Chapter 2

Approximation by Algebraic Polynomials

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 $\|f - p\| < \varepsilon$.*

It follows from the Weierstrass theorem that, for some sequence of polynomials (q_k), we have $\|f - q_k\| \to 0$. We may suppose that $q_k \in \mathcal{P}_{n_k}$ where (n_k) is increasing. (Why?) Whence it follows that $E_n(f) \to 0$; that is, $p_n^* \rightrightarrows f$. (Why?) This is an important first step in determining the exact nature of $E_n(f)$ as a function of f and n. We'll look for much more precise information in later sections.

Now there are many proofs of the Weierstrass theorem (a mere three are outlined in the exercises, but there are hundreds!), and all of them start with one simplification: The underlying interval $[\,a,b\,]$ is of no consequence.

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).

Proof. We'll settle for proving only the first assertion; the second is outlined in the exercises (and uses a similar argument). See Problem 1.

First, notice that the function

$$\sigma(x) = a + (b - a)x, \quad 0 \leq x \leq 1,$$

defines a continuous, one-to-one map from $[0,1]$ onto $[a,b]$. Given $f \in C[a,b]$, it follows that $g(x) = f(\sigma(x))$ defines an element of $C[0,1]$. Moreover,

$$\max_{0\le x\le 1} |g(x)| = \max_{a\le t\le b} |f(t)|.$$

Now, given $\varepsilon > 0$, suppose that we can find a polynomial p such that $\|g - p\| < \varepsilon$; in other words, suppose that

$$\max_{0\le x\le 1} \big|f\big(a + (b-a)x\big) - p(x)\big| < \varepsilon.$$

Then,

$$\max_{a\le t\le b} \left|f(t) - p\left(\frac{t-a}{b-a}\right)\right| < \varepsilon.$$

(Why?) But if $p(x)$ is a polynomial in x, then $q(t) = p\left(\frac{t-a}{b-a}\right)$ is a polynomial in t satisfying $\|f - q\| < \varepsilon$.

The proof of the converse is entirely similar: If $g(x)$ is an element of $C[0,1]$, then $f(t) = g\left(\frac{t-a}{b-a}\right)$, $a \le t \le b$, defines an element of $C[a,b]$. Moreover, if $q(t)$ is a polynomial in t approximating $f(t)$, then $p(x) = q(a + (b-a)x)$ is a polynomial in x approximating $g(x)$. The remaining details are left as an **exercise**. □

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:

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

This, and more, is the content of the following lemma.

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) $\displaystyle\sum_{k=0}^{n}\Big(\frac{k}{n}-x\Big)^2\binom{n}{k}x^k(1-x)^{n-k} = \frac{x(1-x)}{n} \le \frac{1}{4n},$ *if* $0 \le x \le 1$.

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

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

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

To see that $B_n(f_1) = f_1$, first notice that for $k \ge 1$ we have

$$\frac{k}{n}\binom{n}{k} = \frac{(n-1)\,!}{(k-1)\,!\,(n-k)\,!} = \binom{n-1}{k-1}.$$

Consequently,

$$\begin{aligned}\sum_{k=0}^{n}\frac{k}{n}\binom{n}{k}x^k(1-x)^{n-k} &= x\sum_{k=1}^{n}\binom{n-1}{k-1}x^{k-1}(1-x)^{n-k}\\ &= x\sum_{j=0}^{n-1}\binom{n-1}{j}x^j(1-x)^{(n-1)-j} \;=\; x.\end{aligned}$$

Next, to compute $B_n(f_2)$, we rewrite twice:

$$\begin{aligned}\Big(\frac{k}{n}\Big)^2\binom{n}{k} = \frac{k}{n}\binom{n-1}{k-1} &= \frac{n-1}{n}\cdot\frac{k-1}{n-1}\binom{n-1}{k-1} + \frac{1}{n}\binom{n-1}{k-1}, \text{ if } k \ge 1\\ &= \Big(1-\frac{1}{n}\Big)\binom{n-2}{k-2} + \frac{1}{n}\binom{n-1}{k-1}, \text{ if } k \ge 2.\end{aligned}$$

Thus,

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

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

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

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

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

Now we're ready for *the proof of Bernstein's theorem*:

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,

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

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 $(1-t^2)^n \ge 1 - nt^2$, and so we get

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

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

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

provided that n is sufficiently large. □

A third proof of the Weierstrass theorem, due to Lebesgue in 1898, is outlined in the problems at the end of the chapter (see Problem 7). Lebesgue's proof is of historical interest because it inspired Stone's version of the Weierstrass theorem, which we'll discuss in Chapter 11.

Before we go on, let's stop and make an observation or two: While the Bernstein polynomials $B_n(f)$ offer a convenient and explicit polynomial approximation to f, they are by no means the best approximations. Indeed, recall that if $f_1(x) = x$ and $f_2(x) = x^2$, then $B_n(f_2) = (1 - \frac{1}{n})f_2 + \frac{1}{n}f_1 \ne f_2$. Clearly, the best approximation to f_2 out of $\mathcal{P}_n$ should be f_2 itself whenever $n \ge 2$. On the other hand, because we always have

$$E_n(f) \le \|f - B_n(f)\| \qquad \text{(why?)},$$

a detailed understanding of Bernstein's proof will lend insight into the general problem of polynomial approximation. Our next project, then, is to improve upon our estimate of the error $\|f - B_n(f)\|$.

Improved Estimates

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

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

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.

Here are a few easy facts about the modulus of continuity:

Exercise 2.4.

1. We always have $|f(x)-f(y)| \le \omega_f(\,|x-y|\,)$ for any $x \ne y \in [\,a,b\,]$.
2. If $0 < \delta' \le \delta$, then $\omega_f(\delta') \le \omega_f(\delta)$.
3. f is *uniformly continuous* if and only if $\omega_f(\delta) \to 0$ as $\delta \to 0^+$. [Hint: The statement that $|f(x)-f(y)| \le \varepsilon$ whenever $|x-y| \le \delta$ is equivalent to the statement that $\omega_f(\delta) \le \varepsilon$.]
4. If $f\,'$ exists and is bounded on $[\,a,b\,]$, then $\omega_f(\delta) \le K\delta$ for some constant K.
5. We say that f satisfies a *Lipschitz condition* of order α with constant K, where $0 < \alpha \le 1$ and $0 \le K < \infty$, if $|f(x)-f(y)| \le K|x-y|^\alpha$ for all x, y. We abbreviate this statement by writing: $f \in \text{lip}_K\alpha$. Check that if $f \in \text{lip}_K\alpha$, then $\omega_f(\delta) \le K\delta^\alpha$ for all $\delta > 0$.

For the time being, we actually need only one simple fact about $\omega_f(\delta)$:

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,

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

We next repeat the proof of Bernstein's theorem, making a few minor adjustments here and there.

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

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

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

Proof. We first do some term juggling:

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

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

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

Finally,

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

Examples 2.7.

1. If $f \in \mathrm{lip}_K\alpha$, it follows that $\|f - B_n(f)\| \le \frac{3}{2}Kn^{-\alpha/2}$ and hence that $E_n(f) \le \frac{3}{2}Kn^{-\alpha/2}$.

2. As a particular case of the first example, consider $f(x) = \left|x - \frac{1}{2}\right|$ on $[0,1]$. Then $f \in \text{lip}_1 1$, and so $\|f - B_n(f)\| \le \frac{3}{2}\, n^{-1/2}$. But, as Rivlin points out (see Remark 3 on p. 16 of [45]), $\|f - B_n(f)\| > \frac{1}{2}\, n^{-1/2}$. Thus, we can't hope to improve on the power of n in this estimate. Nevertheless, we will see an improvement in our estimate of $E_n(f)$.

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
$$B_n(f) \ge 0 \text{ whenever } f \ge 0.$$

(See Problem 15 for more on this.) As it happens, any positive, linear map $T : C[\,0,1\,] \to C[\,0,1\,]$ is automatically continuous!

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 $T(f) \le T(g)$ whenever $f \le g$. (Why?) Thus, for any $f \in C[\,a,b\,]$, we have

$$-f,\ f \le |f| \Longrightarrow -T(f),\ T(f) \le T(|f|);$$

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

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

Thus,

$$\|T(f)\| \le \|f\|\ \|T(\mathbf{1})\|$$

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

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

Consequently, T is continuous. □

Now positive, linear maps abound in analysis, so this is a fortunate turn of events. What's more, Bernstein's theorem generalizes very nicely when placed in this new setting. The following elegant theorem was proved (independently) by Bohman and Korovkin in, roughly, 1952.

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.

Example 2.10. Let $f \in C[\,0,1\,]$ and, for each n, let $L_n(f)$ be the polygonal approximation to f with nodes at k/n, $k = 0, 1, \ldots, n$. That is, $L_n(f)$ is linear on each subinterval $[\,(k-1)/n,\, k/n\,]$ and agrees with f at each of the endpoints: $L_n(f)(k/n) = f(k/n)$. Then $L_n(f) \to f$ uniformly for each $f \in C[\,0,1\,]$. This is actually an easy calculation all by itself, but let's see why the Bohman-Korovkin theorem makes short work of it.

That $L_n(f)$ is positive and linear is (nearly) obvious; that $L_n(f_0) = f_0$ and $L_n(f_1) = f_1$ are really easy because, in fact, $L_n(f) = f$ for any linear function f. We just need to show that $L_n(f_2) \rightrightarrows f_2$. But a picture will convince you that the maximum distance between $L_n(f_2)$ and f_2 on the interval $[\,(k-1)/n,\, k/n\,]$ is at most

$$\left(\frac{k}{n}\right)^2 - \left(\frac{k-1}{n}\right)^2 = \frac{2k-1}{n^2} \le \frac{2}{n}.$$

That is, $\|f_2 - L_n(f_2)\| \le 2/n \to 0$ as $n \to \infty$.

[Note that L_n is a linear projection from $C[\,0,1\,]$ onto the subspace of polygonal functions based on the nodes k/n, $k = 0, \ldots, n$. An easy calculation, similar in spirit to the example above, will show that $\|f - L_n(f)\| \le 2\,\omega_f(1/n) \to 0$ as $n \to \infty$ for any $f \in C[\,0,1\,]$. See Problem 8.]

Problems

▷∗ 1. Define $\sigma : [\,0,1\,] \to [\,a,b\,]$ by $\sigma(t) = a + t(b-a)$ for $0 \le t \le 1$, and define a transformation $T_\sigma : C[\,a,b\,] \to C[\,0,1\,]$ by $(T_\sigma(f))(t) = f(\sigma(t))$. Prove that T_σ satisfies:

(a) $T_\sigma(f+g) = T_\sigma(f) + T_\sigma(g)$ and $T_\sigma(cf) = c\,T_\sigma(f)$ for $c \in \mathbb{R}$.

(b) $T_\sigma(fg) = T_\sigma(f)\,T_\sigma(g)$. In particular, T_σ maps polynomials to polynomials.

(c) $T_\sigma(f) \le T_\sigma(g)$ if and only if $f \le g$.

(d) $\|T_\sigma(f)\| = \|f\|$.

(e) T_σ is both one-to-one and onto. Moreover, $(T_\sigma)^{-1} = T_{\sigma^{-1}}$.

▷∗ 2. Bernstein's Theorem shows that the polynomials are *dense* in $C[\,0,1\,]$. Using the results in Problem 1, conclude that the polynomials are also dense in $C[\,a,b\,]$.

▷∗ 3. How do we know that there are non-polynomial elements in $C[\,0,1\,]$? In other words, is it possible that every element of $C[\,0,1\,]$ agrees with some polynomial on $[\,0,1\,]$?

4. Let (Q_n) be a sequence of polynomials of degree m_n, and suppose that (Q_n) converges uniformly to f on $[\,a,b\,]$, where f is *not* a polynomial. Show that $m_n \to \infty$.

5. Use induction to show that $(1+x)^n \ge 1 + nx$, for all $n = 1, 2, \ldots$, whenever $x \ge -1$. Conclude that $(1-t^2)^n \ge 1 - nt^2$ whenever $-1 \le t \le 1$.

A *polygonal function* is a piecewise linear, continuous function; that is, a continuous function $f : [\,a,b\,] \to \mathbb{R}$ is a polygonal function if there are finitely many distinct points $a = x_0 < x_1 < \cdots < x_n = b$, called *nodes*, such that f is linear on each of the intervals $[\,x_{k-1}, x_k\,]$, $k = 1, \ldots, n$.

Fix distinct points $a = x_0 < x_1 < \cdots < x_n = b$ in $[\,a,b\,]$, and let S_n denote the set of all polygonal functions having nodes at the x_k. It's not hard to see that S_n is a vector space. In fact, it's relatively clear that S_n must have dimension exactly $n+1$ as there are $n+1$ "degrees of freedom" (each element of S_n is completely determined by its values at the x_k). More convincing, perhaps, is the fact that we can easily display a *basis* for S_n. (see Natanson [41]).

∗ 6. (a) Show that S_n is an $(n+1)$-dimensional subspace of $C[\,a,b\,]$ spanned by the constant function $\varphi_0(x) = 1$ and the "angles" $\varphi_{k+1}(x) = |x - x_k| + (x - x_k)$ for $k = 0, \ldots, n-1$. Specifically, show that each $h \in S_n$ can be uniquely written as $h(x) = c_0 + \sum_{i=1}^n c_i \varphi_i(x)$. [Hint: Because each side of the equation is an element of S_n, it's enough to show that the system of equations $h(x_0) = c_0$ and $h(x_k) = c_0 + 2\sum_{i=1}^k c_i(x_k - x_{i-1})$ for $k = 1, \ldots, n$ can be solved (uniquely) for the c_i.]

(b) Each element of S_n can be written as $\sum_{i=1}^{n-1} a_i |x - x_i| + bx + d$ for some choice of scalars $a_1, \ldots, a_{n-1}$, b, d.

∗ 7. Given $f \in C[\,0,1\,]$, show that f can be uniformly approximated by a polygonal function. Specifically, given a positive integer n, let $L_n(x)$ denote the unique polygonal function with nodes at $(k/n)_{k=0}^n$ that agrees with f at each of these nodes. Show that $\|f - L_n\|$ is small provided that n is sufficiently large.

8. (a) Let f be in $\text{lip}_C 1$; that is, suppose that f satisfies $|f(x) - f(y)| \le C|x - y|$ for some constant C and all x, y in $[0,1]$. In the notation of Problem 7, show that $\|f - L_n\| \le 2C/n$. [Hint: Given x in $[k/n, (k+1)/n)$, check that $|f(x) - L_n(x)| = |f(x) - f(k/n) + L_n(k/n) - L_n(x)| \le |f(x) - f(k/n)| + |f((k+1)/n) - f(k/n)|$.]

   (b) More generally, prove that $\|f - L_n(f)\| \le 2\,\omega_f(1/n) \to 0$ as $n \to \infty$ for any $f \in C[0,1]$.

In light of the results in Problems 6 and 7, Lebesgue noted that he could fashion a proof of Weierstrass's Theorem provided he could prove that $|x - c|$ can be uniformly approximated by polynomials on any interval $[a,b]$. (Why is this enough?) But thanks to the result in Problem 1, for this we need only show that $|x|$ can be uniformly approximated by polynomials on the interval $[-1,1]$.

∗ 9. Here's an elementary proof that there is a sequence of polynomials (P_n) converging uniformly to $|x|$ on $[-1,1]$.

   (a) Define (P_n) recursively by $P_{n+1}(x) = P_n(x) + [x - P_n(x)^2]/2$, where $P_0(x) = 0$. Clearly, each P_n is a polynomial.

   (b) Check that $0 \le P_n(x) \le P_{n+1}(x) \le \sqrt{x}$ for $0 \le x \le 1$. Use Dini's theorem to conclude that $P_n(x) \rightrightarrows \sqrt{x}$ on $[0,1]$.

   (c) $P_n(x^2)$ is also a polynomial, and $P_n(x^2) \rightrightarrows |x|$ on $[-1,1]$.

10. If $f \in C[-1,1]$ is an *even* function, show that f may be uniformly approximated by even polynomials (that is, polynomials of the form $\sum_{k=0}^{n} a_k x^{2k}$).

11. If $f \in C[0,1]$ and if $f(0) = f(1) = 0$, show that the sequence of polynomials $\sum_{k=0}^{n} \left[\binom{n}{k} f(k/n)\right] x^k (1-x)^{n-k}$ having *integer* coefficients converges uniformly to f (where $[x]$ denotes the greatest integer in x). The same trick works for any $f \in C[a,b]$ provided that $0 < a < b < 1$.

12. If p is a polynomial and $\varepsilon > 0$, prove that there is a polynomial q with *rational* coefficients such that $\|p - q\| < \varepsilon$ on $[0,1]$. Conclude that $C[0,1]$ is *separable* (that is, $C[0,1]$ has a countable dense subset).

13. Let (x_i) be a sequence of numbers in $(0,1)$ such that $\lim\limits_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} x_i^k$ exists for every $k = 0, 1, 2, \ldots$. Show that $\lim\limits_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} f(x_i)$ exists for every $f \in C[0,1]$.

14. If $f \in C[0,1]$ and if $\int_0^1 x^n f(x)\,dx = 0$ for each $n = 0, 1, 2, \ldots$, show that $f \equiv 0$. [Hint: Using the Weierstrass theorem, show that $\int_0^1 f^2 = 0$.]

▷ 15. Show that $|B_n(f)| \le B_n(|f|)$, and that $B_n(f) \ge 0$ whenever $f \ge 0$. Conclude that $\|B_n(f)\| \le \|f\|$.

16. If f is a bounded function on $[0,1]$, show that $B_n(f)(x) \to f(x)$ at each point of continuity of f.

17. Find $B_n(f)$ for $f(x) = x^3$. [Hint: $k^2 = (k-1)(k-2) + 3(k-1) + 1$.] The same method of calculation can be used to show that $B_n(f) \in \mathcal{P}_m$ whenever $f \in \mathcal{P}_m$ and $n > m$.

∗ 18. Let f be continuously differentiable on $[a,b]$, and let $\varepsilon > 0$. Show that there is a polynomial p such that $\|f - p\| < \varepsilon$ *and* $\|f' - p'\| < \varepsilon$.

19. Suppose that $f \in C[\,a,b\,]$ is twice continuously differentiable and has $f'' > 0$. Prove that the best linear approximation to f on $[\,a,b\,]$ is $a_0 + a_1 x$ where $a_0 = f'(c)$, $a_1 = [f(a) + f(c) + f'(c)(a+c)]/2$, and where c is the unique solution to $f'(c) = (f(b) - f(a))/(b-a)$.

20. Let $f : [\,a,b\,] \to \mathbb{R}$ be a bounded function. Prove that

$$\omega_f([\,a,b\,];\delta) = \sup\big\{\,\mathrm{diam}\big(f(I)\big) : I \subset [\,a,b\,],\ \mathrm{diam}(I) \le \delta\,\big\}$$

where I denotes a closed subinterval of $[\,a,b\,]$ and where $\mathrm{diam}(A)$ denotes the diameter of the set A.

21. If the graph of $f : [\,a,b\,] \to \mathbb{R}$ has a jump of magnitude $\alpha > 0$ at some point x_0 in $[\,a,b\,]$, then $\omega_f(\delta) \ge \alpha$ for all $\delta > 0$.

22. Calculate ω_g for $g(x) = \sqrt{x}$.

23. If $f \in C[\,a,b\,]$, show that $\omega_f(\delta_1 + \delta_2) \le \omega_f(\delta_1) + \omega_f(\delta_2)$ and that $\omega_f(\delta) \downarrow 0$ as $\delta \downarrow 0$. Use this to show that ω_f is continuous for $\delta \ge 0$. Finally, show that the modulus of continuity of ω_f is again ω_f.

▷ 24. (a) Let $f : [-1,1\,] \to \mathbb{R}$. If $x = \cos\theta$, where $-1 \le x \le 1$, and if $g(\theta) = f(\cos\theta)$, show that $\omega_g([-\pi,\pi\,],\delta) = \omega_g([\,0,\pi\,],\delta) \le \omega_f([-1,1\,];\delta)$.

    (b) If $h(x) = f(ax+b)$ for $c \le x \le d$, show that $\omega_h([\,c,d\,];\delta) = \omega_f([\,ac+b, ad+b\,]; a\delta)$.

25. (a) Let f be continuously differentiable on $[\,0,1\,]$. Show that $(B_n(f)')$ converges uniformly to f' by showing that $\|B_n(f') - (B_{n+1}(f))'\,\| \le \omega_{f'}(1/(n+1))$.

    (b) In order to see why this is of interest, find a uniformly convergent sequence of polynomials whose derivatives fail to converge uniformly.

    [Compare this result with Problem 18.]