CHAPTER FIVE

APPROXIMATION BY POLYNOMIALS

§1. NORMS

We have seen in the fourth chapter that we can interpolate a polynomial function at any finite set of points, so that polynomials are sufficiently numerous to reflect functional values at a finite set of points. This raises the question: *given an arbitrary continuous function on a set S of real or complex numbers, how closely can we approximate it by a polynomial?*

We need to specify exactly what we mean by approximation. If we try to match a function by a polynomial at a finite number of points, then we have poor control over how the polynomial relates to the function away from these points. To speak reasonably of approximation of a function by a polynomial, we want to have the polynomial close to the function on an infinite domain of points, usually an interval of the real line, or some set with a non-void interior, such as a disc, in the complex plane.

The first recourse that might come to the mind of an undergraduate is the production of a Taylor polynomial for the function. There are two difficulties with this approach. First, the production of Taylor polynomials of arbitrarily high degree depends on the function being infinitely differentiable. Secondly, the Taylor polynomial of the nth degree matches the first n derivatives of the function at a particular real or complex point, and so is an excellent *local* approximation. However, as we move away from the point, the closeness of the two functions deterioriates, and indeed the Taylor series fails to converge outside of an interval or disc of convergence.

It is customary to discuss approximation by polynomials in terms of a normed linear space that contains all the polynomials along with the function to be approximated. Such a space is a vector space over **R** or **C** for which each of its entries f possesses a norm $\|f\|$ with the following properties:

(1) $\|f\| \geq 0$, with equality if and only if $f = 0$;

(2) $\|cf\| = |c|\|f\|$ for each scalar c and function f;

(3) $\|f + g\| \leq \|f\| + \|g\|$ for any pair of functions f and g.

The distance between two functions f and g is defined as $\|f - g\|$, in much the same way as we describe the distance between two complex numbers in terms of the absolute value of the difference.

The Uniform Norm. Let V be a vector space of continuous functions defined on a closed subset K of **C**. Then each function is bounded on K and we can defined the *uniform* norm:

$$\|f\|_\infty = \sup\{|f(z)| : z \in K\} \ .$$

A sequence $\{f_n\}$ of functions in V is said to *converge uniformly* to f if and only if $\lim_{n\rightarrow\infty} \|f - f_n\|_\infty = 0$. This agrees with the definition of uniform convergence on a set given in an advanced calculus course.

The L_p Norm. Let $1 \leq p \leq \infty$. Consider the space $L_p(K)$ of complex functions defined on a closed complex set K for which $|f|^p$ is Lebesgue integrable. In this setting, we regard two functions as being equal if and only if the set of points upon which they differ has Lebesgue measure 0. Then

$$\|f\|_p = \begin{cases} (\int_K |f|^p dt)^{1/p}; & \text{if } 1 \leq p < \infty \\ \text{esssup } \{|f(t)| : t \in K\}; & \text{if } p = \infty \end{cases}$$

defines a norm on $L_p(K)$. If $p = 2$, then $L_p(K)$ is actually a Hilbert space whose norm is given by the inner product

$$\langle f, g\rangle = \int_K \bar{f} g dt \ .$$

§2. WEIERSTRASSS APPROXIMATION THEOREM

Let us consider the problem of approximation by polynomials of real functions defined on a closed real interval $[a, b]$.

Weierstrass Approximation Theorem. *Let f be a continuous complex function on $[a, b]$. There exists a sequence $\{p_n\}$ of polynomials that converges uniformly to the function f on $[a, b]$. If f takes real values, then the polynomials can be determined so they take real values.*

The use of an approximate identity. [**8**] The first strategy that we shall consider is one that is used in many approximation situations, and that is to convolve the function f with an "approximate identity" that embraces the property that we wish the approximants to have, in this case, of being polynomials. First, note that we can restrict the argument to approximating a function which vanishes at the end points 0 and 1. (If we can approximate $f(x) - f(0) - x[f(1) - f(0)]$ by a sequence of polynomials, then by adding the polynomial $f(0) + x[f(1) - f(0)]$ to each approximant, we get the desired sequence of approximants.)

Thus, let f be a continuous function for which $f(0) = f(1) = 0$ and extend this function to all of $\mathbf{R}$ by making it vanish everywhere off $[0, 1]$. The function, so extended, is uniformly continuous.

Define the polynomial of degree $2n$,

$$q_n(x) = c_n(1 - x^2)^n$$

where c_n is chosen so that

$$\int_{-1}^{1} q_n(x)dx = 1 \ .$$

Using the fact that $(1 - x^2)^n \geq 1 - nx^2$, we can get an estimate for the value of c_n:

$$\begin{aligned} 1 = 2c_n \int_0^1 (1 - x^2)^n dx &\geq 2c_n \int_0^{1/\sqrt{n}} (1 - x^2)^n dx \\ &\geq 2c_n \int_0^{1/\sqrt{n}} (1 - nx^2)dx > \frac{c_n}{\sqrt{n}} \ , \end{aligned}$$

whence $c_n < \sqrt{n}$.

For any $\delta > 0$,

$$q_n(x) < \sqrt{n}(1 - \delta^2)^n$$

when $\delta \leq x \leq 1$. Thus, the uniform limit of $\{q_n\}$ is 0 on any closed interval of $[-1, 1]$ that does not contain 0. We see that the graph of the q_n spikes more at the origin as n increases, while the area under the graph maintains the value 1. Another way of putting it, is that the "weight" of q_n is more and more concentrated at the origin.

We set

$$p_n(x) = \int_{-1}^{1} f(x + t)q_n(t)dt$$

for $0 \leq x \leq 1$. It is not clear at this stage that p_n is indeed a polynomial. However, noting that f vanishes off $[0, 1]$ and making a simple linear change of variables, we have that,

$$p_n(x) = \int_{-x}^{1-x} f(x + t)q_n(t)dt = \int_0^1 f(t)q_n(x - t)dt \ ,$$

which is manifestly a polynomial in x. Since the integral of q_n is 1, we can think of the integral defining p_n as a weighted average of the values of f; as n increases, more and more of the weight is concentrated near the value of t for which $x - t$ is zero, so that we are increasing the weight of values of f close to $f(x)$. All that remains are the epsilonics.

Let $\epsilon > 0$ be given. Because f is uniformly continuous, we can select $\delta > 0$ for which $|f(y) - f(x)| < \epsilon/2$ whenever $|y - x| < \delta$. Let $M = \sup |f(x)|$. Since, for each fixed $\delta > 0$,

$$\lim_{n\to\infty} \sqrt{n}(1-\delta^2)^n = 0 \ ,$$

we can select an integer N for which $\sqrt{n}(1-\delta^2)^n < \epsilon/(8M)$ for all $n > N$. Then, for $n > N$ and $0 \le x \le 1$, we have that

$$\begin{aligned}
|p_n(x) - f(x)| &= \left| \int_{-1}^{1} [f(x+t) - f(x)] q_n(t) dt \right| \\
&\le \int_{-1}^{1} |f(x+t) - f(x)| q_n(t) dt \\
&\le 2M \int_{-1}^{\delta} q_n(t) dt + \frac{\epsilon}{2} \int_{-\delta}^{\delta} q_n(t) dt + 2M \int_{\delta}^{1} q_n(t) dt \\
&\le 4M\sqrt{n}(1-\delta^2)^n + \frac{\epsilon}{2} < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon \ .
\end{aligned}$$

The desired result follows. ♠

§3. BERNSTEIN POLYNOMIALS

Another approach to the approximation problem is through Bernstein polynomials. If f is a continuous function, define the *Bernstein polynomial* of order n:

$$B(f, n; x) = \sum_{k=0}^{n} f\left(\frac{k}{n}\right) \binom{n}{k} x^k (1-x)^{n-k} \ .$$

For example, $B(\mathbf{1}, n; x) = \mathbf{1}\mathbf{1}$, $B(x, n; x) = x$ and $B(x^2, n; x) = [(n-1)x^2 + x]/n$ for each positive integer n.

It turns out that $f(x)$ is the uniform limit of $B(f, n; x)$ as n tends to infinity. This can be proved using a more general result of Korovkin [**4**]. Recall that a linear operator is positive iff Tf is everywhere positive when f is so. We have then that $f \le g$ implies $Tf \le Tg$. A positive linear operator T is bounded and $\|T\| = \|T\mathbf{1}\|_\infty$.

Theorem. *Let L_n be a positive linear operator from $C[a, b]$ into $C[a.b]$ and suppose that $L_n(f)$ converges uniformly to f when $f(x)$ is any one of the three functions $\mathbf{1}$, x, x^2. Then $L_n(f)$ converges to f uniformly for every continuous function f on $[a, b]$.*

Comments. If we consider $C[a, b]$, the space of all continuous functions (real or complex) on $[a, b]$ as a Banach space, then the positive linear operator L_n must be bounded. Positive simply means that $L_n(f) \ge 0$ whenever $f \ge 0$,

Lemma 1. *Let X be a compact metric space, and let α and β be positive continuous functions defined on X whose zero sets are $Z(\alpha) = \{x : \alpha(x) = 0\}$ and $Z(\beta) = \{x : \beta(x) = 0\}$. Suppose further that $Z(\beta) \subseteq Z(\alpha)$, and that ϵ is any positive number.*

Then there exists a positive number $M = M(\epsilon)$ such that

$$\alpha(x) \le \epsilon + M\beta(x)$$

for all $x \in X$.

Proof. Define the function γ for which $\gamma(x) = 0$ whenever $\beta(x) = 0$ and

$$\gamma(x) = \max\left\{\frac{\alpha(x) - \epsilon}{\beta(x)}, 0\right\}$$

when $\beta(x) > 0$. Then $\gamma(x)$ is a continuous function on the compact space X and so is bounded by some number M. The result follows. ♣

Before formulating the second lemma, we need two definitions. For a continuous function on X, define $\Delta(f) = \{(x,t) \in X \times X : f(x) = f(t)\}$. The continuous function γ defined on $X \times X$ is a *boundary function for* f if and only if $Z(\gamma) \subseteq \Delta(f)$. For each $t \in X$, we define the function γ_t on X by $\gamma_t(x) = \gamma(x,t)$.

Lemma 2. *Suppose that X is a compact metric space, that f is a continuous function on X and that γ is a boundary function for f. Suppose that L_n is a sequence of positive linear operators from $C(X)$ into $C(X)$ for which*

(i) $L_n(\mathbf{1}) \longrightarrow \mathbf{1}$ uniformly;

(ii) $L_n(\gamma_t)(t) \longrightarrow 0$ uniformly in t;

Then $L_n(f) \longrightarrow f$ uniformly on X.

Proof. Let $\alpha(x,t) = |f(x) - f(t)|$. Let $\epsilon > 0$. By Lemma 1 applied to α and γ, we can determine a constant M for which

$$|f(x) - f(t)| \le (\epsilon/2) + M\gamma(x,t)$$

for all (x,t) in $X \times X$. Let t be fixed and apply the positive linear operator L_n to obtain

$$|L_n(f)(x) - f(t)L_n(\mathbf{1})(x)| \le (\epsilon/2)L_n(\mathbf{1})(x) + ML_n(\gamma_t)(x)$$

for all $x \in X$. Now let $x = t$:

$$\begin{aligned}|L_n(f)(t) - f(t)| &\le |L_n(f)(t) - f(t)L_n(\mathbf{1})(t)| + |f(t)||L_n(\mathbf{1})(t) - 1| \\ &\le (\epsilon/2)L_n(\mathbf{1})(t) + ML_n(\gamma_t)(t) + \|f\|_\infty |L_n(\mathbf{1})(t) - 1| \ .\end{aligned}$$

We can now select N_ϵ such that, for $n > N_\epsilon$, the right side is less than ϵ. The proof can now be concluded. ♣

Proof of the Theorem. Now let $X = [a,b]$ and $\gamma(x,t) = (x-t)^2$. Then $\gamma_t(x) = (x-t)^2 = x^2 - 2tx + t^2$, so that

$$L_n(\gamma_t) = L_n(x^2) - 2tL_n(x) + t^2 L_n(\mathbf{1})$$

converges uniformly in X to $x^2 - 2tx + t^2 = (x-t)^2$.

Let $y, t \in [a,b]$ and $c = |a| + |b|$. Then

$$\begin{aligned}|L_n(\gamma_t)(y) - (y-t)^2| &= |(L_n(x^2)(y) - y^2) - 2t(L_n(x)(y) - y) + t^2(L_n(\mathbf{1})(y) - \mathbf{1}(y)| \\ &\le \|L_n(x^2) - x^2\|_\infty + 2c\|L_n(x) - x\|_\infty + c^2\|L_n(\mathbf{1}) - \mathbf{1}\|_\infty \ ,\end{aligned}$$

and the right side is less than $\epsilon > 0$ for n sufficiently large. If we take $y = t$, we find that $L_n(\gamma_t)(t) \to 0$ uniformly and the result follows from Lemma 2. ♠

Back to Bernstein polynomials. We apply Korovkin's Theorem, taking $[a,b] = [0,1]$, $\gamma(x,t) = (x-t)^2$ and $L_n(f) = B(f,n)$.

$$B((x-t)^2, n; x) = \left(1 - \frac{1}{n}\right)x^2 + \left(\frac{1}{n} - 2t\right)x + t^2 \ ,$$

so that

$$B(\gamma_t, n; t) = \left(1 - \frac{1}{n}\right)t^2 + \left(\frac{1}{n} - 2t\right)t + t^2 = \frac{t(1-t)}{n} \ .$$

Hence

$$|B(\gamma_t, n; t)| \le \frac{1}{4n} \ .$$

If follows that f is the uniform limit of $B(f,n)$, as $n \to \infty$.

One can see intuitively that the Bernstein functions might converge. Let

$$b_{n,k}(x) = \binom{n}{k} x^k (1-x)^{n-k} \ .$$

Taking the derivative, one can easily note that this function assumes its maximum value at $x = k/n$, and that as n increases, this maximum value becomes more pronounced as the function takes smaller values for x away from k/n. Thus, as n increases, the value $B(f, n; x)$ is a weighted average of the numbers $f(k/n)$ with the weight concentrated at those values of $f(k/n)$ for which k/n is close to x. In fact, a proof of the uniform convergence of $B(f, n; x)$ to $f(x)$ can be constructed in a way similar to that used for our first argument for the Weierstrass Approximation Theorem by splitting the sum

$$f(x) - B(f, n; x) = \sum_{k=0}^{n} (f(x) - f(k/n))b_{n,k}(x)$$

into three parts according as k/n is less than x and sufficiently far from it, close to x or greater than x and sufficiently far from it.

Another way to look at the Bernstein approximation is to think of the points $(k/n, f(k/n))$ in the plane as "control points" that are exactly given as points on the graph of the function $f(x)$, but which we wish to have, at least approximately, as points on the graph of a function that is easily calculated. This can be generalized to a set P of points p_k $(0 \leq k \leq n)$ in $d-$dimensional space $\mathbf{R}^n$. If we join these control points consecutively by line segments, we get a polygon, which of course will be nonsmooth at the points. However, we can use Bernstein polynomials to get a close and reasonably computable approximation to the curves that they suggest. We can define the *Bézier curve* in space parametrically by

$$B_n(P; t) = \sum_{k=0}^{n} p_k b_{n,k}(t) \ ,$$

This method of relating a consecutive set of points in space to a curve has many attractive features. It is affinely invariant: if we transform the set P to another set Q by an affine transformation, then the Bézier curve for Q is the same affine transformation of the Bézier curve for P. The Bézier curve will collapse to a point if and only if the points in P coincide. The points on the Bézier curve lie in the convex hull of P. The Bézier curve starts at p_0 and ends at p_n. [**5**, pp. 757-773]

§4. THE ABSOLUTE VALUE FUNCTION

As a special case, let us examine the approximation of the absolute value function $|x|$ on the interval $[-1, 1]$ by polynomials. One way to do this is through a binomial expansion. Let $t = 1-x^2$, so that $0 \leq t \leq 1$. Then

$$\begin{aligned} |x| &= (x^2)^{1/2} = [1 - (1 - x^2)]^{1/2} = (1 - t)^{1/2} \\ &= \sum_{n=0}^{\infty} \binom{1/2}{n} (-t)^n = 1 - \frac{1 - x^2}{2} - \sum_{n=2}^{\infty} \frac{(2n - 3)!(1 - x^2)^n}{2^{2n-2}(n - 2)!n!} \ . \end{aligned}$$

Since this series converges uniformly on $[-1, 1]$, its partial sums will provide a convergent approximating series of polynomials.

A second approach is through Fourier series. Every point on the interval $[-1, 1]$ can be represented in the form $\cos\theta$ for some value of θ in $[0, \pi]$. We recall that the Fourier series of a continuous function $f(\theta)$ defined on the closed interval $[-\pi, \pi]$ is given by

$$\frac{1}{2}a_0 + \sum_{k=0}^{\infty} (a_k \cos k\theta + b_k \sin k\theta) \ ,$$

where

$$a_n = \frac{1}{\pi} \int_{-\pi}^{\pi} f(\theta) \cos n\theta d\theta$$

and

$$b_n = \frac{1}{\pi} \int_{-\pi}^{\pi} f(\theta) \sin n\theta d\theta \ .$$

In general, the Fourier series need not converge to its parent function, even a continuous one. However, if f is not only continuous, but piecewise monotonic, we will have convergence.

We compute the Fourier series of $|\cos\theta|$ on the interval $[-\pi,\pi]$. Because $|\cos\theta|\sin k\theta$ is an odd function, $b_k = 0$ for each value of k. Because $|\cos\theta|\cos k\theta$ is even,

$$\begin{aligned}
a_k &= \frac{2}{\pi}\int_0^{\pi} |\cos\theta|\cos k\theta d\theta \\
&= \frac{2}{\pi}\left[\int_0^{\pi/2}\cos\theta\cos k\theta d\theta - \int_{\pi/2}^{\pi}\cos\theta\cos k\theta d\theta\right] \\
&= \frac{2}{\pi}\left[\int_0^{\pi/2}\cos\theta\cos k\theta d\theta + (-1)^k\int_0^{\pi/2}\cos\theta\cos k\theta d\theta\right] .
\end{aligned}$$

Thus $a_k = 0$ when k is odd, $a_0 = 4/\pi$ and a_k is equal to

$$\frac{2}{\pi}\int_0^{\pi/2}[\cos(k+1)\theta + \cos(k-1)\theta]d\theta$$

when k is positive and even. Thus, for $k = 2r$ with $r \geq 1$, we have that

$$a_{2r} = \frac{2}{\pi}(-1)^{r+1}\left(\frac{1}{2r-1} - \frac{1}{2r+1}\right) = \frac{4}{\pi}\left[\frac{(-1)^{r+1}}{4r^2-1}\right] .$$

Hence, the Fourier series for $|\cos\theta|$ is

$$\frac{4}{\pi}\left[\frac{1}{2} + \sum_{r=1}^{\infty}\frac{(-1)^{r+1}}{4r^2-1}\cos 2r\theta\right] .$$

Let us now relate this to functions on $[-1,1]$. By De Moivre's Theorem, for any nonnegative integer n, we find that

$$\cos n\theta + i\sin n\theta = (\cos\theta + i\sin\theta)^n = T_n(\cos\theta) + i\sin\theta U_n(\cos\theta).$$

where $T_n(x)$ and $U_n(x)$ are polynomials that satisfies the pair of recursions

$$T_{n+1}(x) = xT_n(x) + (x^2-1)U_n(x)$$

$$U_{n+1}(x) = T_n(x) + xU_n(x) .$$

This can be deduced from the equation

$$\cos(n+1)\theta + i\sin(n+1)\theta = (\cos\theta + i\sin\theta)(\cos n\theta + i\sin n\theta) .$$

We have that $T_0(x) = 1$, $U_0(x) = 0$, $T_1(x) = x$, $U_1(x) = 1$, $T_2(x) = 2x^2-1$, $U_2(x) = 2x$. The recursions have the form $\rho_{n+1} = M\rho_n$, where $\rho_n = (T_n(x), U_n(x))^{\mathbf{t}}$ and M is the matrix

$$\begin{pmatrix} x & x^2-1 \\ 1 & x \end{pmatrix} .$$

This matrix has the characteristic polynomial $\lambda^2 - 2x\lambda + 1$ and so satisfies $M^2 = 2xM - I$. Hence

$$T_{n+1}(x) = 2xT_n(x) - T_{n-1}(x)$$

and

$$U_{n+1}(x) = 2xU_n(x) - U_{n-1}(x) .$$

We call $T_n(x)$ the *Chebyshev polynomial* of degree n, or sometimes the *Chebyshev polynomial of the first kind* and $U_n(x)$ the *Chebyshev polynomial of the second kind* of degree $n-1$. Recalling our series expansion of $|\cos\theta|$ in terms of cosines of multiple angles, we obtain that

$$|x| = \frac{4}{\pi}\left(\frac{1}{2} + \sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{4k^2-1} T_{2k}(x)\right) .$$

If we define $p_n(x) = (4/\pi)((1/2) + \sum_{k=0}^{n}(-1)^{k+1}(4k^2-1)^{-1}T_{2k}(x)$, then, since $|T_{2k}(x)| \le 1$ on $[-1,1]$, we have that

$$\begin{aligned} \big||x| - p_k(x)\big| &= \frac{4}{\pi}\left| \sum_{k=n+1}^{\infty} \frac{(-1)^{k+1}}{4k^2-1} T_{2k}(x)\right| \\ &\le \frac{4}{\pi} \sum_{k=n+1}^{\infty} \frac{1}{4k^2-1} = \frac{2}{\pi(2n-1)} . \end{aligned}$$

Hence $|x|$ is the uniform limit of the sequence $\{p_n\}$ of polynomials on $[-1,1]$.

There is a famous conjecture due to Bernstein concerning $|x|$ that was recently settled. For any continuous real function on $[a,b]$, define the *modulus of continuity* $\omega(\delta; f)$ for $\delta > 0$ by

$$\omega(\delta; f) = \sup\{|f(x) - f(y)| : |x-y| \le \delta\} .$$

Let

$$E_n(f) = \inf\{\|f-p\|_\infty : p \text{ a polynomial of degree not exceeding } n\} .$$

Then, it is known that

$$E_n(f) \equiv E_n(f : [-1,1]) \le 6\omega(1/n; f)$$

so that f can be uniformly approximated by polynomials. In the case of $f(x) = |x|$, the above calculation shows that

$$2nE_{2n}(|x|) < \frac{2}{\pi} .$$

In 1913, Bernstein proved that there is a positive constant β such that

$$\lim_{n\to\infty} 2nE_{2n}(|x|) = \beta$$

with $0.278 < \beta < 0.286$, and conjectured that $\beta = 1/(2\sqrt{\pi})$. This was settled negatively in 1985 by Varga and Carpenter. [**5**, pp. 749-757]

§5. THE STONE-WEIERSTRASS THEOREM

The Weierstrass Approximation Theorem has a beautiful generalization that is worth presenting. We need a few definitions. Let X be a space. A set of real (or complex) functions defined on X is said to be an *algebra* if and only if it is a linear space that is also closed under multiplication. A family of functions *separates the points of* X if and only if, given two distinct points x_1 and x_2 of X, there is a function in the family that takes different values at these points.

The Stone-Weierstrass Theorem. *Suppose that X is a compact Hausdorff topological space and $\mathbf{A}$ is an algebra of continuous real-valued functions defined on X that separates points and for which there is no point in X at which all the functions in $\mathbf{A}$ vanish. Then the uniform closure of $\mathbf{A}$ is the set of all continuous real functions on X.*

The proof of this has a number of steps:

I. For any pairs (x_1, x_2) of points of X and (c_1, c_2) of real numbers, there is a function $f \in \mathbf{A}$ for which

$$f(x_1) = c_1 \qquad \text{and} \qquad f(x_2) = c_2 .$$

II. Let $\mathbf{B}$ be the uniform closure of $\mathbf{A}$. Then, if $f \in \mathbf{B}$, then so also does $|f| \in \mathbf{B}$. To see this, suppose that the range of f lies inside a closed interval I. By the Weierstrass Approximation Theorem, we can find a sequence $\{p_n(t)\}$ of polynomials that converges uniformly to $|t|$ on I. Then the sequence $\{p_n \circ f\}$ converges uniformly to $|f|$.

III. If $f, g \in \mathbf{B}$, then max (f,g) and min (f,g) both belong to $\mathbf{B}$. This follows from the identities

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

and

$$2 \min (f,g) = (f+g) - |f-g| \ .$$

This closure extends to finite maxima and minima of functions.

IV. For any $\epsilon > 0$ and any real-valued continuous function f on X and any point $x \in X$, there exists a function g_x for which $g_x(x) = f(x)$ and $g_x(y) > f(y) - \epsilon$ for all $y \in X$. [By step I, given any $y \neq x$, we can find a function $h_y \in \mathbf{B}$ for which $h_y(x) = f(x)$ and $h_y(y) = f(y)$. There exists an open neighbourhood U_y of y for which $h_y(t) > f(t) - \epsilon$ for $t \in U_y$. There is a similar open neighbourhood U_x of x. Use the compactness of X to cover X with finitely many such neighbourhoods and take the maximum of the corresponding h_y.]

V. For any $\epsilon > 0$ and any real-valued continuous function f on X, there is a function $h \in \mathbf{B}$ for which $|f(x) - h(x)| < \epsilon$. [For each $x \in X$, construct the function g_x as in step IV, and select an open neighbourhood V_x for which $g_x(y) < f(y) + \epsilon$ for all $y \in V_x$. Use the compactness of X to find a finite cover of neighbourhoods V_x and let h be the minimum of the corresponding g_x.] The theorem now follows. ♠

§6. NEAREST APPROXIMANT.

Now that we know that every continuous function f on a closed interval can be uniformly approximated by polynomials, the question arises as to what is the best approximant to f by polynomials whose degree is less than n. One way of detecting this best approximant is through an *alternation property*.

Suppose that f is a continuous function on $[a,b]$ and that p is a polynomial of degree less than n. Suppose that $K = \sup\{|f(x) - p(x)| : a \leq x \leq b\}$ and that there are $n+1$ points a_i in increasing order in $[a,b]$ for which $|f(a_i) - p(a_i)| = K$ $(0 \leq i \leq n)$. Suppose further that $f(a_i) - p(a_i)$ alternate in sign, so that $f(a_i) - p(a_i) = -(f(a_{i-1}) - p(a_{i-1}))$ for $1 \leq i \leq n$.

Suppose that $q(x)$ is a polynomial for which $\|f - q\|_\infty < K$. Then, by checking a diagram, we can see that the graphs of $f(x) - p(x)$ and $f(x) - q(x)$ must cross in each of the n intervals $[a_{i-1}, a_i]$ $(1 \leq i \leq n)$, so that $p(x) - q(x)$ is a nonzero polynomial with n roots, one in each interval $[a_{i-1}, a_i]$. Thus, $p(x) - q(x)$ and, hence, $q(x)$ are polynomials of degree at least n. Therefore, $p(x)$ is the best approximant to $f(x)$ from among the polynomials of degree less than n, with respect to the uniform norm.

The case of the function x^n approximate on $[-1,1]$ be polynomials of degree less than n is of particular interest. Using the alternation property, we see that the best approximant of x by constants is the function $\mathbf{0}$, and the best approximant of x^2 by polynomials of degree not exceeding 1 is the constant function $\mathbf{1/2}$.

In general, if for x^n, we can find a polynomial $p(x)$ of degree less than n for which $x^n - p(x)$ assumes its maximum absolute value $n+1$ times in $[-1,1]$ with the alternation property, then $p(x)$ must be the best uniform approximant to x^n among polynomials of degree less than n. It is straightforward to construct a polynomial of degree not exceeding n to serve as $x^n - p(x)$.

For nonnegative integers n, let $T_n(x)$ be the function introduced in Section 5.4, to wit

$$T_n(x) = \cos(n \arccos x)$$

for $-1 \leq x \leq 1$. With $\theta = \arccos x$, it is a consequence of $\cos n\theta + i \sin n\theta = (\cos\theta + i \sin\theta)^n$ and $\sin^2\theta = 1 - \cos^2\theta$ that $\cos n\theta$ is a polynomial of degree n in $\cos\theta$; thus $T_n(x)$ is a polynomial of degree n.

Indeed,

$$\begin{aligned} T_n(x) &= \frac{1}{2}[(x+\sqrt{x^2-1})^n + (x-\sqrt{x^2-1})^n] \\ &= \frac{n}{2}\sum_{k=0}^{\lfloor n/2 \rfloor}(-1)^k\frac{(n-k-1)!}{k!(n-2k)!}(2x)^{n-2k} = 2^{n-1}x^n + \cdots \\ &= \sum_{k=0}^{\lfloor n/2 \rfloor}\binom{n}{2k}x^{n-2k}(x^2-1)^k \ . \end{aligned}$$

Since $|\cos n\theta| = 1$ when $n\theta$ is an even multiple of π, $T_n(x)$ assumes its maximum and minimum values with alternating signs when $x = \cos(k\pi/n)$ for $k = 0, 1, \cdots, n$, *i.e.* $n+1$ times. Thus the polynomial $x^n - 2^{1-n}T_n(x)$ is that polynomial of degree $n-1$ that best approximates x^n uniformly.

The Chebyshev polynomials have a number of interesting properties that are listed in (**1**, 29-41). They constitute a family of polynomials that commute under composition. Moreoever, according to a result of Block and Thielman, if $\{p_n\}$ is a sequence of polynomials for which p_n has degree n for $n = 1, 2, 3, \cdots$, and $p_n \circ p_m = p_{mn}$ for all n, m, then there is a linear polynomial $q(x)$ for which either $q^{-1} \circ p \circ q = x^n$ or $q^{-1} \circ p \circ q = T_n$. The generating function for the Chebyshev polynomials is

$$\frac{1-tx}{1-2tx+t^2} = \sum_{k=0}^{\infty} T_n(x)t^n$$

and they satisfy the differential equation

$$(1-x^2)T_n''(x) - xT_n'(x) + n^2T_n(x) = 0 \ .$$

§7. ORTHOGONALITY AND APPROXIMATE INTEGRATION

The familiar orthogonality relation

$$2\int_0^{\pi} \cos m\theta \cos n\theta d\theta = \int_{-\pi}^{\pi} \cos m\theta \cos n\theta d\theta = \pi\delta_{m,n}$$

can be transformed by the subsitution $x = \cos\theta$ to the relations

$$\frac{2}{\pi}\int_{-1}^{1}\frac{T_n(x)T_m(x)}{\sqrt{1-x^2}}dx = \delta_{m,n} \ .$$

This suggests another context in which to consider the approximation problem, that of inner product space. The background theory is reviewed in (**2**, 41-56). Let $w(x)$ be a nonnegative Lebesgue integrable function on $[a, b]$ that is positive almost everywhere. The space $C([a, b])$ of all continuous complex functions on the interval $[a, b]$ can be equipped with the inner product

$$\langle f, g \rangle = \int_a^b \overline{f(x)}g(x)w(x)dx \ .$$

The norm $\|f\|_w$ is given by

$$\|f\|_w = \sqrt{\int_a^b |f(x)|^2 w(x)dx}$$

and the completion of $C([a, b])$ with respect to this norm is denoted by $L_2(w; [a, b])$ or just $L_2(w)$ when the underlying interval is understood.

Once the weight function is given, we can start with the linearly independent set $\{1, x, x^2, \cdots, x^n, \cdots\}$ of polynomials on $[a, b]$ and use the Gram-Schmidt orthogonalization process to determine an on orthogonal (or orthonormal) set of real polynomials that will serve as a basis for the space of polynomials. We can then complete the space with respect to the norm induced by w. If $\{p_n\}$ is such a basis, with the degree of p_n equal to n, then we have that $p_n(x) = \gamma_n x^n + r_{n-1}(x)$ for some polynomial r_{n-1} of degree less than n and constant γ_n. From this, we can derive a recursion of the form

$$xp_n(x) = a_n p_{n+1}(x) + b_n p_n(x) + a_{n-1}p_{n-1}(x) \tag{$*$}$$

where $p_{-1}(x) \equiv 0$, $a_{-1} = 0$ and $a_n = \gamma_n/\gamma_{n+1}$. To see this, note that the degree of $xp_n(x)$ is $n+1$, so that it admits a representation of the form

$$xp_n(x) = \sum_{k=0}^{n+1} d_k p_k(x) \ .$$

Then $\langle xp_n, p_k\rangle = \langle p_n, xp_k\rangle = 0$ for $0 \leq k \leq n-2$, from which we have that $d_k = 0$ for $0 \leq k \leq n-2$. Comparing the leading coefficient on the two sides of the equation $(*)$, we find that $\gamma_n = d_{n+1}\gamma_{n+1}$. The remaining relation is a little more complicated and can be derived from taking the inner product of the equation with p_{n-1}.

Quite a bit more can be said about the polynomials $p_n(x)$. Noting that $p_0(x) = 1$ and $0 = \langle p_n, p_0\rangle = \int_a^b p_n(x)w(x)dx$, we see that the polynomial $p_n(x)$ must change sign at least once in the open interval (a, b). It cannot change signs more than n times because of its degree. Let $\{r_i : 1 \leq i \leq m\}$ be the set of all points within (a, b) where the polynomial $p_n(x)$ changes sign, indexed in increasing order of magnitude. Since $p_n(x)(x-r_1)\cdots(x-r_m)$ is a polynomial of constant sign on (a, b), $\int_a^b p_n(x)(x-r_1)\cdots(x-r_m)w(x)dx \neq 0$. This entails that $(x-r_1)\cdots(x-r_m)$ cannot lie within the linear span of the set $\{p_0, p_1, \cdots, p_{n-1}\}$ and so it must have degree n. Therefore, $m = n$ and we deduce that all the zeros of $p_n(x)$ are simple and lie within the open interval (a, b).

Furthermore, if we let $a = r_0$ and $b = r_{n+1}$, then it can be shown that each open interval (r_i, r_{i+1}) $(0 \leq i \leq n)$ contains a zero of p_{n+1}.

The zeros r_i of $p_n(x)$ figure in an approximate integration formula.

Theorem. *Let n be a positive integer and let $r_1, r_2, \cdots, r_n$ be the zeros of $p_n(x)$. Then there are real numbers λ_1, λ_2, $\cdots$, λ_n for which*

$$\int_a^b f(x)w(x)dx = \lambda_1 f(r_1) + \lambda_2 f(r_2) + \cdots + \lambda_n f(r_n) \ ,$$

for every polynomial $f(x)$ of degree not exceeding $2n-1$.

Proof. For $1 \leq i \leq n$, let

$$q_i(x) = \frac{(x-r_1)\cdots\widehat{(x-r_i)}\cdots(x-r_n)}{(r_i-r_1)\cdots(r_i-r_{i-1})(r_i-r_{i+1})\cdots(r_i-r_n)} = \frac{p_n(x)}{p_n'(r_i)(x-r_i)} \ .$$

Suppose that $f(x)$ is a polynomial of degree not exceeding $2n-1$. Then the Lagrange Polynomial of degree less than n that agrees with $f(x)$ at $x = r_i$ $(1 \leq i \leq n)$ is $g(x) = \sum_{i=1}^n f(r_i)q_i(x)$. Therefore

$$f(x) - g(x) = p_n(x)s(x)$$

where $s(x)$ is a polynomial of degree not exceeding $n-1$. Since $s(x)$ is in the linear span of $\{p_i(x) : 0 \leq i \leq n-1\}$, $\int_a^b p_n(x)s(x)w(x)dx = 0$. Therefore

$$\int_a^b f(x)w(x)dx = \int_a^b g(x)w(x)d(x) = \sum_{i=1}^n f(r_i)\int_a^b q_i(x)w(x)dx \ .$$

Example. Let $[a,b]=[-1,1]$ and $w(x)=1$. Then

$$(p_0(x),p_1(x),p_2(x),p_3(x))=\left(1,x,x^2-\frac{1}{3},x^3-\frac{3}{5}x\right)$$

and

$$\int_{-1}^{1} f(x)dx = 2f(0) \qquad \text{for deg } f(x)\le 1\ ;$$

$$\int_{-1}^{1} f(x)dx = f(-1/\sqrt{3})+f(1/\sqrt{3}) \qquad \text{for deg } f(x)\le 3\ ;$$

and

$$\int_{-1}^{1} f(x)dx = \frac{5}{9}f\left(-\sqrt{\frac{3}{5}}\right)+\frac{8}{9}f(0)+\frac{5}{9}f\left(\sqrt{\frac{3}{5}}\right) \qquad \text{for deg } f(x)\le 5\ .$$

In fact, if we know the formula for the particular interval $[-1,1]$, we can obtain the formula for a general interval $[a,b]$ using

$$\int_a^b f(x)dx = \frac{b-a}{2}\int_{-1}^{1} f\left(\frac{1}{2}a(1-t)+\frac{1}{2}b(1+t)\right)dt\ .$$

This device helps us to obtain a generalization of the trapezoidal and Simpson's rules which give an exact value of the integral over an interval in terms of values at $n+1$ points for polynomials of degrees up to $2n-1$. The condition that two of the evaluation points be endpoints of the interval requires the insertion of an additional evaluation point.

Recall that the trapezoidal rule

$$\int_a^b f(x)d(x) = \frac{b-a}{2}(f(a)+f(b))$$

holds for polynomials of degrees not exceeding 1 and that Simpson's Rule

$$\int_a^b f(x)dx = \frac{b-a}{6}\left(f(a)+4f\left(\frac{a+b}{2}\right)+f(b)\right)$$

holds for polynomials of degrees not exceeding 3. At the next stage, we consider the interval $[-1,1]$ and exploit the symmetry to seek a formula of the form

$$\int_{-1}^{1} f(x)dx = uf(-1)+vf(-w)+vf(w)+uf(1)$$

where u,v,w are real numbers with $0<w<1$. This formula always holds for $f(x)=x^k$ when k is odd, and imposing it on $f(x)=x^k$ when $k=0,2,4$ leads to

$$1=u+v \qquad \frac{1}{3}=u+vw^2 \qquad \frac{1}{5}=u+vw^4\ .$$

Hence

$$2=3v(1-w^2)=15vw^2(1-w^2)$$

from which $w=1/\sqrt{5}$ and $(u,v)=(\frac{1}{6},\frac{5}{6})$. Thus, for polynomials $f(x)$ of degree not exceeding 5,

$$\int_{-1}^{1} f(x)dx = \frac{1}{6}[f(-1)+5f(-1/\sqrt{5})+5f(1/\sqrt{5})+f(1)]\ .$$

This in turn yields that, for such polynomials,

$$\int_a^b f(x)dx = \frac{b-a}{12}\left[f(a) + 5f\left(\left(\frac{5+\sqrt{5}}{10}\right)a + \left(\frac{5-\sqrt{5}}{10}\right)b\right) + 5f\left(\left(\frac{5-\sqrt{5}}{10}\right)a + \left(\frac{5+\sqrt{5}}{10}\right)b\right) + f(b)\right] .$$

§8. PROBLEMS AND INVESTIGATIONS

1. Suppose that a and b are positive integers. Prove that, if

$$\frac{a^2+b^2}{ab+1}$$

is an integer, then it is a square.

2. Let P_n be the vector space of real polynomials of degree not exceeding n and C be the vector space of real continuous funtions, both defined on the closed unit interval $[0,1]$. With respect to the sup-norm

$$\|f\|_\infty \equiv \sup\{|f(x)| : 0 \leq x \leq 1\} ,$$

C is a complete normed linear space and P_n a closed linear subspace of dimension n. Let $0 = a_0 < a_2 < \cdots < a_n = 1$ and define, for $f \in C$, $L_n(f)$ to be that polynomial whose values equal those of f at a_0, a_1, $\cdots$, a_n. The norm of L_n is defined by

$$\|L_n\| = \sup\{\|L_n(f)\|_\infty : \|f\|_\infty \leq 1\} .$$

What can be said about the norm $\|L_n\|$?

3. The $n+1$ Bernstein polynomials of degree n are defined by

$$b_{n,k}(x) = \binom{n}{k} x^k (1-x)^{n-k}$$

for $0 \leq k \leq n$. When all $n+1$ polynomials are plotted on the same graph for large fixed n over the interval $0 \leq x \leq 1$, an "upper envelope" begins to be seen. Let

$$\beta(x) = \lim_{n\to\infty} \sqrt{n} \max\{b_{n,k}(x) : 0 \leq k \leq n\} .$$

Find a closed form expression for $\beta(x)$.

Hints and comments

1. This is an International Mathematical Olympiad problem. A simpler version appeared on the 1998 Canadian Mathematical Olympiad. In that, one was required to show that, if the sequence $\{a+n\}$ was defined recursively by $a_0 = 0$, $a_1 = m$ and $a_{n+1} = m^2 a_n - a_{n-1}$ for $n \geq 1$, then the positive integers a, b satisfy $(a^2+b^2)/(ab+1) = m^2$ if and only if a and b are two consecutive terms of the sequence. To solve this problem, we use a descent argument. Let $a \geq b$ and $f(a,b) = (a^2+b^2).(ab+1)$. Suppose that $a = nb - r$. Then $n - 1 < f(a,b) < n$, so that, if $f(a,b)$ is an integer, then $f(a,b) = n$ and $f(b,r) = n$. Thus, we can define $a_k > \cdots > a_2 > a_1 > 0$ with $a_k = a$, $a_{k-1} = b$, $a - i + 1 = na_i - a_{i+1}$ for $2 \leq i \leq k-1$, and $a_2 = na_1$. This can be related to the Chebshev polynomials in the following way. If $V_n(x) = U_n(x/2)$, then $f(a,b)$ is an integer if and only if $n = m^2$ and

$$(a,b) = (mv_k(m^2), mv_{k-1}(m^2)) .$$

4. $\beta(x) = [2\pi x(1-x)]^{-1/2}$. [AMM #10990: 110:1 (January, 2003), 59; 111:9 (November, 2004), 825-826.]

References

1. Edward J. Barbeau, *Pell's equation.* Springer, 2003
2. Peter Borwein, Tamás Erdélyi, *Polynomials and polynomial inequalities.* Springer, 1995
3. Konrad Knopp, *Theory and application of infinite series.* Dover, 1957, 1989.
4. H.E. Lomelí & C.L. Garciá, Variations on a theme of Korovkin. *Amer. Math. Monthly* 113:8 (Oct., 2006), 744-749
5. G.V. Milovanović, D.S. Mitrinović & Th. M. Rassias, *Topics in polynomials: extremal problems, inequalities, zeros.* World Scientific, 1994.
6. L.M. Milne-Thomson, *The Calculus of Finite Differences.* Macmillan, London, 1933
7. N. E. Norlund, *Leçons sur les séries d'interpolation.* Gauthier-Villars, Paris, 1926
8. Walter Rudin, *Principles of mathematical analysis.* Third edition. McGraw-Hill, 1964, 1976
9. Gabor Szegö, *Orthogonal polynomials.* Collloquium 23, American Mathematical Society, 1939 Chapter III