THE STONE-WEIERSTRASS THEOREM

1. *The function-lattice version of the theorem.*

The really new thing about Stone's approach to the approximation theorem was the approach via lattices of continuous functions, although Lebesgue had noticed the importance of approximating the absolute-value function earlier. There is a segment of the mathematical community formed of people who are as likely to encounter a lattice in their work as an algebra and for whom the lattice version of the theorem is at least as important as the algebra version. I'm one of them, but I also thought it more important to expose Stone's insight (about the density of a lattice with the two-point interpolation property) first.

The proof given below will work in any compact Hausdorff space, but because this course is set in the metric-space context, the reader needs only to think of the case in which X is a compact metric space (X, d).

Definition: A set L of real-valued functions on a set X is called a (**function**) **lattice** if it is closed under forming (finite) pointwise suprema and infima, *i.e.*, if whenever f, $g \in L$ the functions $f \vee g$ and $f \wedge g$ defined by

$$(f \vee g)(x) = \max\{f(x), g(x)\}$$
$$(f \wedge g)(x) = \min\{f(x), g(x)\}$$

belong to L.

Theorem [M. H. Stone]: Let X be a compact Hausdorff space (with at least two points) and $L \subseteq \mathcal{C}(X, \mathbb{R})$ a lattice of continuous functions with the property that for each pair of distinct points $x \neq y$ in X and pair of real numbers a, b there exists $f \in L$ for which $f(x) = a$ and $f(y) = b$. Then for every $\varphi \in \mathcal{C}(X, \mathbb{R})$ and $\epsilon > 0$ there is a function $h \in L$ for which $\|\varphi - h\|_\infty < \epsilon$; *i.e.*, the closure of L in the uniform norm is all of $\mathcal{C}(X, \mathbb{R})$.

Proof. Let $\varphi \in \mathcal{C}(X, \mathbb{R})$ and $\epsilon > 0$ be given. For each pair of points x, $y \in X$ we can by hypothesis find $f_{x,y} \in L$ for which $f_{x,y}(x) = \varphi(x)$ and $f_{x,y}(y) = \varphi(y)$ (even including the possibility that $x = y$). For fixed x, the open set $V_{x,y} = \{z \in X : \varphi(z) - \epsilon < f_{x,y}(z)\}$ is an open neighborhood of y, so the family of all such sets covers X and by compactness one can find a finite subfamily $\{V_{x,y_1(x)}, \dots, V_{x,y_{n(x)}(x)}\}$ that covers X (both the y's and the number of y's [and therefore V's] will depend on x, as we have indicated with the subscripts). Let $g_x = f_{x,y_1(x)} \vee \cdots \vee f_{x,y_{n(x)}(x)} \in L$; then $\varphi(z) - \epsilon < g_x(z)$ holds for all $z \in X$ and $g_x(x) = \varphi(x)$ holds for each $x \in X$. Now play the same game from above: for each $x \in X$ the open set $U_x = \{z \in X : g_x(z) < \varphi(z) + \epsilon\}$ is a neighborhood of x, so the family of all such U_x is an open cover of X from which one can extract a finite subcover $\{U_{x_1}, \dots, U_{x_n}\}$. Let $h = g_{x_1} \wedge \cdots \wedge g_{x_n} \in L$; then $\varphi(z) - \epsilon < h(z) < \varphi(z) + \epsilon$ holds for all $z \in X$, *i.e.*, $|h(z) - \varphi(z)| < \epsilon$ holds for all $z \in X$, or $\|h - \varphi\|_\infty < \epsilon$, as desired.

The "two-point interpolation" condition on L can be weakened slightly in an obvious way and the proof will still be valid: simply require that for each pair of distinct points $x \neq y$ in X and pair of real numbers a, b there exist for each $\epsilon > 0$ some $g \in L$ for which $|g(x) - a| < \epsilon$ and $|g(y) - b| < \epsilon$ and the proof will still go through (use $\epsilon/2$ instead of ϵ throughout). This fact is occasionally handy in complex-function-theory contexts. The same proof also adapts to a situation in which the uniform closure of L may not be all of $\mathcal{C}(X, \mathbb{R})$. Suppose that $L \subseteq \mathcal{C}(X, \mathbb{R})$ is a lattice that does not necessarily satisfy a "two-point interpolation" condition, but instead make a hypothesis about its relation to the given $\varphi \in \mathcal{C}(X, \mathbb{R})$; namely, that for each pair of distinct points $x \neq y$ in X there exist for each $\epsilon > 0$ some $f \in L$ for which $|f(x) - \varphi(x)| < \epsilon$ and $|f(y) - \varphi(y)| < \epsilon$. The proof will still go through (using $\epsilon/2$ instead of ϵ throughout), and we shall have proved the following corollary.[1]

Corollary: Let X be a compact Hausdorff space (with at least two points) and $L \subseteq \mathcal{C}(X, \mathbb{R})$ be a lattice of continuous real-valued functions on X. Then in order that $\varphi \in \mathcal{C}(X, \mathbb{R})$ belong to the closure of L in the uniform-norm topology, it is necessary and sufficient that for each pair of distinct points $x \neq y$ in X there exist for each $\epsilon > 0$ some $f \in L$ for which $|f(x) - \varphi(x)| < \epsilon$ and $|f(y) - \varphi(y)| < \epsilon$.

[1] This is essentially Lemma (4.48) in Folland's textbook.

2. *Ways to manufacture lattices of continuous functions.*

Well, first of all, lattices that arise naturally are usually vector lattices in the following sense.

Definition: A set $\mathcal{L}$ of real-valued functions on a set X is called a (**function**) **vector lattice** if it is a vector space under the pointwise operations (*i.e.*, a set of functions that contains with each pair of its elements f, g the function $f+g$ and contains αf for every $\alpha \in \mathbb{R}$ and $f \in \mathcal{L}$) that is also closed under forming (finite) pointwise suprema and infima, *i.e.*, for each pair of f, $g \in \mathcal{L}$, the functions $f \vee g$ and $f \wedge g$ defined by

$$(f \vee g)(x) = \max\{f(x), g(x)\}$$
$$(f \wedge g)(x) = \min\{f(x), g(x)\}$$

also belong to $\mathcal{L}$.

Vector lattices of continuous functions in a $\mathcal{C}(X, \mathbb{R})$ frequently arise from a **wedge** that is closed under taking (finite) suprema or infima. A subset $W \subseteq \mathcal{C}(X, \mathbb{R})$ is called a **wedge** if it is closed under addition and under multiplication by nonnegative real numbers. An example is furnished by the **convex functions** on a closed bounded interval $X \subseteq \mathbb{R}$ (or, more generally, on a compact convex set $X \subseteq \mathbb{R}^n$), *i.e.*, the functions that for each pair of points x_0, $x_1 \in X$ satisfy $f((1-\lambda)x_0 + \lambda x_1) \leq (1-\lambda)f(x_0) + \lambda f(x_1)$ for all $0 \leq \lambda \leq 1$. It is not difficult to show that these functions are necessarily continuous, and easy to show that the set of all such functions is closed under the operations of addition, multiplication by nonnegative scalars, and forming pointwise suprema ($f \vee g$). If W is a wedge of real-valued functions on X (which can now be any set) then $W - W = \{f - g : f, g \in W\}$ is obviously a vector space of functions, and if W was closed under one of the operations $\wedge$ or $\vee$ then $W - W$ is a lattice. For example, if W was closed under $\vee$ then the fact that the order relation on $\mathbb{R}$ is invariant under translation implies that for f's and g's in W

$$(f_1 - g_1) \vee (f_2 - g_2) + (g_1 + g_2) = (f_1 + g_2) \vee (f_2 + g_1)$$
$$(f_1 - g_1) \vee (f_2 - g_2) = (f_1 + g_2) \vee (f_2 + g_1) - (g_1 + g_2) \in W - W$$

so $W - W$ is closed under $\vee$, and since it is a vector space (and thus contains the negative of each of its elements) it must also be closed under $\wedge$, for $f \wedge g = -[(-f) \vee (-g)]$. Looking at the particular examples of the wedges of convex functions on compact convex subsets X of $\mathbb{R}$ or $\mathbb{R}^n$, we see that Stone's theorem implies that every continuous real-valued function on X can be uniformly approximated arbitrarily closely by a difference of two convex functions.[2]

Vector lattices can alternatively be characterized in the following way:

Proposition: For a vector space $\mathcal{L}$ of real-valued functions on a set X to be a vector lattice, it is necessary and sufficient that it contain with each of its elements f the absolute-value function of f (where $|f|(x) = (x \mapsto |f(x)|)$).

Proof. This condition is necessary because $|f| = f \vee (-f)$ and sufficient because

$$f \vee g = \frac{f+g}{2} + \left|\frac{f-g}{2}\right|, \qquad f \wedge g = \frac{f+g}{2} - \left|\frac{f-g}{2}\right|$$

showing that the lattice operations can be expressed in terms of the absolute-value function. {These relations are geometrically obvious: the greater of two numbers equals their average plus half the distance between them, while the smaller equals their average minus half the distance between them. They can also be proved without considering pointwise values by remembering that $|\varphi| = \varphi \vee (-\varphi)$: this relation makes $|f - g| = (f - g) \vee (g - f)$ and then the translation-invariance of the order makes $(f + g) + |f - g| = (f+g)+[(f-g)\vee(g-f)] = (2f)\vee(2g)$; now divide by 2 and use the invariance of order under multiplication by positive numbers. The relation one needs for infima then follows from $f \wedge g = -[(-f) \vee (-g)]$.}

Producing the absolute value of a given bounded real-valued function can be achieved by algebraic rather than order-theoretic constructions. The basic fact is the following proposition: although I state it in

[2] This fact has the useful consequence that a countably additive signed measure on (the Borel sets of) X is uniquely determined by its integrals against the convex functions; many uniqueness theorems for "representing measures" arise in a similar way.

the generality that the construction will give, one generally employs it only with continuous functions on a compact space.

Proposition: Let f be a bounded real-valued function on a set X. Then $|f|$ can be approximated arbitrarily closely uniformly on X by polynomials in f.

The basic idea is the following: there is no loss of generality in assuming that $\|f\|_\infty \le 1$, since one can always multiply f by a small positive constant α and deduce the result for f from the one for αf. If one writes

$$|f| = \sqrt{f^2} = \sqrt{1 + (f^2 - 1)} = \sum_{n=0}^{\infty} \binom{1/2}{n} (f^2 - 1)^n$$

and one knows that $\sum_{n=0}^{\infty} \left| \binom{1/2}{n} \right| < \infty$, then one can approximate $|f|$ uniformly as closely as one wishes—one simply takes a large partial sum of the series, because $\|f^2 - 1\|_\infty \le 1$ and thus

$$\| \sum_{n>N} \binom{1/2}{n} (f^2 - 1)^n \|_\infty \le \sum_{n>N}^{\infty} \left| \binom{1/2}{n} \right| \to 0 \quad \text{as } N \to \infty\ .$$

So to prove the proposition we need to investigate the convergence of this binomial-series-with-fractional-exponent. The most elementary approach to this problem involves two elementary results that everybody should know anyway.

Lemma [the ratio comparison test]: if $\sum a_n$ and $\sum b_n$ are two series of positive terms for which the quotient $(a_n/a_{n-1})/(b_n/b_{n-1})$ is ≤ 1 for all sufficiently large n, then $\sum b_n < \infty$ implies $\sum a_n < \infty$.

Proof of the lemma. If $(a_n/a_{n-1})/(b_n/b_{n-1}) \le 1$ holds for all $n > N$, then

$$\begin{aligned} a_{N+p} &= \frac{a_{N+p}}{a_{N+p-1}} \cdots \frac{a_{N+2}}{a_{N+1}} \cdot \frac{a_{N+1}}{a_N} \cdot a_N \\ &\le \frac{b_{N+p}}{b_{N+p-1}} \cdots \frac{b_{N+2}}{b_{N+1}} \cdot \frac{b_{N+1}}{b_N} \cdot a_N = \frac{a_N}{b_N}\, b_{N+p} \end{aligned}$$

so the terms of $\sum a_n$ are majorized by (a constant multiple of) the corresponding terms of the convergent series $\sum b_n$ for all sufficiently large indices.

The standard ratio test of freshman calculus is obtained by using $b_n = r^n$ for some nonnegative ratio $r < 1$, *i.e.*, using a convergent geometric series as a "comparison series." However, other choices are possible: a useful family of choices is given by the series $\sum_{n=1}^{\infty} 1/n^\beta$ for $\beta > 1$. As everybody learned in freshman calculus,[3] these series converge for (and only for) $\beta > 1$. The ratio of consecutive terms (with a tasteful choice of consecutive indices) for these series is[4]

$$\left(\frac{n-1}{n}\right)^\beta = \left(1 - \frac{1}{n}\right)^\beta = 1 - \frac{\beta}{n} + o\left(\frac{1}{n}\right) \qquad (*)$$

where as usual the symbol "$o(1/n)$" stands for a remainder term with the property that $o(1/n)/(1/n) \to 0$ as $n \to \infty$. The relation $(*)$ now produces a test for convergence that enables one to decide many interesting "inconclusive cases" of the freshman-calculus ratio test: the test is called Raabe's test.

Lemma [Raabe's test]: Let $\sum_{n=0}^{\infty} a_n$ be a series of positive terms. If the ratio of consecutive terms has the form

$$\frac{a_n}{a_{n-1}} = 1 - \frac{\alpha}{n} + o\left(\frac{1}{n}\right)$$

with $\alpha > 1$, then $\sum_{n=0}^{\infty} a_n < \infty$.

[3] One established this by using the integral test.

[4] To establish this, one needs to do nothing more complicated than apply the definition of the derivative at x=0 to the function $(1+x)^\beta$.

Proof of Raabe's test. Choose β with $\alpha > \beta > 1$. Then for sufficiently large n one has

$$\frac{a_n}{a_{n-1}} = 1 - \frac{\alpha}{n} + o\left(\frac{1}{n}\right) \leq 1 - \frac{\beta}{n} + o\left(\frac{1}{n}\right) = \left(\frac{n-1}{n}\right)^{\beta}$$

because the $o(1/n)$ terms are vanishingly small in comparison with the term $(\alpha - \beta)/n$. The series $\sum_{n=0}^{\infty} a_n$ thus converges by the ratio comparison test [compared with the series $\sum (1/n)^{\beta}$].

{Remark: Of course there is a "pessimistic" ratio comparison test too: if $\sum a_n = \infty$ then the same is true of $\sum b_n$. The fact that $\sum (1/n)^{\alpha}$ diverges for $\alpha < 1$ then implies that if the ratio of consecutive terms of a series (with positive terms) has the form $1 - \dfrac{\beta}{n} + o(1/n)$ with $\beta < 1$ then the series diverges. The case $\beta = 1$ is inconclusive, as the examples $\sum 1/(n \log n)$ and $\sum 1/(n \log^2 n)$ demonstrate. For our present purposes we are only interested in getting a series to converge.}

Lemma: The binomial series $(1+x)^{\alpha} = \displaystyle\sum_{n=0}^{\infty} \binom{\alpha}{n} x^n$ converges absolutely and uniformly for $|x| \leq 1$ when $\alpha \geq 0$.

Proof of the lemma. We can assume that $\alpha \notin \mathbb{N}$ since nonnegative integer powers of $(1+x)$ are polynomials. Then the binomial coefficients for $n > 0$ are

$$\binom{\alpha}{n} = \frac{\alpha(\alpha-1)\cdots(\alpha-(n-1))}{n!}$$

which are never zero, and the absolute value of the ratio of two consecutive coefficients is

$$\left|\frac{n-\alpha-1}{n}\right| = 1 - \frac{\alpha+1}{n}$$

provided n is sufficiently large (α is real). Raabe's test then gives us absolute convergence of the series of coefficients provided that $\alpha > 0$.[5]

Proof of the proposition. The series of coefficients in $\sqrt{1+(f^2-1)} = \displaystyle\sum_{n=0}^{\infty} \binom{1/2}{n} (f^2-1)^n$ has now been shown to converge absolutely, and that justifies the computation and approximation argument that we made for this series above.

This proposition now tells us that we can make vector lattices out of algebras in $\mathcal{C}(X, \mathbb{R})$ where X is compact (Hausdorff, or a compact metric space). Recall that a subset $\mathcal{A} \subseteq \mathcal{C}(X, \mathbb{R})$ is a **subalgebra** (**with identity**) if it is a vector subspace that is closed under multiplication (of its own elements and by scalars in $\mathbb{R}$) and that also contains 1 (and thus contains all constant functions). The following proposition is easily checked, and the reader should check it:

Proposition: If $\mathcal{A} \subseteq \mathcal{C}(X, \mathbb{R})$ is a subalgebra, then so is its uniform closure $\overline{\mathcal{A}}$ (*i.e.*, its closure in the uniform-norm metric).

From what we did with the binomial series, it follows that

Proposition: If $\mathcal{A} \subseteq \mathcal{C}(X, \mathbb{R})$ is a subalgebra, then its uniform closure $\overline{\mathcal{A}}$ is an algebra *and* a vector lattice.

Theorem [Stone-Weierstraß Theorem for algebras]: Let X be a compact (Hausdorff, or metric) space and $\mathcal{A} \subseteq \mathcal{C}(X, \mathbb{R})$ a subalgebra of the continuous real-valued functions that contains the constants. Then its closure $\overline{\mathcal{A}}$ equals $\mathcal{C}(X, \mathbb{R})$—*i.e.*, $\mathcal{A}$ is dense in $\mathcal{C}(X, \mathbb{R})$—if and only if for every pair of distinct points x, $y \in X$

[5] Loosely: if the series "ought to converge" because the function makes sense at both endpoints, then it does converge. Never tell freshman calculus students anything like this, of course: it only encourages them.

there is some $f \in \mathcal{A}$ for which $f(x) \neq f(y)$. (It is usually said that $\mathcal{A}$ **separates points** of X or that $\mathcal{A}$ is a **separating subalgebra** of $\mathcal{C}(X, \mathbb{R})$.)

Proof. Suppose $\mathcal{A}$ separates points of X and contains the constant functions. Then $\mathcal{A}$ automatically has the two-point interpolation property that the lattice version of Stone's theorem requires, because if $x \neq y$ in X, $f(x) \neq f(y)$, and β_1 and β_2 are given, then the system of equations in the unknowns α_1 and α_2 $\alpha_1 \cdot 1 + \alpha_2 f(x) = \beta_1$, $\alpha_1 \cdot 1 + \alpha_2 f(y) = \beta_2$ has determinant $\det \begin{pmatrix} 1 & f(x) \\ 1 & f(y) \end{pmatrix} = f(y) - f(x) \neq 0$ and thus possesses a solution. Since $\overline{\mathcal{A}}$ is a closed subalgebra, it contains with each of its elements f the function $|f|$; it follows that $\overline{\mathcal{A}}$ is a vector lattice with the two-point interpolation property. $\overline{\mathcal{A}}$ is therefore dense in $\mathcal{C}(X, \mathbb{R})$; but since it is closed, $\overline{\mathcal{A}} = \mathcal{C}(X, \mathbb{R})$, which says exactly that the original $\mathcal{A}$ was dense in $\mathcal{C}(X, \mathbb{R})$. The converse is easy, since $\mathcal{C}(X, \mathbb{R})$ separates points of X (use the metric, or in the compact-Hausdorff-space setting, use Uryson's lemma).

3. *Specializations (applications) to standard settings.*

The first thing is the original Weierstraß approximation theorem.

Theorem [K. Weierstraß]: Any continuous real-valued function on a real interval $[a, b]$ can be approximated uniformly arbitrarily closely by a polynomial function.

Proof. The algebra of polynomial functions (with real coefficients) on $[a, b]$ contains the constants, and the identity function "x" separates points. Apply Stone-Weierstraß for algebras.

More generally, if $X \subseteq \mathbb{R}^n$ is a compact set, then the polynomial functions of the coördinates—*e.g.*, in two dimensions the polynomials $(x, y) \mapsto \sum a_{jk} x^j y^k$— form an algebra that separates points and contains the constants, so every continuous function on X can be approximated uniformly by polynomials as closely as one wishes. No topological hypotheses other than compactness need be placed on X.

Of course, there are unheard-of numbers of proofs of the Weierstraß approximation theorem in its original form. A proof via classical Fourier-series convergence theorems can be found on p. 238 of Wheeden & Zygmund. The **trigonometric polynomials**, *i.e.*, functions of the form $\theta \mapsto a_0 + \sum_{k=1}^{n} \{a_k \cos k\theta + b_k \sin k\theta\}$ form a separating subalgebra of $\mathcal{C}(\mathbb{T}, \mathbb{R})$ where $\mathbb{T}$ is the unit circle in the plane; so Stone-Weierstraß implies that these are dense in $\mathcal{C}(\mathbb{T}, \mathbb{R})$, which is what the classical Fourier-series convergence theorem shows. There are recent results[6] of the type of the Weierstraß approximation theorem; a sample is

Proposition: Let $[-a, 1] \subseteq \mathbb{R}$ be a closed real interval, with $a > 1$ (so the interval extends to the left of -1). Then any continuous real-valued f on $[-a, 1]$ that vanishes on $[-1, 1]$ can be approximated uniformly arbitrarily closely by polynomial functions $x \mapsto \sum_k a_k x^k$ all of whose coefficients $a_k \geq 0$.

One can prove the Tietze extension theorem, at least in the compact-metric-space case, by using the Stone-Weierstraß theorem:

Proposition [Tietze extension theorem for compact metric spaces]: Let Y be a closed subset of the compact (metric) space (X, d). Then every element of $\mathcal{C}(Y, \mathbb{R})$ is the restriction of an element of $\mathcal{C}(X, \mathbb{R})$ of equal uniform norm.

Proof. The algebra-with-1 of restrictions $\mathcal{C}(X, \mathbb{R})|Y$ to the little space of continuous functions on the big space contains the constants and separates points (use the distance function to give examples—in the compact Hausdorff space setting, use Uryson's lemma), so it is dense in $\mathcal{C}(Y, \mathbb{R})$. Let $\varphi \in \mathcal{C}(Y, \mathbb{R})$ be given. By density of $\mathcal{C}(X, \mathbb{R})|Y$ in $\mathcal{C}(Y, \mathbb{R})$ we can find $f_1 \in \mathcal{C}(X, \mathbb{R})$ with $\|\varphi - f_1\|_{Y,\infty} < (1/2)\|\varphi\|_{Y,\infty}$, and without loss of generality one can assume that $\|f_1\|_{X,\infty} \leq \|\varphi\|_{Y,\infty}$ because replacing f_1 by $(f_1 \wedge \|\varphi\|_{Y,\infty}) \vee (-\|\varphi\|_{Y,\infty})$ will only bring it closer to φ (the reader should check this). One can repeat this process inductively, generating a sequence $\{f_k\}_{k=1}^{\infty}$ of elements of $\mathcal{C}(X, \mathbb{R})$ with $\|\varphi - f_1 - \cdots - f_k\|_{Y,\infty} < (1/2)^k \|\varphi\|_{Y,\infty}$ and

[6] J. Toland, *Selfadjoint operators and cones*, J. London Math. Soc. (2) 53 (1996), 167–183; R. D. Nussbaum & B. Walsh, *Approximation by polynomials with nonnegative coefficients ...* , Trans. Amer. Math. Soc. 350 (1998), 2367–2391.

$\|f_k\|_{X,\infty} \leq \|\varphi - f_1 - \cdots - f_{k-1}\|_{Y,\infty} < (1/2)^{k-1}\|\varphi\|_{Y,\infty}$. Evidently $\sum_k \|f_k\|_{X,\infty} < \infty$, so the function $f = \sum_k f_k$ is a well-defined element of $\mathcal{C}(X, \mathbb{R})$. It is obvious from the construction that $f|Y = \varphi$; replacing f by $g = (f \wedge \|\varphi\|_{Y,\infty}) \vee (-\|\varphi\|_{Y,\infty})$ will only bring it closer to φ, so $g \in \mathcal{C}(X, \mathbb{R})$ has the same uniform norm as φ and $g|Y = \varphi$, as required.