $g(x) = g(y)$ satisfied by every function in χ_0. The first characterization of the closed linear subrings of χ given in the statement of the theorem follows. Now, for the second characterization, let's define a stronger equality. $x \equiv y$ only if $f(x) = f(y)$ for all $f \in \chi_0$. We can partition the space into mutually disjoint subsets, each being the maximal set of mutually equal elements. We know that the partition space containing a x is closed. If x isn't in the same space as y, that means there is some f such that $f(x) \neq f(y)$. At most one partition space can contain a point where for all f then $f(z) = 0$, since any other space would be equivalent. And all elements in that space share that property. That means that we can distinctly partition them. Q.E.D. ∎

4. Sources

1. Random Wikipedia pages to get a basic understanding including the one on Closure (topology), .

2. Stone, M.H. "The Generalized Weierstrass Approximation Theorem". Mathematics Magazine, Vol. 21, No. 4 (Mar. - Apr., 1948), pp. 167-184