كلية العلوم للبنات
قسم الرياضيات
المرحلة الرابعة

APPROXIMATION THEORY

The Weierstrass Theorem

Let's begin with some notation. Throughout this chapter, we'll be concerned with the problem of best (uniform) approximation of a given function $f \in C[a, b]$ by elements from $\mathcal{P}_n$, the subspace of algebraic polynomials of degree at most n in $C[a, b]$. We know that the problem has a solution (possibly more than one), which we've chosen to write as p_n^*. We set

$$E_n(f) = \min_{p \in \mathcal{P}_n} \|f - p\| = \|f - p_n^*\|.$$

Because $\mathcal{P}_n \subset \mathcal{P}_{n+1}$ for each n, it's clear that $E_n(f) \geq E_{n+1}(f)$ for each n. Our goal in this chapter is to prove that $E_n(f) \to 0$. We'll accomplish this by proving:

Theorem 2.1. (The Weierstrass Approximation Theorem, 1885) *Let $f \in C[a, b]$. Then, for every $\varepsilon > 0$, there is a polynomial p such that $\|f - p\| < \varepsilon$.*

Lemma 2.2. *If the Weierstrass theorem holds for $C[0, 1]$, then it also holds for $C[a, b]$, and conversely. In fact, $C[0, 1]$ and $C[a, b]$ are, for all practical purposes, identical: They are linearly isometric as normed spaces, order isomorphic as lattices, and isomorphic as algebras* (rings).

The point to our first result is that it suffices to prove the Weierstrass theorem for any interval we like; $[0,1]$ and $[-1,1]$ are popular choices, but it hardly matters which interval we use.

Bernstein's Proof

The proof of the Weierstrass theorem we present here is due to the great Russian mathematician S. N. Bernstein in 1912. Bernstein's proof is of interest to us for a variety of reasons; perhaps most important is that Bernstein actually *displays* a sequence of polynomials that approximate a given $f \in C[0,1]$. Moreover, as we'll see later, Bernstein's proof generalizes to yield a powerful, unifying theorem, called the Bohman-Korovkin theorem (see Theorem 2.9).

If f is any *bounded* function on $[0,1]$, we define the sequence of *Bernstein polynomials* for f by

$$\big(B_n(f)\big)(x) = \sum_{k=0}^{n} f(k/n) \cdot \binom{n}{k} x^k (1-x)^{n-k}, \qquad 0 \le x \le 1.$$

Please note that $B_n(f)$ is a polynomial of degree at most n. Also, it's easy to see that $\big(B_n(f)\big)(0) = f(0)$, and $\big(B_n(f)\big)(1) = f(1)$. In general, $\big(B_n(f)\big)(x)$ is an *average* of the numbers $f(k/n)$, $k = 0, \ldots, n$. Bernstein's theorem states that $B_n(f) \rightrightarrows f$ for each $f \in C[0,1]$. Surprisingly, the proof actually only requires that we check three easy cases:

$$f_0(x) = 1, \quad f_1(x) = x, \quad \text{and} \quad f_2(x) = x^2.$$

Lemma 2.3. (i) $B_n(f_0) = f_0$ *and* $B_n(f_1) = f_1$.

(ii) $B_n(f_2) = \left(1 - \frac{1}{n}\right) f_2 + \frac{1}{n} f_1$, *and hence* $B_n(f_2) \rightrightarrows f_2$.

(iii) $$\sum_{k=0}^{n} \left(\frac{k}{n} - x\right)^2 \binom{n}{k} x^k (1-x)^{n-k} = \frac{x(1-x)}{n} \leq \frac{1}{4n}, \quad \text{if } 0 \leq x \leq 1.$$

(iv) *Given* $\delta > 0$ *and* $0 \leq x \leq 1$, *let* F *denote the set of* k *in* $\{0, \ldots, n\}$ *for which* $\left|\frac{k}{n} - x\right| \geq \delta$. *Then* $\sum_{k \in F} \binom{n}{k} x^k (1-x)^{n-k} \leq \frac{1}{4n\delta^2}$.

Proof. That $B_n(f_0) = f_0$ follows from the binomial formula:

$$\sum_{k=0}^{n} \binom{n}{k} x^k (1-x)^{n-k} = [x + (1-x)]^n = 1.$$

since $f_0(x) = 1,$ then

$$B_n(f_0) = \sum_{k=0}^{n} f_0(x) \binom{n}{k} x^k (1-x)^{n-k} = 1 = f_0$$

Now, since $f_1(x) = x$

T.P. $Bn(f_1) = f_1$,

For $k \geq 1$, then:

$$\frac{k}{n}\binom{n}{k} = \frac{k}{n}\left(\frac{n!}{k!(n-k)!}\right) = \frac{k}{n}\left(\frac{n(n-)!}{k(k-1)!(n-k)!}\right) = \left(\frac{(n-1)!}{(k-1)!(n-k)!}\right) = \binom{n-1}{k-1}$$

Consequently,

$$\sum_{k=0}^{n} \frac{k}{n}\binom{n}{k} x^k(1-x)^{n-k} = x\sum_{k=1}^{n}\binom{n-1}{k-1}x^{k-1}(1-x)^{n-k}$$

$$= x\sum_{j=0}^{n-1}\binom{n-1}{j}x^j(1-x)^{(n-1)-j} = x. \quad \text{where } j=k-1$$

Next , to compute $\mathrm{B_n}(f_2)$

$$\left(\frac{k}{n}\right)^2\binom{n}{k} = \left(\frac{k}{n}\right)\left[\frac{k}{n}\binom{n}{k}\right] = \frac{k}{n}\binom{n-1}{k-1}$$

$$= \frac{k-1+1}{n}\binom{n-1}{k-1} = \frac{k-1}{n}\binom{n-1}{k-1} + \frac{1}{n}\binom{n-1}{k-1} = \frac{n-1}{n}\frac{k-1}{n-1}\binom{n-1}{k-1} + \frac{1}{n}\binom{n-1}{k-1}$$

$$= \frac{n-1}{n}\frac{k-1}{n-1}\frac{(n-1)!}{(k-1)!(n-k)!} + \frac{1}{n}\binom{n-1}{k-1} = \frac{n-1}{n}\frac{k-1}{n-1}\frac{(n-1)(n-2)!}{(k-1)(k-2)!(n-k)!} + \frac{1}{n}\binom{n-1}{k-1} \quad \text{for } k\geq 1$$

$$=\left(1-\frac{1}{n}\right)\binom{n-2}{k-2} + \frac{1}{n}\binom{n-1}{k-1} \quad \text{for } k\geq 2$$

Thus,

$$\sum_{k=0}^{n}\left(\frac{k}{n}\right)^2\binom{n}{k}x^k(1-x)^{n-k}$$
$$=\left(1-\frac{1}{n}\right)\sum_{k=2}^{n}\binom{n-2}{k-2}x^k(1-x)^{n-k} + \frac{1}{n}\sum_{k=1}^{n}\binom{n-1}{k-1}x^k(1-x)^{n-k}$$
$$=\left(1-\frac{1}{n}\right)x^2+\frac{1}{n}x,$$

which establishes (ii) because $\|B_n(f_2)-f_2\| = \frac{1}{n}\|f_1-f_2\| \to 0$ as $n \to \infty$.

To prove (iii) we combine the results in (i) and (ii) and simplify. Because $((k/n)-x)^2 = (k/n)^2 - 2x(k/n) + x^2$, we get

$$\begin{aligned}\sum_{k=0}^{n}\left(\frac{k}{n}-x\right)^2\binom{n}{k}x^k(1-x)^{n-k} &= \left(1-\frac{1}{n}\right)x^2+\frac{1}{n}x-2x^2+x^2 \\ &= \frac{1}{n}x(1-x) \;\le\; \frac{1}{4n},\end{aligned}$$

for $0 \le x \le 1$.

Finally, to prove (iv), note that $1 \le ((k/n)-x)^2/\delta^2$ for $k \in F$, and hence

$$\begin{aligned}\sum_{k\in F}\binom{n}{k}x^k(1-x)^{n-k} &\le \frac{1}{\delta^2}\sum_{k\in F}\left(\frac{k}{n}-x\right)^2\binom{n}{k}x^k(1-x)^{n-k} \\ &\le \frac{1}{\delta^2}\sum_{k=0}^{n}\left(\frac{k}{n}-x\right)^2\binom{n}{k}x^k(1-x)^{n-k} \\ &\le \frac{1}{4n\delta^2}, \quad \text{from (iii).}\end{aligned}$$

□

Theorem 2.1. (The Weierstrass Approximation Theorem, 1885) *Let $f \in C[a,b]$. Then, for every $\varepsilon > 0$, there is a polynomial p such that $\|f - p\| < \varepsilon$.*

Proof. Let $f \in C[0,1]$ and let $\varepsilon > 0$. Then, because f is uniformly continuous, there is a $\delta > 0$ such that $|f(x) - f(y)| < \varepsilon/2$ whenever $|x - y| < \delta$. Now we use the previous lemma to estimate $\|f - B_n(f)\|$. First notice that because the numbers $\binom{n}{k} x^k (1-x)^{n-k}$ are nonnegative and sum to 1, we have

$$\begin{aligned} |f(x) - B_n(f)(x)| &= \left| f(x) - \sum_{k=0}^{n} \binom{n}{k} f\left(\frac{k}{n}\right) x^k (1-x)^{n-k} \right| \\ &= \left| \sum_{k=0}^{n} \left(f(x) - f\left(\frac{k}{n}\right)\right) \binom{n}{k} x^k (1-x)^{n-k} \right| \\ &\le \sum_{k=0}^{n} \left| f(x) - f\left(\frac{k}{n}\right)\right| \binom{n}{k} x^k (1-x)^{n-k}, \end{aligned}$$

Now fix n (to be specified in a moment) and let F denote the set of k in $\{0, \ldots, n\}$ for which $|(k/n) - x| \ge \delta$. Then $|f(x) - f(k/n)| < \varepsilon/2$ for $k \notin F$, while $|f(x) - f(k/n)| \le 2\|f\|$ for $k \in F$. Thus,

$$\begin{aligned} &|f(x) - (B_n(f))(x)| \\ &\quad \le \frac{\varepsilon}{2} \sum_{k \notin F} \binom{n}{k} x^k (1-x)^{n-k} + 2\|f\| \sum_{k \in F} \binom{n}{k} x^k (1-x)^{n-k} \\ &\quad < \frac{\varepsilon}{2} \cdot 1 \ + \ 2\|f\| \cdot \frac{1}{4n\delta^2}, \quad \text{from (iv) of the Lemma,} \\ &\quad < \varepsilon, \quad \text{provided that } n > \|f\|/\varepsilon\delta^2. \end{aligned}$$

□

Landau's Proof

Just because it's good for us, let's give *a second proof of Weierstrass's theorem*. This one is due to Landau in 1908. First, given $f \in C[0,1]$, notice that it suffices to approximate $f - p$, where p is any polynomial. (Why?) In particular, by subtracting the *linear* function $f(0) + x(f(1) - f(0))$, we may suppose that $f(0) = f(1) = 0$ and, hence, that $f \equiv 0$ outside $[0,1]$. That is, we may suppose that f is defined and uniformly continuous on all of $\mathbb{R}$.

Again we will display a sequence of polynomials that converge uniformly to f; this time we define

$$L_n(x) = c_n \int_{-1}^{1} f(x+t)\,(1-t^2)^n\,dt,$$

where c_n is chosen so that

$$c_n \int_{-1}^{1} (1-t^2)^n\,dt = 1.$$

Note that by our assumptions on f, we may rewrite $L_n(x)$ as

$$L_n(x) = c_n \int_{-x}^{1-x} f(x+t)\,(1-t^2)^n\,dt = c_n \int_0^1 f(t)\,(1-(t-x)^2)^n\,dt.$$

Written this way, it's clear that L_n is a polynomial in x of degree at most n.

We first need to estimate c_n. An easy induction argument will convince you that $(1-t^2)^n \geq 1 - nt^2$, and so we get

$$\int_{-1}^{1} (1-t^2)^n\,dt \geq 2\int_0^{1/\sqrt{n}} (1-nt^2)\,dt = \frac{4}{3\sqrt{n}} > \frac{1}{\sqrt{n}},$$

from which it follows that $c_n < \sqrt{n}$. In particular, for any $0 < \delta < 1$,

$$c_n \int_{\delta}^{1} (1-t^2)^n\,dt < \sqrt{n}\,(1-\delta^2)^n \to 0 \qquad (n \to \infty),$$

which is the inequality we'll need.

Next, let $\varepsilon > 0$ be given, and choose $0 < \delta < 1$ such that

$$|f(x) - f(y)| \le \varepsilon/2 \text{ whenever } |x - y| \le \delta.$$

Then, because $c_n(1-t^2)^n \ge 0$ and integrates to 1, we get

$$\begin{aligned}
|L_n(x) - f(x)| &= \left| c_n \int_{-1}^{1} [f(x+t) - f(x)](1-t^2)^n \, dt \right| \\
&\le c_n \int_{-1}^{1} |f(x+t) - f(x)|(1-t^2)^n \, dt \\
&\le \frac{\varepsilon}{2} c_n \int_{-\delta}^{\delta} (1-t^2)^n \, dt + 4\|f\| \, c_n \int_{\delta}^{1} (1-t^2)^n \, dt \\
&\le \frac{\varepsilon}{2} + 4\|f\| \sqrt{n} \, (1-\delta^2)^n < \varepsilon,
\end{aligned}$$

provided that n is sufficiently large. □

To begin, we will need a bit more notation. The *modulus of continuity* of a bounded function f on the interval $[a, b]$ is defined by

$$\omega_f(\delta) = \omega_f([a,b];\delta) = \sup\{ \, |f(x) - f(y)| : x, y \in [a,b], \ |x - y| \le \delta \, \}$$

for any $\delta > 0$. Note that $\omega_f(\delta)$ is a measure of the "ε" that goes along with δ (in the definition of uniform continuity); literally, we have written $\varepsilon = \omega_f(\delta)$ as a function of δ.

Lemma 2.5. *Let f be a bounded function on $[a,b]$ and let $\delta > 0$. Then, $\omega_f(n\delta) \le n\,\omega_f(\delta)$ for $n = 1, 2, \ldots$. Consequently, $\omega_f(\lambda\delta) \le (1+\lambda)\,\omega_f(\delta)$ for any $\lambda > 0$.*

Proof. Given $x < y$ with $|x - y| \le n\,\delta$, split the interval $[x, y]$ into n pieces, each of length at most δ. Specifically, if we set $z_k = x + k(y-x)/n$, for $k = 0, 1, \ldots, n$, then $|z_k - z_{k-1}| \le \delta$ for any $k \ge 1$, and so

$$\begin{aligned} |f(x) - f(y)| &= \left|\sum_{k=1}^{n} f(z_k) - f(z_{k-1})\right| \\ &\le \sum_{k=1}^{n} |f(z_k) - f(z_{k-1})| \\ &\le n\,\omega_f(\delta). \end{aligned}$$

Thus, $\omega_f(n\delta) \le n\,\omega_f(\delta)$.

The second assertion follows from the first (and one of our exercises). Given $\lambda > 0$, choose an integer n so that $n - 1 < \lambda \le n$. Then,

$$\omega_f(\lambda\delta) \le \omega_f(n\,\delta) \le n\,\omega_f(\delta) \le (1+\lambda)\,\omega_f(\delta). \qquad \square$$

Theorem 2.6. *For any bounded function f on $[0,1]$ we have*

$$\|f - B_n(f)\| \leq \frac{3}{2}\,\omega_f\left(\frac{1}{\sqrt{n}}\right).$$

In particular, if $f \in C[0,1]$, then $E_n(f) \leq \frac{3}{2}\,\omega_f(\frac{1}{\sqrt{n}}) \to 0$ as $n \to \infty$.

Proof. We first do some term juggling:

$$\begin{aligned}
|f(x) - B_n(f)(x)| &= \left|\sum_{k=0}^{n}\left(f(x) - f\left(\frac{k}{n}\right)\right)\binom{n}{k}x^k(1-x)^{n-k}\right| \\
&\leq \sum_{k=0}^{n}\left|f(x) - f\left(\frac{k}{n}\right)\right|\binom{n}{k}x^k(1-x)^{n-k} \\
&\leq \sum_{k=0}^{n}\omega_f\left(\left|x - \frac{k}{n}\right|\right)\binom{n}{k}x^k(1-x)^{n-k} \\
&\leq \omega_f\left(\frac{1}{\sqrt{n}}\right)\sum_{k=0}^{n}\left[1 + \sqrt{n}\left|x - \frac{k}{n}\right|\right]\binom{n}{k}x^k(1-x)^{n-k} \\
&= \omega_f\left(\frac{1}{\sqrt{n}}\right)\left[1 \;+\; \sqrt{n}\sum_{k=0}^{n}\left|x - \frac{k}{n}\right|\binom{n}{k}x^k(1-x)^{n-k}\right],
\end{aligned}$$

where the third inequality follows from Lemma 2.5 (by taking $\lambda = \sqrt{n}\,\left|x - \frac{k}{n}\right|$ and $\delta = \frac{1}{\sqrt{n}}$). All that remains is to estimate the sum, and for this we'll use Cauchy-Schwarz (and our earlier observations about Bernstein polynomials). Because each of the terms $\binom{n}{k}x^k(1-x)^{n-k}$ is nonnegative, we have

$$\begin{aligned}
&\sum_{k=0}^{n}\left|x-\frac{k}{n}\right|\binom{n}{k}x^k(1-x)^{n-k} \\
&\qquad = \sum_{k=0}^{n}\left|x-\frac{k}{n}\right|\left[\binom{n}{k}x^k(1-x)^{n-k}\right]^{1/2}\cdot\left[\binom{n}{k}x^k(1-x)^{n-k}\right]^{1/2} \\
&\qquad \le \left[\sum_{k=0}^{n}\left|x-\frac{k}{n}\right|^2\binom{n}{k}x^k(1-x)^{n-k}\right]^{1/2}\cdot\left[\sum_{k=0}^{n}\binom{n}{k}x^k(1-x)^{n-k}\right]^{1/2} \\
&\qquad \le \left[\frac{1}{4n}\right]^{1/2} = \frac{1}{2\sqrt{n}}.
\end{aligned}$$

Finally,

$$|f(x) - B_n(f)(x)| \le \omega_f\left(\frac{1}{\sqrt{n}}\right)\left[1 \;+\; \sqrt{n}\cdot\frac{1}{2\sqrt{n}}\right] = \frac{3}{2}\,\omega_f\left(\frac{1}{\sqrt{n}}\right). \qquad \square$$

The Bohman-Korovkin Theorem

The real value to us in Bernstein's approach is that the map $f \mapsto B_n(f)$, while providing a simple formula for an approximating polynomial, is also *linear* and *positive*. In other words,

$$B_n(f+g) = B_n(f) + B_n(g),$$
$$B_n(\alpha f) = \alpha B_n(f), \ \alpha \in \mathbb{R},$$
and
$$B_n(f) \geq 0 \text{ whenever } f \geq 0.$$

Lemma 2.8. *If $T : C[a,b] \to C[a,b]$ is both positive and linear, then T is continuous.*

Proof. First note that a positive, linear map is also *monotone*. That is, T satisfies $T(f) \leq T(g)$ whenever $f \leq g$. (Why?) Thus, for any $f \in C[a,b]$, we have

$$-f, f \leq |f| \implies -T(f), T(f) \leq T(|f|);$$

that is, $|T(f)| \leq T(|f|)$. But now $|f| \leq \|f\| \cdot \mathbf{1}$, where $\mathbf{1}$ denotes the constant 1 function, and so we get

$$|T(f)| \leq T(|f|) \leq \|f\| \, T(\mathbf{1}).$$

Thus,

$$\|T(f)\| \leq \|f\| \, \|T(\mathbf{1})\|$$

for any $f \in C[a,b]$. Finally, because T is linear, it follows that T is Lipschitz with constant $\|T(\mathbf{1})\|$:

$$\|T(f) - T(g)\| = \|T(f-g)\| \leq \|T(\mathbf{1})\| \, \|f-g\|.$$

Consequently, T is continuous. □

Theorem 2.9. *Let* $T_n : C[0,1] \to C[0,1]$ *be a sequence of positive, linear maps, and suppose that* $T_n(f) \to f$ *uniformly in each of the three cases*

$$f_0(x) = 1, \quad f_1(x) = x, \quad \text{and} \quad f_2(x) = x^2.$$

Then, $T_n(f) \to f$ *uniformly for every* $f \in C[0,1]$.

The proof of the Bohman-Korovkin theorem is essentially identical to the proof of Bernstein's theorem except, of course, we write $T_n(f)$ in place of $B_n(f)$. For full details, see [12]. Rather than proving the theorem, let's settle for a quick application.