# THE (STONE-)WEIERSTRASS APPROXIMATION THEOREM

PETE L. CLARK

You are **not** responsible for any of this material on your final exam.

## 1. The statement of the theorem

**Theorem 1.** *(Weierstrass Approximation Theorem) Let $f : [a, b] \to \mathbb{R}$ be a continuous function and $\epsilon > 0$ be any positive number. Then there exists a polynomial $P = P(\epsilon)$ such that for all $x \in [a, b]$, $|f(x) - P(x)| < \epsilon$. In other words, any continuous function defined on a closed interval is the uniform limit of a sequence of polynomials.*

Remark: As we have seen in the homework problems, the result is false when $[a, b]$ is replaced by $\mathbb{R}$: a sequence of polynomials defined on the entire real line can only be uniformly convergent if it is ultimately constant.

It is interesting to compare this result with Taylor's theorem, which gives conditions for a function to be equal to its Taylor series. Note that any such function must be $C^\infty$ (i.e., it must have derivatives of all orders), whereas in the Weierstrass Approximation Theorem we can get *any* continuous function. An important difference is that the Taylor polynomials $T_N(x)$ have the property that $T_{N+1}(x) = T_N(x) + a_{N+1}x^n$, so that in passing from one Taylor polynomial to the next, we are not changing any of the coefficients from 0 to $N$ but only adding a higher order term. In contrast, for the sequence of polynomials $P_n(x)$ uniformly converging to $f$ in Theorem 1, $P_{n+1}(x)$ is not required to have any simple algebraic relationship to $P_n(x)$.

The Weierstrass Approximation Theorem is quite deep: in other words, a very careful understanding of the statement does not in any way suggest a proof. It might be useful to compare it to the following result, which sounds similar but is much easier.

Definition: We say a function $f : [a, b] \to \mathbb{R}$ is **polygonal** if it is a continuous function made up out of finitely many straight line segments. More formally, there exists a partition $P = \{a = x_0 < x_1 \ldots < x_n = b\}$ such that for $1 \leq i \leq n$, the restriction of $f$ to $[x_{i-1}, x_i]$ is a linear function. For instance, the absolute value function is polygonal.

It is no problem to approximate continuous functions by polygonal functions:

**Proposition 2.** *("Polygonal approximation theorem") Let $f : [a, b] \to \mathbb{R}$ be a continuous function and $\epsilon > 0$ be any positive number. Then there exists a polygonal function $P$, depending upon $\epsilon$, such that for all $x \in [a, b]$, $|f(x) - P(x)| < \epsilon$.*

Proof: The obvious interpolation argument works. Namely, since any continuous function on $[a,b]$ is uniformly continuous, there exists a $\delta > 0$ such that whenever $|x-y| < \delta$, $|f(x)-f(y)| < \epsilon$. Choose $n$ large enough so that $\frac{b-a}{n} < \delta$, and consider the uniform partition $P_n = \{a < a + \frac{b-a}{n} < \ldots < b\}$ of $[a,b]$. There is a polygonal function $P$ such that $P(x_i) = f(x_i)$ for all $x_i$ in the partition, and is defined in between by "connecting the dots" (more formally, by linear interpolation). For any $i$ and for all $x \in [x_{i-1}, x_i]$, we have $f(x) \in (f(x_{i-1})-\epsilon, f(x_{i-1})+\epsilon)$. The same is true for $P(x)$ at the endpoints, and one of the nice properties of a linear function is that it is either increasing or decreasing, so its values are always in between its values at the endpoints. Thus for all $x$ in $[x_{i-1}, x_i]$ we have $P(x)$ and $f(x)$ both lie in an interval of length $2\epsilon$, so it follows that $|P(x)-f(x)| < 2\epsilon$ for all $x \in [a,b]$. Good enough.

The proof we are going to give of Theorem 1 is different from the proof in your text. We will actually prove a more general result, but the extra generality makes some of the proof more conceptual.

We need some preliminary definitions. First, consider that the class of polynomial functions $\mathcal{P}[a,b]$ on the closed interval $[a,b]$ has a certain algebraic structure: namely, if $P_1$ and $P_2$ are polynomials and $\alpha$ is any real number, $P_1 + P_2$, $\alpha P_1$ and $P_1 \cdot P_2$ are all polynomials. The first two properties express the fact that polynomials form a real vector space. But the third property says that the product of two polynomials is also a polynomial; a real vector space endowed with a product operation is called a **real algebra**.

Let $\mathcal{A}$ be any algebra of bounded functions on $[a,b]$. We can consider the collection of all functions $f : [a,b]$ which are uniform limits of sequences $\{f_i\}$, where each $f_i \in f$. Of course each element $f$ of $\mathcal{A}$ is a uniform limit of a sequence of functions in $\mathcal{A}$ – just take the constant sequence $f_i = f$ for all $i$. We aspire to get more however; recall that what we want to show is that every continuous function arises as the uniform limit of a sequence of functions in $\mathcal{P}[a,b]$. So why not consider more generally the set $\overline{\mathcal{A}}$ of all uniform limits of sequences of functions in our algebra $\mathcal{A}$? The following result gives some indication that this might be a sensible thing to do.

**Proposition 3.** *For any algebra $\mathcal{A}$ of bounded functions on $[a,b]$, $\overline{\mathcal{A}}$ is also an algebra of bounded functions on $[a,b]$.*

Proof: We know that a uniform limit of bounded functions is bounded, so any element of $\overline{\mathcal{A}}$ is bounded. Similarly, if $f_n \stackrel{u}{\rightarrow} f$ and $g_n \stackrel{u}{\rightarrow} g$, we showed that $f_n + g_n \stackrel{u}{\rightarrow} f+g$; we also showed that if $f$ and $g$ are bounded, then $f_n g_n \stackrel{u}{\rightarrow} fg$. It is quite easy to show that if $f_n \stackrel{u}{\rightarrow} f$ and $\alpha$ is any real number, then $\alpha f_n \stackrel{u}{\rightarrow} \alpha f$. Thus if $f,\ g \in \overline{\mathcal{A}}$ then $f+g$, $\alpha f$ and $fg$ are all in $\overline{\mathcal{A}}$, so $\overline{\mathcal{A}}$ forms an algebra of bounded functions.

We call $\overline{\mathcal{A}}$ the **uniform closure** of $\mathcal{A}$. Thus we can restate the Weierstrass Approximation Theorem as: show that the uniform closure of the algebra of polynomial functions on $[a,b]$ is the algebra of all continuous functions on $[a,b]$.

Consider two more properties of algebras of functions on $[a,b]$. Namely, we say that an algebra $\mathcal{A}$ **vanishes at no point of** $[a,b]$ if for all $x \in [a,b]$, there exists some $f \in \mathcal{A}$ such that $f(x) \neq 0$. Also we say that an algebra $\mathcal{A}$ **separates points** if for any distinct points $x_1 \neq x_2$ of $[a,b]$, there exists a function $f \in \mathcal{A}$ such that $f(x_1) \neq f(x_2)$.

Note that it is completely obvious that $\mathcal{P}[a,b]$ satisfies these two properties. Indeed, the collection of all linear functions has these properties.

**Lemma 4.** *Suppose $\mathcal{A}$ is an algebra of functions on $[a,b]$ which separates points and vanishes at no point of $[a,b]$. Then for distinct points $x_1 \neq x_2$ and any two real numbers $c_1$ and $c_2$, there exists $f \in \mathcal{A}$ such that $f(x_1) = c_1$ and $f(x_2) = c_2$.*

Proof: By assumption, $\mathcal{A}$ contains functions $g$, $h$, $k$ such that $g(x_1) \neq g(x_2)$, $h(x_1) \neq 0$ and $k(x_2) \neq 0$. Put

$$u = gk - g(x_1)k, \ \ v = gh - g(x_2)h.$$

Then $u$ and $v$ are elements of $\mathcal{A}$ such that $u(x_1) = v(x_2) = 0$ and $u(x_2), v(x_1)$ are both nonzero. It follows that

$$f = \frac{c_1 v}{v(x_1)} + \frac{c_2 u}{u(x_2)}$$

has $f(x_1) = c_1$ and $f(x_2) = c_2$.

We are now ready for the "abstract" version of the Weierstrass Approximation Theorem that we will prove.

**Theorem 5.** *(Stone-Weierstrass Theorem) Let $\mathcal{A}$ be an algebra of continuous functions on $[a,b]$. Suppose that $\mathcal{A}$ separates points and vanishes at no point of $[a,b]$. Then $\overline{\mathcal{A}} = \mathcal{C}[a,b]$, the algebra of all continuous functions.*

Remarks: The separation and nonvanishing conditions are *necessary* as well as sufficient for the uniform closure to contain all the continuous functions. For instance, if every $f \in \mathcal{A}$ had $f(x_0) = 0$, every element $f \in \overline{\mathcal{A}}$ will also have $f(x_0) = 0$, because the limit (even a pointwise limit, let alone a uniform limit) of a sequence of functions which are equal to zero at a point will also be equal to zero at that point. A similar comment holds for separation of points.

What is interesting about this theorem is that it somehow "explains" why every continuous function can be approximated by polynomials. Also it gives many generalizations:

Exercise 81*: Consider the set $\mathcal{P}_2$ of functions on $[-1,1]$ which are of the form $P(x^2)$ where $P(x)$ is a polynomial, and similarly the set $\mathcal{P}_3$ of functions of the form $P(x^3)$. Show that $\mathcal{P}_2$ and $\mathcal{P}_3$ are both algebras of functions. Exactly one of them has as its uniform closure the set of all continuous functions on $[a,b]$; what is the uniform closure of the other?

## 2. Proof of the Stone-Weierstass Theorem

We need first a technical fact about open intervals covering a closed interval. Supppose that we have a collection $\{I_i\}$ of open intervals $I_i = (a_i, b_i)$ and a closed interval $[a,b]$ such that every point $x$ of $[a,b]$ lies in at least one $I_i$. (We say the intervals $I_i$ **cover** $[a, b]$.) The following result says that we can throw away all but finitely many of the intervals and still cover all of $[a, b]$.

**Lemma 6.** *(Compactness Lemma) Suppose that* $\{I_i\}$ *is a covering of* $[a, b]$ *by open intervals. Then there exists a finite set of indices* $i_1, \ldots, i_N$ *such that*

$$[a, b] \subset I_{i_1} \cup \ldots \cup I_{i_n}.$$

Proof: Let us consider the set $S$ of $c \in [a, b]$ for which at least the subinterval $[a, c]$ can be covered many finitely many of the open intervals $I_i$. This set is nonempty because $[a, a] = \{a\}$ is in it: by hypothesis, $a$ lies in at least one subinterval $I_i$. Let $C$ be the supremum of $S$. It is not possible that $C = a$, because the interval $I_i$ which contains $a$, being open, cannot have $a$ as a right endpoint – it must extend at least a little bit farther. We claim that $C$ is an element of $S$, i.e., that if it is possible to cover $[a, C - \epsilon]$ by finitely many $I_i$'s for all $\epsilon > 0$, it is also possible to cover $[a, C]$ by finitely many $I_i$'s. We can see this because $C$ must be covered by some interval $I_{i_0}$, which, being open, again contains $[C - \delta, C]$ for some $\delta > 0$. Choosing $\epsilon < \delta$, we find that by adding if necessary one more subinterval $I_{i_0}$, we can cover $[a, C - \epsilon]$ together with $I_N$ (finitely many $I_i$'s plus $I_N$ still makes only finitely many), so we get $C$ as well. And again, once we have covered $C$ by an open interval, we must get values at least a little bit larger than $C$, unless $C = b$. This completes the proof of the lemma.

Remark: If we want the intervals $I_i$ to be contained in $[a, b]$, we need to allow them to contain $a$ as a left endpoint or $b$ as a right endpoint: e.g., for the purposes of the lemma we need to regard $[a, c)$ and $(c, b]$ as "open subintervals of $[a, b]$." With this convention, the proof still works.

Ironically, to prove the theorem, we need to know the following special case of the Weierstrass Approximation Theorem which we have mentioned already:

**Proposition 7.** *For any* $a > 0$*, the function* $f(x) = |x|$ *on* $[-a, a]$ *can be uniformly approximated by polynomials.*

It seems instructive to first prove the Stone-Weierstrass Theorem assuming Proposition 7 and then come back to prove Proposition 7: we'll see that its proof is about as difficult as the proof of the rest of the theorem!

Anyway, suppose that we have an algebra $\mathcal{A}$ of continuous functions $f : [a, b] \to \mathbb{R}$ which separates points and vanishes identically at no point. We want to show that any continuous function is the uniform limit of a sequence of functions in $\mathcal{A}$.

Step 1: If $f \in \mathcal{A}$, then $|f| \in \mathcal{A}$.

Proof: Take $|a| = ||f||$. Fix $\epsilon > 0$. Proposition guarantees that there exists a polynomial $P(x) = \sum_{i=1}^{n} c_i x^n$ such that

$$||P(x) - |x||| < \epsilon. \tag{1}$$

Since $\mathcal{A}$ is an algebra, the function $g = P \circ f = \sum_{i=1}^{n} c_i f^n$ is also in $\mathcal{A}$, and since $|f(x)| \leq a$ for all $x \in [a, b]$, replacing $x$ with $y = f(x)$ in (1), we get

$$|P(y) - |y|| = |g(x) - |f(x)|| < \epsilon$$

for all $x \in [a,b]$, so that $|f|$ can be uniformly approximated by elements of $\mathcal{A}$.

Step 2: If $f, g \in \mathcal{A}$, then so are the functions $\max(f,g)$ and $\min(f,g)$. (E.g., the definition of the first function is that for any given $x$, $x \mapsto \max(f(x), g(x))$.)

Proof: This follows immediately from Step 1, using the formulas

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

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

which you'll want to check if you've never seen them before.

Remark: By a trivial induction argument, we can in fact take maxes or mins of any finite number of functions and stay inside $\overline{\mathcal{A}}$.

Step 3: Given any continuous function $f : [a,b] \to \mathbb{R}$, a point $x \in [a,b]$ and an $\epsilon > 0$, there exists a function $g_x \in \overline{\mathcal{A}}$ such that $g_x(x) = f(x)$ and $g_x(t) > f(t) - \epsilon$ for all $t \in [a,b]$.[1]

Proof: Since $\mathcal{A}$ satisfies the hypotheses of Lemma 6 by assumption and $\mathcal{A} \subset \overline{\mathcal{A}}$, then it follows that $\overline{\mathcal{A}}$ satisfies the hypotheses as well. Thus for any $y \in [a,b]$ we can find a function $h_y \in \overline{\mathcal{A}}$ such that $h_y(x) = f(x)$ and $h_y(y) = f(y)$. Because $h_y$ is continuous, there exists an open subinterval[2] $I_y$ containing $y$ such that $h_y(t) > f(t) - \epsilon$ for all $t \in I_y$. Now we have a collection $\{I_y\}$ of open subintervals of $[a,b]$ which covers $[a,b]$ (because for each $y \in [a,b]$ we have one subinterval guaranteed to contain $y$, namely $I_y$), so by Lemma 6 we know that there exists a finite set $\{y_1, \ldots, y_n\}$ such that every $x \in [a,b]$ lies in $I_{y_1} \cup \ldots \cup I_{y_n}$. Take $g_x = \max(h_{y_1}, \ldots, h_{y_n})$. Then $g_x \in \overline{\mathcal{A}}$ and by construction it has the required property.

Step 4: Finally, by a similar procedure we can get the $g_x$'s not to lie too much above $f$. Namely, for any $x$, continuity of $g_x$ means that there exists an open interval $I_x$ containing $x$ such that $g_x(t) < f(t) + \epsilon$ for all $t \in I_x$. Again, since the $\{I_x\}$ form a covering of $[a,b]$ by open subintervals, there exitss a finite set $\{x_1, \ldots x_n\}$ such that $[a,b] \subset I_{x_1} \cup \ldots \cup I_{x_n}$. As in Step 3, taking $h = \min(g_{x_1}, \ldots, g_{x_n})$, we get at last that $h(t) > f(t) - \epsilon$ and $h(t) < f(t) + \epsilon$, both for all $t \in [a,b]$. In other words, we get $||h - f|| < \epsilon$, proving the result.

But wait: we must still prove Proposition 6, namely that $|x|$ can be a approximated uniformly by polynomials on $[-a,a]$. We will take $a = 1$, leaving the general case as an exercise (for instance, using the linear function $L(x) = ax$, one could just rescale everything to get from $[-1,1]$ to $[-a,a]$). In this case we will define a seuqnece of polynomials $P_n(x)$ by the following recursion: $P_0 = 0$, and for all

[1] In other words, we are constructing a function which cannot dip too far below $f$; this is half of what we want to show. and the other half is accomplished in the next step.

[2] When $y = a$ we are getting an interval of the form $[y, c)$ and when $y = b$ we are getting an interval of the form $(c, y]$.

$n \geq 0$,

$$P_{n+1}(x) = P_n(x) + \frac{x^2 - P_n(x)^2}{2}.$$

From this definition it is not hard to check the following identity:

$$|x| - P_{n+1}(x) = (|x| - P_n(x)) \left(1 - \frac{|x| + P_n(x)}{2}\right). \tag{2}$$

One can show by induction that $0 \leq P_n(x) \leq P_{n+1} \leq |x|$ for all $x \in [-1, 1]$. Briefly, assuming that $P_n(x) \leq |x|$ and using the recursion, one sees that $P_{n+1} - P_n(x) = \frac{x^2 - P_n(x)^2}{2}$, and this is non-negative since $P_n(x) \leq |x|$ (part of the induction hypothesis). Similarly, one shows that $P_{n+1}(x) \leq |x|$ by considering signs in (2); the first factor in the righthandside is non-negative by the induction hypothesis, and the non-negativity of the second factor is equivalent to $|x| + P_n(x) \leq 2$, which is true because both terms are at most 1. The nonnegativity can be checked by setting $P_{n+1}(x) = \frac{-P_n(x)^2 + 2P_n(x) + x^2}{2}$ equal to zero and observing that there are no roots in the interior of $[-1, 1]$.

Using these inequalities, one can derive

$$|x| - P_n(x) \leq |x| \left(1 - \frac{|x|}{2}\right)^n < \frac{2}{n+1}$$

where the last inequality follows from using good old differential calculus to maximze the function $x(1 - \frac{x}{2})^n$ on $[0, 1]$. Since the difference between $|x|$ and $P_n(x)$ converges to zero via a bound that is independent of $x$, this shows that $P_n(x) \xrightarrow{u} |x|$.

This completes the proof of the Stone-Weierstrass Theorem!