CONVOLUTIONS AND THE WEIERSTRASS APPROXIMATION THEOREM

MATTHEW BOND

ABSTRACT. The famous **Weierstrass approximation theorem** states that any continuous function $f : [0,1] \to \mathbb{R}$ can be approximated by a polynomial with a maximum error as small as one likes. There are several approaches to proving this theorem, but in this paper, we present a proof which is essentially that of Walter Rudin's **Principles of Mathematical Analysis**. It is my opinion that the presentation of this proof could have benefited from a discussion of its main tool, the **convolution** of two functions, which plays an important role in Harmonic Analysis, PDEs, and Probability Theory, just to name a few. In particular, the convolution with an **approximate identity** is the important case on which this theorem is based.

1. Introduction

We wish to prove the following:

Theorem 1. Weierstrass approximation theorem

Let $f : [0,1] \to \mathbb{R}$ be a continuous function. Then for each $\varepsilon > 0$, there exists a polynomial function P such that for all $x \in [0,1]$, $|f(x) - P(x)| < \varepsilon$.

Equivalently, for any such f, there exists a sequence P_n of polynomials such that $P_n \to f$ uniformly on $[0,1]$.

First of all, let us say what this theorem does NOT say: first of all, it does not say that every smooth (i.e., infinitely differentiable) function is equal to its Taylor series (see, for example, the counterexample $f(x) = e^{-\frac{1}{x^2}}$, whose Taylor series is well known to vanish at the origin, by direct computation by applying L'Hopital's rule to the function and its derivatives, for example). The theorem also **does not** say that every continuous function is given by a power series, and that the partial sums define P_n. We will be free to allow all the coefficients to vary as n increases.

The proof is fairly easy to remember in outline, once one has mastery over its ingredients, all of which prove to be indispensible tools as one goes farther in the study of analysis:

1) We can subtract off a linear term and assume that $f(0) = f(1) = 0$.

2) If one takes the convolution of a continuous function on $[0, 1]$ with an approximate identity, then the result is a uniformly good approximation.

3) If the approximate identity is a polynomial restricted to the domain $[-1, 1]$, then the resulting convolution is a polynomial.

Thus it suffices to construct an approximate identity as described in (3). We will define and justify (2) and (3) below, but for one who is already familiar with the terms, the insights of (2) and (3) should be enough to figure out a proof without further severe difficulties.

2. Convolution by an approximate identity

Let $f, g : \mathbb{R} \to \mathbb{R}$. Whenever the following integral is well-defined[1], let the **convolution** of f and g, $f * g$, be defined by

$$(f * g)(x) := \int_{\mathbb{R}} f(x - t)g(t)dt.$$

The convolution operator is commutative and associative[2]. It is hopeless to look for anything like an inverse under convolution, since in some sense convolution by g takes the values of f and dilutes them by a weighted averaging process corresponding to a distribution shaped like g. It is NOT even true that there is a function which plays the role of an identity element under convolution, but there are **approximate identities**, sequences of functions g_n such that for each f, $f * g_n \to f$ in some sense.

Recall that for a set E, the **characteristic function** χ_E is defined as

$$\chi_E(x) = \begin{cases} 1, x \in E \\ 0, x \notin E \end{cases}.$$

Our first example of an approximate identity is

$$g_n := \frac{n}{2}\chi_{[-\frac{1}{n}, \frac{1}{n}]}.$$

Theorem 2. *Let f be a continuous function which vanishes outside of $[0, 1]$. Then $\forall x \in [0, 1]$, $g_n * f \to f$ uniformly as $n \to \infty$.*

---

[1]For example, convolutions are well-defined for almost every x if f and g belong to L^1, the space of all functions h such that $\int_{\mathbb{R}} |h(x)|dx < \infty$.

[2]You can prove this directly from the definition!

Proof.

$$f * g_n(x) = \int_{\mathbb{R}} g_n(t)f(x-t)dt = \frac{n}{2}\int_{-\frac{1}{n}}^{\frac{1}{n}} f(x-t)dt \qquad (2.1)$$

$$f * g_n(x) = \frac{n}{2}\int_{x-\frac{1}{n}}^{x+\frac{1}{n}} f(t)dt. \qquad (2.2)$$

So convolution by g_n acts on f at each point by averaging f over a $\frac{1}{n}$-neighborhood, and we expect such averages of a continuous function f to converge to f. Indeed, since $[0,1]$ is compact, f is uniformly continuous. Let us compare the difference:

$$(f * g_n)(x) - f(x) = \frac{n}{2}\int_{x-\frac{1}{n}}^{x+\frac{1}{n}} (f(t) - f(x))dt. \qquad (2.3)$$

We used that $f(x)$ does not depend on t. Note that $|t - x| < \frac{1}{n}$ in the domain of integration. Now let $\varepsilon > 0$. Use the uniform continuity of f. We may choose n such that $\forall x,\ |f(t) - f(x)| < \varepsilon$. So 2.3 gives us

$$|(f * g_n)(x) - f(x)| \leq \frac{n}{2}\int_{x-\frac{1}{n}}^{x+\frac{1}{n}} |f(t) - f(x)|dt \leq \frac{n}{2}\cdot\frac{2}{n}\cdot\varepsilon = \varepsilon. \qquad (2.4)$$

$\square$

The above was a sandbox example of the idea behind approximate identities. A general approximate identity will contain MOST of its mass on a small interval, so while we won't be able to throw away everything outside of a small interval, we WILL be able to control its contribution, as well. The idea is that instead of taking the usual average over a small interval, convolution by a general approximate identity will take a weighted average that successively concentrates more and more of its mass nearer and nearer to the origin as we increase the parameter n.

Definition:

An **approximate identity** g_n is a sequence of functions with the following properties[3]:

a) $\int_{\mathbb{R}} g_n = 1$

b) $\int_{\mathbb{R}} |g_n| < M$, for some constant M which does depend on n.

c) $\forall \delta, \varepsilon > 0,\ \exists N\colon |g_n(x)| < \varepsilon\ \forall x \geq \delta$ and $\forall n \geq N$.

d) $g_n(x) = 0$ whenever $|x| > 1$.

---

[3]Definitions may vary. If g is required to be positive, for example, we may do away with (b). (c) and (d) may be replaced by, and in fact imply, the condition (for all $\delta > 0$) $\int_{|t|>\delta} |g_n(t)|dt \to 0$ as $n \to \infty$

Then

Theorem 3. *Let g_n be any approximate identity. Then Theorem 2 is still true for g_n.*

Proof. Let $\varepsilon > 0$. We want $|(f * g_n)(x) - f(x)| < \varepsilon$.

$$(f * g_n)(x) - f(x) = \int_{\mathbb{R}} [f(x - t)g_n(t) - f(x)g_n(t)]dt,$$

by the definition of convolution, property (a), and the fact that $f(x)$ is constant with respect to t. Let us split up the above and take absolute values:

$$|(f * g_n)(x) - f(x)| \leq \int_{-\delta}^{\delta} |f(x - t) - f(x)||g_n(t)|dt + \int_{|t| > \delta} |f(x - t) - f(x)||g_n(t)|dt.$$

First, choose $\delta > 0$ so that $|f(x) - f(x - t)| < \frac{\varepsilon}{2M}$ whenever $|t| < \delta$. $\forall x$, $|f(x)| \leq C$ some constant C. So for the chosen δ, let N be such that $|g_n(t)| < \frac{\varepsilon}{8C}$ $\forall |t| > \delta$ when $n \geq N$. So we get:

$$|(f * g_n)(x) - f(x)| \leq \frac{\varepsilon}{2M} \int_{-\delta}^{\delta} |g_n(t)|dt + 2C \int_{|t| > \delta} |g_n(t)|dt < \varepsilon.$$

$\square$

3. **A polynomial approximate identity**

First, one should verify that if Q is a polynomial restricted to $[-1, 1]$ and f is a continuous function on $[0, 1]$ vanishing at the endpoints, then $f * Q$ is again a polynomial. Indeed, since convolution is commutative, we may write

$$f * Q(x) = Q * f(x) = \int_{\mathbb{R}} Q(x - t)f(t)dt.$$

One can write out $Q(x - t)$, take careful consideration of the domains of f and Q, and verify that one gets an integral (with respect to t over all of $\mathbb{R}$) of a sum of terms of the type $c_j x^k t^l f(t)$. If one supplies the details, then indeed we get

Theorem 4. *Let $f : [0, 1] \to \mathbb{R}$ be continuous and vanishing at its endpoints, and let Q be a polynomial restricted to $[-1, 1]$. Then $f * Q$ is a polynomial function.*

So to prove Theorem 1, all we need is Theorems 3, 4 and the following Theorem 5:

Theorem 5. *Let $Q_n(x) = C_n(1 - x^2)^n$ be functions restricted to $[-1, 1]$, with C_n chosen to make Q_n satisfy (a). Then Q_n is an approximate identity.*

Proof. (b) follows because Q_n is positive, and (d) follows by definition. We need only prove (c).

In fact,

$$\frac{1}{C_n} = 2 \int_0^1 (1 - x^2)^n dx.$$

Let $\delta > 0$. Then

$$\frac{1}{C_n} = 2 \int_0^1 (1 - x^2)^n dx \geq 2 \int_0^{\delta/2} (1 - (\frac{\delta}{2})^2)^n dx \geq \delta(1 - (\frac{\delta}{2})^2)^n \geq \delta \alpha^n (1 - \delta^2)^n,$$

where $\alpha > 1$ depends only on δ ($\alpha = \frac{1 - (\frac{\delta}{2})^2}{1 - \delta^2}$ works).

So whenever $|x| > \delta$, we get

$$Q_n(x) = C_n(1 - x^2)^n \leq C_n \cdot (1 - \delta^2)^n \leq \frac{1}{\delta \alpha^n},$$

which goes to 0 as $n \to \infty$, uniformly in x for all x outside of $(-\delta, \delta)$.

$\square$

To be quite explicit, here is one way to get polynomial functions $P_n \to f$ uniformly on $[0, 1]$ as $n \to \infty$:

First, let $\tilde{f}(x) = f(x) - (1 - x)f(0) - xf(1)$ (i.e., subtract off the equation of the line through $(0, f(0))$ and $(1, f(1))$). Because the modified function $\tilde{f}$ is contionuous on $[0, 1]$ and vanishes at the endpoints, it is approximable on $[0, 1]$ by the convolution method.

$\tilde{P}_n(x) = \tilde{f} * Q_n(x)$, where Q_n is defined exactly as in Theorem 5. $\tilde{P}_n(x) \to \tilde{f}(x)$ uniformly in x as $n \to \infty$ by Theorems 3 and 5. So let $P_n(x) = \tilde{P}(x) + (1 - x)f(0) + xf(1)$. Then P_n satisfies the conclusion of Theorem 1, concluding the proof. $\square$

4. REMARKS

The Weierstrass theorem generalizes considerably; see Walter Rudin's **Principles of Mathematical Analysis** for a relatively constructive approach to a generalization he refers to as the Stone-Weierstrass Theorem.

Convolutions have many other applications. For example, the probability density function of a sum of continuous, independent random variables is given by the convolution of their probability density functions.

In PDEs, the heat equation, among others, is given by a certain convolution of the intial values f by an approximate unit g_t. The time t plays the role of n, and in this case, $u = f * g_t$ converges to the initial values f as the time t decreases toward the initial time, $t = 0$. Then as t increases, the heat disperses as g_t widens and flattens.

Convolutions are indispensible to the theory of Harmonic Analysis, as well. A certain convolution along the unit circle allows one to approximate a function $f : \mathbb{S}^1 \to \mathbb{C}$ by exponential polynomials.

If time permits, a slightly longer brief but shallow discussion of these applications will be added to this paper for flavor.

A careful treatment of the well-definedness of convolutions requires a bit of the Lebesgue theory of integration, especially when f isn't as nice as we have required it to be here. However, we got a somewhat impressive classical result in this paper without having to use anything stronger than Riemann integration. As always, the books of Rudin and of Royden give much insight, and books on Harmonic Analysis consider even more such problems.

Matthew Bond, Department of Mathematics, Michigan State University bondmatt@msu.edu