

APPROXIMATION OF FUNCTIONS

Ronald A. DeVore

Approximation Theory began at the end of the last century with the study of the approximation of functions by polynomials and rational functions. It is a broad subject which interacts with various aspects of real, complex and functional analysis. Some of its recent popularity comes from its importance in the development of numerical algorithms and the solution of problems of optimization.

One hundred years ago, Weierstrass proved his famous theorem on the approximation of continuous functions by algebraic polynomials. Undoubtedly, everyone of you has seen this theorem. But, in order to guarantee that we are all starting at the same point, let's begin with a formulation of this theorem which uses the notation of Approximation Theory.

We want to approximate functions f which are continuous on an interval $I := [a, b]$. We let $C(I)$ denote the set of all such functions and let

$$\|f\| := \sup_{x \in I} |f(x)|,$$

be its norm. We are interested in approximating f by algebraic polynomials $P(x) = a_0 + a_1x + \dots + a_nx^n$ of degree at most n . If Π_n denotes the set of all such polynomials, we let $E_n(f)$ be the error of approximation to f from Π_n :

$$(0.1) \quad E_n(f) := \inf_{P \in \Pi_n} \|f - P\|.$$

With this, we have

THEOREM 0.1. (Weierstrass [W]) If $f \in C(I)$, then $E_n(f) \rightarrow 0$ as $n \rightarrow \infty$.

In other words, each continuous function can be approximated arbitrarily well in the uniform norm by polynomials. There are many wonderful proofs of

Weierstrass' theorem. We shall give one of these a little later (§7).

Important theorems often open more doors than they close. This is certainly the case with the Weierstrass theorem. Now that we know polynomial approximation is possible, we are confronted with questions like:

Are there polynomials $P^* \in P_n$ which attain the infimum in (0.1)?

If such a P^* exists, is it unique?

How can we calculate P^* ?

Can we say anything about how fast $E_n(f)$ tends to zero?

These fundamental questions about polynomial approximation were studied around the turn of the century and their solution forms the foundation of Approximation Theory. It is appropriate therefore in a course on approximation, whether it be short or long, that we begin with a look at the solution to these questions.

§1. Best Approximation. The polynomials P^* are called polynomials of best approximation to f of degree n . Let us begin with the questions of existence and uniqueness of the P^* . These can be discussed in the following more general setting. We have a normed linear space X , $\|\cdot\|$ and one of its finite dimensional subspaces Y . We are interested in approximating $x \in X$ by the elements Y . For this, we introduce the distance $d(x, Y)$ of x to Y :

$$(1.1) \quad \text{dist}(x, Y) := \inf_{y \in Y} \|x - y\|.$$

If $y^* \in Y$ assumes the infimum in (1.1), then we say that y^* is a best approximation to \bar{x} from Y and we let $B(x)$ denote the collection of all such best approximants. A simple but useful remark is that $B(\alpha x) = \alpha B(x)$ for any scalar α . It is rather easy to see that best approximants always exist.

THEOREM 1.1. For each $x \in X$, $B(x) \neq \emptyset$.

Proof. The infimum in (1.1) can be restricted to those $y \in Y$ with $\|y\| \leq 2\|x\|$. Indeed any other y gives $\|x - y\| \geq \|y\| - \|x\| \geq \|x\|$, so that 0 is a better approximation than y . Now the closed ball $\|y\| \leq 2\|x\|$ is a compact set (because Y is finite dimensional) [L, p.16], hence the continuous function $\|x - y\|$ attains its infimum on this set. In this way, we see that best approximants exist.

A more subtle question is to decide when best approximation is unique, that is, when does $B(x)$ have just one element? This depends very much on the

geometry in the space X , namely, on the shape of the unit ball $U := \{x : \|x\| \leq 1\}$. This ball is always convex: if $x, x' \in U$, then the line segment $[x, x'] := \{z \in X : z = \alpha x + (1-\alpha)x', \alpha \geq 0\}$ is contained in U . We say U is strictly convex if

$$(1.2) \quad \|\alpha x + (1-\alpha)x'\| < 1, \text{ for all } x, x' \in U, x \neq x', \text{ and all } 0 < \alpha < 1.$$

Strict convexity means that the interior of the line segment $[x, x']$ is contained in the interior of U , for all $x, x' \in U$.

THEOREM 1.2 If the unit ball U of X is strictly convex, then best approximation from a finite dimensional space Y is unique, that is, $B(x)$ is a singleton for all $x \in X$.

Proof. Suppose $y, y' \in B(x)$:

$$\|x - y\| = \|x - y'\| = \text{dist}(x, Y).$$

If $\text{dist}(x, Y) = 0$, then $x = y = y'$ as desired. Otherwise, we can rescale (that is multiply x, y, y' by the same constant) so that $\text{dist}(x, Y) = 1$. Then $\frac{1}{2}(y + y')$ is in Y . If $y \neq y'$, then from (1.2),

$$\|x - \frac{1}{2}(y + y')\| < \text{dist}(x, Y),$$

which is an obvious contradiction. Hence $y = y'$ and best approximation is unique.

The most important normed spaces X with strictly convex unit balls are the $L_p(I)$ spaces. This space consists of all Lebesgue integrable functions on I for which the following norm is finite:

$$(1.3) \quad \|f\|_p(I) := \left\{ \int_I |f(x)|^p dx \right\}^{1/p}, \quad 1 \leq p < \infty.$$

When $p = \infty$, the right side of (1.3) is replaced by the essential supremum of f . The spaces L_p are strictly convex when $1 < p < \infty$. This is proved by examining when equality can hold in the triangle inequality for the L_p norm.

From the strict convexity of the L_p spaces, $1 < p < \infty$, it follows that any function $f \in L_p(I)$ has a unique best approximation P^* from Π_n .

Analogous to the L_p are the ℓ_p^n spaces which consist of all sequences $\underline{x} = (x_i)_1^m$ for which the norm

$$\|\underline{x}\|_p := \left(\sum_{i=1}^m |x_i|^p \right)^{1/p}, \quad 1 \leq p < \infty$$

$$\max_{1 < i < m} |x_i|, \quad p = \infty$$

is finite. These spaces also have strictly convex unit ball if $1 < p < \infty$ but not when $p=1$ or ∞ . This is easily seen when $m=2$ where the boundaries of the unit balls are depicted in Figure 1 for the values $p=1, 2, \infty$.

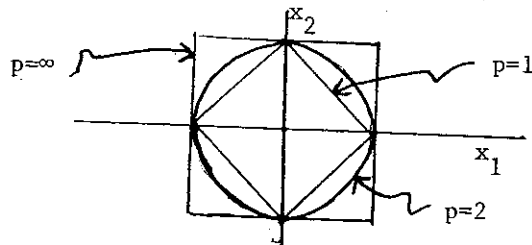


Figure 1
Unit balls of ℓ_p^2 for $p=1, 2, \infty$

Unfortunately, the space $C(I)$, of most immediate interest to us, does not have a strictly convex unit ball. For example, the functions $\phi(x)=x$ and $\psi(x)=x^2$ each have norm one in $C[0,1]$ but $\frac{1}{2}(\phi+\psi)$ also has norm one. Even more to the point, it is easy to construct finite dimensional subspaces Y of $C[0,1]$ from which best approximation is not unique. Consider, for example, the space $Y := \text{span}\{\phi, \psi\}$. Any non-negative function in Y of norm at most two is a best approximation to $f(x) := 1$.

Nevertheless, the situation is not as bad as it seems. It was shown by P.L. Chebyshev that each $f \in C(I)$ has a unique best approximation from Π_n . The special properties of Π_n that make this true is the next subject for discussion.

2. Chebyshev's Theorem. This theorem gives that best approximation P^* from Π_n to a continuous function f is unique. To prove this theorem, P.L. Chebyshev analyzed the behavior of the error function $E(x) := f(x) - P^*(x)$. He showed that there are many points where E alternately takes on the values $\pm \|E\|$.

THEOREM 2.1 (Chebyshev) If $f \in C(I)$ and P^* is its best approximation from Π_n , then there are points x_0, \dots, x_{n+1} , and a value $\eta = \pm 1$ such that

$$(2.1) \quad E(x_i) = (-1)^i \eta \|E\|, \quad i=0, \dots, n+1.$$

Hence, this theorem shows that there are at least $n+2$ points where the error E alternately takes on its maximum ($\|E\|$) and its minimum ($-\|E\|$).

Proof of Theorem 2.1. Let x_0 be the first point x (from the left) on I where $|E(x)| = \|E\|$; such a point exists because E is continuous. Then, let x_1 be the first point $x > x_0$ where $E(x) = -E(x_0)$. Continuing in this way, we create a sequence of points $x_0 < \dots < x_m$ where E alternately takes on the values $\pm \|E\|$.

We claim that $m > n$, as desired. Indeed, if $m \leq n$, then because of the continuity of E , we can find points ξ_1, \dots, ξ_m with $x_0 < \xi_1 < x_1 < \dots < \xi_m < x_m$ such that $[\xi_i, \xi_{i+1}]$ contains no points x with $E(x) = -E(x_i)$. Now, for a proper choice of γ , the polynomial $P(x) := \gamma(x - \xi_1) \dots (x - \xi_m)$ is of degree $\leq n$ and agrees in sign with E at each of the points x_0, \dots, x_m . If $E(x_i) > 0$, then $P(x) > 0$ on (ξ_i, ξ_{i+1}) and also $E(x) > -E(x_i)$ on $[\xi_i, \xi_{i+1}]$. Hence, for $\eta > 0$ sufficiently small, $|E(x) - \eta P(x)| < ||E||$, for $x \in [\xi_i, \xi_{i+1}]$. The same is true when $E(x_i) < 0$ and also on the end intervals. Since there are only a finite number of intervals we can choose one $\eta > 0$ and obtain $||f - \eta P|| < ||E||$. But then, this means that $P + \eta P$ is a better approximation to f than P . This contradiction means that $m > n$ and proves Chebyshev's theorem.

From the Chebyshev alternation theorem, it is easy to prove the uniqueness of best polynomial approximants.

THEOREM 2.2. If f is in $C(I)$, then f has a unique best approximation $P \in \Pi_n$.

Proof. If P_1 and P_2 are two best approximants to f from Π_n then so is $P := \frac{1}{2}(P_1 + P_2)$. Let x_0, \dots, x_n be alternation points for $f - P$. Then,

$$(2.2) \quad \frac{1}{2}(f(x_i) - P_1(x_i)) + \frac{1}{2}(f(x_i) - P_2(x_i)) = f(x_i) - P(x_i) = \pm ||E||.$$

Since $||f - P_1|| \leq ||E||$, and likewise for $f - P_2$, the only way that (2.2) can hold is if $f(x_i) - P_1(x_i) = f(x_i) - P_2(x_i)$. That is $P_1(x_i) = P_2(x_i)$, $i = 0, \dots, n$. This means that the polynomial $P_1 - P_2$ which is of degree $\leq n$ has $n+1$ zeros and hence must be the zero polynomial. Therefore, $P_1 = P_2$.

Actually the proof of Theorem 2.2 shows more. Namely, we have the following Chebyshev characterization of when a polynomial P is the best approximation to f .

THEOREM 2.3. If $f \in C(I)$ and $P \in \Pi_n$ are such that $f - P$ alternately takes on the values $\pm M$ at least $n+2$ times with $M := ||f - P||$, then $P = P^*$ is the best approximation to f and $M = E_n(f)$.

Proof. Let x_i , $i = 0, \dots$ be the alternation points of $f - P$. If Q is any other polynomial with $||f - Q|| < M$, then $Q - P = f - P - (f - Q)$ has the same sign as $f - P$ at each of the x_i . Hence $Q - P$ has at least $n+1$ zeros and $Q = P$. This is the desired contradiction.

In view of this theorem, the search for the best approximation is reduced to finding a polynomial P such that $f - P$ has sufficiently many alternations.

What are the essential properties of polynomials which were used in the above proof of uniqueness? Well, in the Chebyshev alternation theorem, we constructed a polynomial which changed sign precisely at the points ξ_1, \dots, ξ_m , and in the proof of uniqueness we used the fact that any non-trivial polynomial of degree n has at most n zeros. There are other $n+1$ dimensional subspaces X_{n+1} of $C(I)$ which have these properties. They are called Haar spaces and any basis ϕ_0, \dots, ϕ_n for X_{n+1} is called a Haar system (sometimes called a Chebyshev system).

DEFINITION 2.4. An n dimensional subspace X_n of $C(I)$ is called a Haar subspace if each function $\phi \in X_n$ has at most $n-1$ zeros on I unless ϕ is identically zero.

Of course, Π_n are the most important Haar spaces. Some other interesting examples are the span of the exponentials $e^{\alpha_i x}$, $i=1, \dots, n$ or the span of the power functions x^{α_i} , $i=1, \dots, n$. Here, $\alpha_1, \dots, \alpha_n$ can be any non-negative real numbers.

Haar spaces are much like polynomial spaces. For example, we have.

THEOREM 2.5. If X_n is a Haar space and $f \in C(I)$, then f has a unique best approximation from X_n .

The proof is essentially the same as that given above for polynomials except that now one has to work much harder to show that there is a function ϕ which changes sign at any prescribed points ξ_1, \dots, ξ_m , $m < n$, from the interior of I .

Remarkably, the notion of Haar space actually characterizes the Chebyshev spaces of $C(I)$ (i.e. the spaces from which best approximation is unique). Indeed, we have the following theorem of Haar.

THEOREM 2.6. If every $f \in C(I)$ has a unique best approximation from the subspace X_n , then X_n is a Chebyshev space.

A proof of this theorem can be found in the book of Lorentz [L]. Haar systems are important in many fields other than approximation. The interested reader should consult the seminal paper of Krein [Kr] or the book of Karlin and Studden [K-S].

3. Trigonometric polynomial approximation. It is not necessary for the interval I to be closed in the definition of a Haar system. In fact, one of the most important Haar systems is the space τ_n of trigonometric polynomials of degree $\leq n$. A trigonometric polynomial of degree n is an expression of the form

$$T(x) = a_0 + \sum_1^n (a_k \cos kx + b_k \sin kx)$$

with $a_n^2 + b_n^2 > 0$. Any trigonometric polynomial T of degree n has at most $2n$ zeros on $[0, 2\pi)$. Hence τ_n is a Haar system on this interval.

Approximation by trigonometric polynomials is quite similar to algebraic polynomial approximation except that now we approximate functions f which are 2π periodic. We let $C(T)$ denote the space of all such continuous functions and let $||\cdot||$ be the supremum norm on $(-\infty, \infty)$ or equivalently any interval of length 2π . If $f \in C(T)$, then f has a unique best approximation $T^* \in \tau_n$:

$$||f - T^*|| = \inf_{T \in \tau_n} ||f - T||.$$

The error of approximation in this case is denoted by $E_n^*(f) = ||f - T^*||$.

There is a very useful and important connection between trigonometric and algebraic polynomial approximation which is obtained by using the transformation $x = \cos \theta$ to identify points on $[-1, 1]$ with points on $[0, \pi]$. If $f \in C(I)$, $I = [-1, 1]$, then the function $g(\theta) := f(\cos \theta)$ is an even 2π periodic continuous function in $C(T)$. Similarly, if P is an algebraic polynomial of degree n then $T(\theta) := P(\cos \theta)$ is an even trigonometric polynomial of degree n .

We can go the other way as well. Namely, for any even trigonometric polynomial T , the function $P(x) := T(\arccos x)$ is an algebraic polynomial

of degree at most n . In fact, $T(\theta) = \sum_0^n a_k \cos k\theta$ and so $P(x)$ is a linear combination of the functions $C_k(x) := \cos k(\arccos x)$, $k=0, 1, \dots, n$. The C_k are algebraic polynomials of degree k (see §5).

It follows from the uniqueness of best approximations that the best approximation T^* to the even function g is an even trigonometric polynomial. Hence, the above one-to-one correspondence between algebraic polynomials P and even trigonometric polynomials T gives that P^* is the best approximation to f if and only if T^* is the best trigonometric approximation to g . We also have

$$(3.1) \quad E_n(f) = E_n^*(g).$$

This simple remark allows us to prove results about algebraic approximation by considering their analogue in trigonometric approximation.

4. Computing best approximants. It is generally difficult to compute best approximations. An exception is when X has an inner product (\cdot, \cdot) and its induced norm: $||f||^2 = (f, f)$. For example, $L_2(I)$ has the inner product

$$(4.1) \quad (f, g) := \int_I f(x)g(x)dx.$$

Now suppose that X_n is an n -dimensional subspace of X and we wish to compute the best approximation to $f \in X$ from X_n . We take a basis ϕ_1, \dots, ϕ_n for X_n which satisfies the orthonormality conditions

$$(4.2) \quad (\phi_i, \phi_j) = \delta_{ij}, \quad i, j = 1, \dots, n,$$

with δ_{ij} the usual Kronecker δ notation. The best approximation ϕ^* from X_n to f is then given by

$$(4.3) \quad \phi^* = \sum_1^n (f, \phi_k) \phi_k.$$

In fact, since $f - \phi^*$ is orthogonal to each ϕ_k , $k=1, \dots, n$, it is orthogonal to every $\phi \in X_n$. Therefore, we have

$$\|f - \phi^* - \phi\|^2 = (f - \phi^* - \phi, f - \phi^* - \phi) = (f - \phi^*, f - \phi^*) + (\phi, \phi) \geq \|f - \phi^*\|^2.$$

for all $\phi \in X_n$, which clearly says that ϕ^* is the best approximation to f from X_n .

For example, when $X = L_2(\mathbb{T})$, the space of 2π -periodic square integrable functions, then the best approximation to $f \in L_2(\mathbb{T})$ by trigonometric polynomials of degree at most n is $S_n(f)$, the n -th partial sum of the Fourier series of f :

$$(4.4) \quad S_n(f, x) := a_0/2 + \sum_0^n (a_k \cos kx + b_k \sin kx),$$

$$a_k := \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \cos kx \, dx; \quad b_k := \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \sin kx \, dx.$$

In this case, the error of approximation is simply $\left(\sum_{n+1}^{\infty} (a_k^2 + b_k^2) \right)^{1/2}$

For approximation in the space $C(I)$, there are only a few special cases where best approximants can be computed exactly. The simplest of these is for approximation by constants. For any $f \in C(I)$, its best approximation from Π_0 is $a := \frac{1}{2}(m+M)$ with m the minimum of f on I and M the maximum of f on I . Indeed since f takes on both its maximum and minimum on I , $f - a$ has two alternations and the Chebyshev criterion of Theorem 2.3 shows that a is the best approximation.

A similar result holds for the approximation of a convex (or concave) function f by linear polynomials. If Q is the linear polynomial which interpolates f at the end points of I , and $M := ||f-Q||$, then $P^* := Q-M/2$ is the best approximation to f from Π_1 .

Another very important example is the approximation of x^n by polynomials of degree $< n$. This problem was solved by Chebyshev and gave rise to a very important sequence of polynomials which bear his name.

5. Chebyshev Polynomials. We take $I := [-1, 1]$. To find the best approximation to x^n from Π_{n-1} , we need only find a polynomial $Q(x) = x^n + a_{n-1}x^{n-1} + \dots + a_0$ such that Q alternately takes on the values $\pm ||Q||$ at least $n+2$ times on I . From Theorem 2.3, $x^n - Q$ is the best approximation to x^n from Π_{n-1} . Now, the trigonometric polynomial $\cos n\theta$ has such alternation properties. Recalling our discussion in §3 of the transformation $x = \cos \theta$, we see that $C_n(x) := \cos(n \arccos x)$ is an algebraic polynomial of degree n which has norm one and has the required $n+1$ alternations; namely, $C_n(x_k) = (-1)^{n-k}$, for $x_k := \cos(n-k)\pi/n$, $k=0, \dots, n$.

Now the polynomial C_n is not quite the Q we are looking for since it does not have leading coefficient one. But it is easy to compute the leading coefficient of C_n . For this we use the recurrence relation

$$(5.1) \quad C_n(x) = 2xC_{n-1}(x) - C_{n-2}(x)$$

which follows from the corresponding trigonometric identity.

Since $C_0(x) \equiv 1$ and $C_1(x) \equiv x$, it follows by induction from (5.1) that

$$(5.2) \quad C_n(x) = 2^{n-1} x^n + \text{lower order terms.}$$

Hence $Q(x) := 2^{-n+1} C_n(x)$ is our sought after polynomial and $P^*(x) := x^n - Q(x)$ is the best approximation to x^n from Π_{n-1} . This also gives the error of approximation $E_n(x^n) = 2^{-n+1}$.

We do not have time to go into all the wonderful properties of Chebyshev polynomials but we should mention one of their other applications to estimating the size of $E_n(f)$. Let

$$(5.3) \quad x_k := \cos(2k-1)\pi/2n, \quad k=1, \dots, n,$$

be the zeros of C_n . If $f \in C[-1, 1]$, we let $P(x) := P(f, x)$ be the polynomial of degree $n-1$ which interpolates f at the points x_k , that is, $P(x_k) = f(x_k)$, $k=1, \dots, n$. The existence and uniqueness of P is well known and equivalent to the non-vanishing of the Vandermonde determinant. Also, one can represent the error (see [B, p. 9]) of interpolation by

$$(5.4) \quad E(x) := f(x) - P(f, x) = \frac{1}{(n+1)!} f^{(n+1)}(\xi_x) (x-x_1)\dots(x-x_n).$$

We recognize that $(x-x_1)\dots(x-x_n) = C_n(x)$, so that the right side of (5.3) does not exceed $\|f^{(n)}\| 2^{-n+1}/(n+1)!$. This gives

THEOREM 5.1 (Bernstein). If f has n continuous derivatives, then

$$E_n(f) \leq \|f^{(n+1)}\| 2^{-n+1}/(n+1)!.$$

6. Interpolation. Usually, we cannot determine the best polynomial approximants for a given $f \in C(I)$. Instead, we look for polynomials which are "good" rather than best approximations. The typical way of constructing such polynomials is to find linear operators L_n which map $C(I)$ onto Π_n and have good approximation properties. One possibility (others are considered in the next section) is for L_n to satisfy

$$(6.1) \quad \|f - L_n(f)\| \leq c_n E_n(f).$$

with c_n a constant which may depend on n . Then, except for the constant c_n , the polynomial $L_n(f)$ is just as good an approximation to f as is the best approximation. Of course, the smaller the constant c_n , the better the operator L_n and therefore we would like to find L_n which will make c_n as small as possible. It turns out, as we will explain in a little more detail shortly, that the best constants c_n behave like $\text{const.} \log n$; in particular, they tend to infinity with n . Thus unfortunately, the c_n in (6.1) cannot be replaced by a constant c which is independent of n .

Finding operators L_n which satisfy (6.1) is intimately connected with the construction of projectors onto the space Π_n . In fact L_n satisfies (6.1) if and only if it is such a projector, that is, if and only if $L_n(P) = P$ for all $P \in \Pi_n$. Indeed, if (6.1) holds and f is in Π_n , then $E_n(f) = 0$ and hence $L_n(f) = f$. On the other hand if L_n is such a projector then for any $f \in C(I)$ and $P \in \Pi_n$, we have

$$(6.2) \quad \|f - L_n(f)\| = \|L_n(f-P) - (f-P)\| \leq (\|L_n\| + 1) \|f-P\|,$$

where $\|L_n\| := \sup_{f \in C(I)} \|L_n(f)\|/\|f\|$ is the norm of L_n on $C(I)$.

Taking an infimum over all P in (6.2) shows that (6.1) holds with $c_n = \|L_n\| + 1$.

The smallest constant c_n which can be used in (6.1) is roughly speaking $\|L_n\|$. We have seen that we can always take $c_n \leq \|L_n\| + 1$. On the other hand for some appropriate f , and with I the identity operator, we have

$$\|f - L_n(f)\| = \|I - L_n\| \|f\| \geq (\|L_n\| - 1) \|f\| > (\|L_n\| - 1) E_n(f).$$

Hence, whenever (6.1) holds, we must have $c_n \geq \|L_n\| - 1$.

This means that to make c_n small, we had better make $\|L_n\|$ small. We are therefore led to the problem of constructing L_n with the smallest possible norm. This turns out to be a very difficult problem which is only solved in the special cases $n=0,1$. Nevertheless, it is possible to construct operators L_n which have close to the smallest possible norm. One of the simplest and most important methods of doing this is to use polynomial interpolation.

If $X: x_0, \dots, x_n$ are $n+1$ points from the interval I and $f \in C(I)$, then there is a unique polynomial $P_n(f) := P_n(f, X)$ which interpolates f at the points in X . In fact, we have the Lagrange representation for P_n :

$$(6.3) \quad P_n(f, x) = \sum_{k=0}^n f(x_k) l_k(x); \quad l_k(x) := \frac{\prod_{j \neq k} (x - x_j)}{\prod_{j \neq k} (x_k - x_j)}.$$

Then P_n is a linear operator which is a projector from $C(I)$ onto Π_n . It is simple to compute the norm of P_n :

$$(6.4) \quad \|P_n\| = \max_{x \in I} |\Lambda(x)|.$$

where

$$(6.5) \quad \Lambda(x) := \Lambda(X, x) := \sum_{k=0}^n |l_k(x)|$$

is called the Lebesgue function of P_n .

There is no simple description of interpolation points X which will minimize $\|P_n\|$; however, the work of Kilgore [K] and de Boor-Pinkus [B-P] give their uniqueness and some of their properties. The most obvious choice of interpolation points is to space the x_i equally in the interval I . But disappointingly, the norms of the resulting projector are then very large, in fact they grow exponentially with n . A much better choice for interpolation points X is the zeros of the Chebyshev polynomial C_n given in (5.3). In fact, with this choice, $\|P_n\| \leq (2/\pi) \log n + 1$, $n=1, \dots$. Hence, this projector has within constants the smallest possible norm. With this, we have

THEOREM 6.1. If P_n is the projector corresponding to interpolation at the zeros of the Chebyshev polynomial C_n , we have $\|P_n\| \leq (2/\pi) \log n + 1$. For any $f \in C(I)$,

$$(6.6) \quad \|f - P_n(f)\| \leq [(2/\pi) \log n + 1] E_n(f) \quad n=1,2,\dots$$

For a proof of this theorem, we refer the reader to the book of Rivlin [R, p.18] on Chebyshev polynomials.

For n small, $\log n$ is not too large so that the approximation P comparable with the best approximation. On the other hand, there are functions f for which the right hand side does not tend to 0 and even more to the point for which $P_n(f)$ does not converge to f . So in spite of the attractiveness of polynomial interpolation, this type of approximation can not even give a proof of the Weierstrass theorem.

7. Degree of approximation. We have yet to discuss the behavior of $E_n(f)$. We expect that the nicer the function f then the faster $E_n(f)$ converges to zero. One result in this direction is the following:

THEOREM 7.1. If f is r times continuously differentiable on $I=[-1,1]$, then

$$(7.1) \quad E_n(f) \leq C_r \|f^{(r)}\| n^{-r}, \quad n=1,2,\dots$$

Thus for example, we know that $E_n(f)$ tends to zero at least as fast as $1/n$ whenever f is differentiable, $1/n^2$ when it is twice differentiable, and so on.

Estimates of the type given in Theorem 7.1 have a rich history. The first results of this type were given at the beginning of this century by Bernstein [Be]. Later, Favard [F] found the best constant C_r . Jackson [J] then refined (7.1) by using subtler measures of the smoothness of a function f such as its modulus of continuity $\omega(f,t)$ defined for $f \in C(I)$ by

$$\omega(f,t) := \sup_{\substack{|x-y| \leq t \\ x,y \in I}} |f(x) - f(y)|.$$

THEOREM 7.2 (Jackson) Let $r=1,2,\dots$. If f is r times continuously differentiable, then

$$(7.2) \quad E_n(f) \leq C_r n^{-r} \omega(f^{(r)}, n^{-1}), \quad n=1,2,\dots$$

The continuity of f insures that $\omega(f,t) \rightarrow 0$ as $t \rightarrow 0$ and therefore (7.2) with $r=0$ shows that $E_n(f) \rightarrow 0$, $n \rightarrow \infty$. Hence (7.2) contains the Weierstrass theorem as well.

There are now several different techniques for proving Jackson's theorem. One of the most important is to use the transformation $x = \cos \theta$ as described in §3. If $f \in C(I)$, then $g(\theta) := f(\cos \theta)$ satisfies $\omega(g,t) \leq \omega(f,t)$ and if f is r times continuously differentiable so is g . Using these ideas, Theorems 7.1 and 7.2 follow from their counterparts for trigonometric approximation.

To approximate a function $g \in C^*(T)$, we can use convolution operators. Namely, if K_n is a trigonometric polynomial of degree n , then

$$(7.3) \quad L_n(g, \theta) := g * K_n(\theta) := \frac{1}{2\pi} \int_{-\pi}^{\pi} g(\theta-t) K_n(t) dt$$

is likewise a trigonometric polynomial of degree n . In order that L_n preserve constant functions, we shall require that

$$(7.4) \quad \int_{-\pi}^{\pi} K_n(t) dt = 2\pi$$

It is also convenient to take K_n non-negative and even.

If we want $L_n(f)$ to provide a good approximation to f , the kernel K_n should concentrate its mass near the origin (similar to the delta function). In fact, in order to prove Theorem 7.1 for $r=0$ or (7.1) for $r=1$ by using L_n , it is enough to have

$$(7.5) \quad \int_{-\pi}^{\pi} \sin^2 t/2 K_n(t) dt \leq \text{const. } n^{-2}$$

Let us indicate how (7.5) gives a proof of these results. From (7.5) and the Cauchy-Schwarz inequality for positive functionals, we find

$$(7.6) \quad \int_{-\pi}^{\pi} |t| K_n(t) dt \leq \pi \int_{-\pi}^{\pi} |\sin t/2| K_n(t) dt \leq \pi \left[\int_{-\pi}^{\pi} \sin^2 t/2 K_n(t) dt \right]^{1/2} \leq \frac{C}{n}$$

Now if f' is continuous and $M := \|f'\|$, then $|f(\theta-t) - f(\theta)| \leq M |t|$. Since $L_n(f, \theta) = f(\theta)$ (because L_n preserves constants), we have

$$(7.7) \quad |L_n(f, \theta) - f(\theta)| \leq \frac{1}{2\pi} \int_{-\pi}^{\pi} |f(\theta-t) - f(\theta)| K_n(t) dt \leq \int_{-\pi}^{\pi} M |t| K_n(t) dt \leq CM/n,$$

which is (7.1) when $r=1$.

In this same way, we can also prove Theorem 7.2. This requires the inequality $\omega(f, t) < (nt+1)\omega(f, 1/n)$, $t > 0$, which follows from the subadditivity of ω : ($\omega(f, t_1+t_2) \leq \omega(f, t_1) + \omega(f, t_2)$). Using this and the inequality $|f(\theta-t) - f(\theta)| \leq \omega(f, |t|)$ as in (7.7) gives Theorem 7.2 because of (7.4) and (7.6).

There are many choices of kernel K_n which satisfy (7.4). Indeed, since $\sin^2 t/2 = \frac{1}{2}(1 - \cos t)$, (7.4) can be restated as a condition on the first

Fourier coefficients $\hat{K}_n(1) = \int_{-\pi}^{\pi} e^{-it} K_n(t) dt$ of K_n , namely:

$$(7.8) \quad 1 - K_n(1) \leq C n^{-2}.$$

Thus any positive trigonometric kernel K_n which have integral one and satisfies (7.8) has the desired properties.

One of the simplest example of such a kernel was given by Jackson:

$$K_n(t) = \lambda_n \left(\frac{\sin mt/2}{\sin t/2} \right)^4, \quad m := [n/4]+1, \quad n=1,2,\dots,$$

with λ_n chosen so that K_n satisfies (7.4). One easily shows that $\lambda_n \sim n$ and then deduces (7.4) (see [L, p.55]).

8. Piecewise polynomial approximation. Polynomials are not usually the best choice for applied problems or numerical computation. For one thing it is not easy to evaluate a polynomial of high degree. Piecewise polynomial functions of low degree are much more desirable in computation. Indeed, it is the case that most numerical algorithms are based in one sense or another on some form of piecewise polynomial approximation.

We consider once again the interval $I:=[-1,1]$ and let $T: -1=t_0 < t_1 < \dots < t_n=1$ be an increasing sequence of points from I . Then T partitions I into n intervals $I_j := [t_{j-1}, t_j)$, $j=1, \dots, n-1$, $I_n := [t_{n-1}, t_n]$. By $S_r(T)$, we denote the piecewise polynomial functions of degree $r-1$ on T . That is, $S \in S_r(T)$ means that $S(x) = P_j(x)$, $x \in I_j$, with P_j a polynomial of degree $< r$ for $j=1, \dots, n$. Sometimes we are only interested in functions from $S_r(T)$ which have some prescribed continuity at the points t_j . These are called spline functions. For example the continuously differentiable functions in $S_4(T)$ are called C^1 cubic splines; those that are twice differentiable are called C^2 cubic splines, etc. The simplest spline functions are the truncated power functions $(x-c)_+^k$, $k=0, \dots$ with

$$x_+^k := \begin{cases} 0, & x \leq 0 \\ x^k, & x > 0 \end{cases}$$

To describe the error in approximation by piecewise polynomials it is enough to consider how the error in polynomial approximation on a subinterval $J := [a,b]$ of I depends on the length of J . For this, we let

$$(8.1) \quad E_r(f, J) := \inf_{P \in \Pi_{r-1}} \|f - P\|(J)$$

with the norm being the sup norm (on the interval J), as usual.

Now if f has r continuous derivatives we can form the Taylor polynomial T_a of f at a :

$$T_a(x) := f(a) + f'(a)(x-a) + \dots + f^{(r-1)}(a)(x-a)^{r-1}/(r-1)!.$$

We have the well known error formula for

$$(8.2) \quad f(x) - T_a(x) = \frac{1}{(r-1)!} \int_J f^{(r)}(t) (x-t)_+^{r-1} dt.$$

If in the integral on the right side of (8.2), we replace $f^{(r)}$ by $\|f^{(r)}\|$ and then integrate, we see that

$$(8.3) \quad E_r(f, J) \leq \frac{1}{r!} \|f^{(r)}\| |J|^r,$$

with $|J|$ the length of J . Thus, as the length of J tends to zero the error of approximation $E_r(f, J)$ goes to zero like $|J|^r$.

The simple inequality (8.3) already tells us a lot about approximation by the elements of $S_r(T)$. Namely, if

$$\delta_T := \max_{1 \leq j \leq n} |I_j|,$$

then we have

THEOREM 8.1. If f has r continuous derivatives on I , there is a piecewise polynomial $S \in S_r(T)$ such that

$$(8.4) \quad \|f - S\|(I) \leq \frac{1}{r!} \|f^{(r)}\| \delta_T^r.$$

Indeed, we can define S on the interval I_j to be the Taylor polynomial of f for the left end point of I_j so that (8.4) follows from (8.3).

Sometimes it is useful not to assume that $f^{(r)}$ is continuous but only that $f^{(r-1)}$ is absolutely continuous and $f^{(r)}$ is in L_p for some $1 \leq p < \infty$. In this case, if we apply Holder's inequality to the integral in (8.2), we find

$$(8.5) \quad E_r(f, J) \leq \frac{1}{(r-1)!} \|f^{(r)}\|_p(J) |J|^{r-1/p}.$$

Hence, there is a spline $S \in S_r(T)$ which satisfies

$$(8.6) \quad \|f - S\|(I) \leq \frac{1}{(r-1)!} \|f^{(r)}\|_p(I) \delta_T^{r-1/p}.$$

There are a variety of other estimates (see [S]) for the error in spline approximation. For example, $E_r(f, J)$ can be estimated by $\text{const. } \omega(f, \delta_T)$ or by $\text{const. } \|f^{(k)}\| \delta_T^k$, for any $0 < k \leq r$. Remarkably, it is also possible to prove these same estimates for approximation by spline functions which have smoothness. For example, we have

THEOREM 8.2. If f has r continuous derivatives on I , there is a spline function $S \in S_r(T)$ which has $r-2$ continuous derivatives and satisfies

$$(8.7) \quad \|f-S\|(I) \leq C \|f^{(r)}\|(I) \delta_T^r,$$

with C depending only on r .

There are several techniques for proving estimates like (8.7). When $r=1$, (8.7) follows from (8.5) since there is no continuity prescribed. When $r=2$, the case of approximation by piecewise linear functions, we can take S as the continuous piecewise linear function in $S_2(T)$ which interpolates f at each of the t_i , $i=0, \dots, n$. Interpolating splines can also be used for other small values of r but the question of where to place these points gets more and more sticky as r increases and is still not solved for general r .

A more successful method to prove (8.7) was introduced by de Boor and Fix [B-F]. It uses certain linear operators L_T called quasi-interpolants. L_T is a projection from $C(I)$ onto $S_r(T) \cap C^{(r-1)}$. While $L_T(f)$ does not interpolate f in the usual sense, it uses only a finite number of values of f (hence the name quasi-interpolant). Quite surprisingly the norms of the projectors L_T are bounded independent of T . This is in stark contrast to polynomials where as explained in §6, the norms of projectors onto Π_n must tend to infinity with increasing n . Using the fact that the L_T are bounded, one proves (8.7) (see [D₁]).

Unfortunately, we do not have time to describe these powerful approximation methods in more detail but certainly they will be brought up in other lectures. The reader should also consult the books of de Boor [B] and Schumaker [S].

9. Non-linear approximation. Up to this point, we have only discussed approximation by elements from a linear space (polynomials or trigonometric polynomials). But many other families of functions used in approximation are not linear spaces. For example, we have the set R_n of rational functions R of degree $\leq n$ or the set $S_n = S_{n,k}$ of all piecewise polynomials of degree k which have n pieces.

Approximation by such non-linear families can sometimes give dramatic improvement in the error of approximation. For example, approximation to the function $f(x)=|x|$ was extensively studied by S. Bernstein [Be] because it is prototypical for polynomial approximation to one time differentiable functions. Bernstein showed that for the error $E_n(f)$ of polynomial approximation, we have $\lim_{n \rightarrow \infty} n E_n(f)$ exists. Hence $E_n(f)$ behaves like

$\text{const.}/n$ as $n \rightarrow \infty$. It was therefore a great surprise when D.J. Newman [N]

showed that f can be approximated by rational functions of degree n with an error not exceeding $3e^{-\sqrt{n}}$, for n sufficiently large. The reason for this dramatic improvement is that the function $|x|$ is piecewise analytic with a singularity at $x=0$. Both polynomials and rational functions can approximate well where f is analytic but rational functions can go one better and also approximate well near the singularity. This is accomplished by stacking the poles of this rational function near this singularity.

This example illustrates rather well the general advantage of non-linear approximation over its linear counterparts. In the non-linear case, the approximation process has the capacity to concentrate on the places where the function f is least smooth. Here is a simple example of this.

We consider functions f defined on $I := [0,1]$ which are absolutely continuous, that is f has a first derivative which is integrable. For such f , we let $\text{Int}(f', J)$ denote the integral of $|f'|$ over the interval J . We assume that $\text{Int}(f', I) = 1$ and we want to approximate f in the uniform norm. If the approximation is done by algebraic polynomials, we can say nothing more than $E_n(f) \rightarrow 0$ as $n \rightarrow \infty$. That is, by suitably choosing f we can make $E_n(f)$ tend to zero as slow as we wish.

On the other hand, consider the approximation of such an f by step functions S which have at most n steps on I . The set of all such functions which we denote by $S_n := S_{n,0}$ is a non-linear space which clearly depends on n parameters. We let $0 =: x_0 < x_1 < \dots < x_n =: 1$ be a partition of I into intervals $I_j := [x_{j-1}, x_j]$ with

$$(9.1) \quad \text{Int}(f', I_j) = 1/n, \quad j=1, \dots, n.$$

Now, if we let

$$(9.2) \quad S(x) := f(x_j) \quad \text{on } I_j,$$

then clearly

$$(9.3) \quad |f(x) - S(x)| \leq \text{Int}(f', I_j) = 1/n, \quad x \in I_j, \quad j=1, \dots, n.$$

Hence, the functions in S_n approximate f much better than polynomials can. This type of approximation is called optimal knot or free knot spline approximation because the knots x_i of the piecewise polynomial are allowed to depend on the function f . Analogous results hold for approximating r times differentiable functions with $f^{(r)}$ in $L_1(I)$ by piecewise polynomials of degree $r-1$ which have n pieces.

One drawback to the above approximation method is that it is difficult to find the knots x_j such that (9.1) holds. For this reason, there is much

favor in so-called adaptive methods which find a partition of I into interval I_j by subdividing. For example, the adaptive analogue of (9.1-2) would proceed as follows. We would choose some tolerance $\epsilon > 0$ which is the error we are willing to accept in the approximation. We call an interval J "good" if $\text{Int}(f', J) < \epsilon$. Otherwise, J is "bad". We are looking to generate a set G_ϵ of disjoint good intervals which are a partition of I . If I itself is good, we can simply take $G_\epsilon = \{I\}$. On the other hand if I is bad, we divide I in half producing therefore two new intervals. Whichever of these two intervals is good, we put into our set G_ϵ . The bad intervals are further subdivided. We continue in this way and the whole process stops when there are no more bad intervals.

Of interest to us is how many good intervals will appear in the set G_ϵ . Birman and Solomjak [B-S] have shown that when f' is in L_p for some $p > 1$ then G_ϵ contains no more than C_0/ϵ intervals with C_0 depending only on p . Thus for example, if we take $\epsilon = C_0/n$, then G_ϵ will contain at most n intervals I_1, \dots, I_n . If we then take S as in (9.2), (9.3) will again be satisfied. This means that by assuming slightly more about the function f , we can approximate f by the above adaptive scheme to the same accuracy as with the optimal knot approximation. There are a variety of other results on adaptive approximation which show that functions with singularities can be approximated better in this way than by linear methods of approximation (see for example [B-R]).

The piecewise polynomials used in the above approximation are not smooth. While it is possible to modify these methods so that the resulting piecewise polynomial has smoothness $C^{(r-1)}$ (in the case the piecewise polynomials have degree r), it is sometimes of interest to approximate f by smoother functions. It turns out that the same accuracy of approximation is attainable with rational functions. For example, we have the following result of Popov [P]

THEOREM 9.1. If $f' \in L_p(I)$, $p > 1$, there is a rational function R of degree at most n which satisfies

$$(9.4) \quad \|f-R\| \leq c \|f'\|_p n^{-1}.$$

There is a simple technique [D₂] for deriving (9.4) from (9.3). For this, we can assume that $\|f'\|_p = 1$. Then, as in the derivation of (9.3), we choose n intervals I_j , such that

$$(9.5) \quad \int_{I_j} |f'|^p \leq 1/n$$

By refining these intervals if necessary, we can further require that $|I_j| \leq 1/n$ and still there are at most $2n$ of these intervals. We let ξ_j be a point in I_j and define

$$\psi_j(x) := \frac{|I_j|^2}{(x - \xi_j)^2 + |I_j|^2}.$$

If $\Psi := \sum \psi_j$, then the functions

$$\phi_j := \psi_j / \Psi, \quad j=1, \dots, n$$

are a partition of unity:

$$\sum \phi_j(x) = 1, \quad x \in I.$$

It can be shown [D₂] that for suitably chosen ξ_j , the rational function

$$R := \sum f(\xi_j) \phi_j$$

has degree at most $4n$ and satisfies (9.4).

BIBLIOGRAPHY

[Be₁] S. Bernstein, Sur l'ordre de la meilleure approximation des fonctions continues par des polynomes de degré donné, Memoires publiés par la classe des sci. Acad. de Belgique (2)4(1912), 1-103.

[Be₂] S. Bernstein, Sur la meilleure approximation de $|x|$ par de polynomes de degrés donnés, Acta Math., 37(1914), 1-57.

[B-S] M.S. Birman- M. Solomjak, Piecewise polynomial approximation of functions of the class W_p^α , Math. USSR, Vol.2 No. 3(1967), 295-317.

[B] C. de Boor, A Practical Guide to Splines, Applied Math. Science, Springer-Verlag vol. 27, New York, 1978

[B-F] C. de Boor- G. Fix, Spline approximation by quasi-interpolants, J. Approx. Th. 8(1973), 19-45.

[B-P] C. de Boor- A. Pinkus, Proof of the conjecture of Bernstein and Erdos concerning the optimal nodes for polynomial interpolation, J. Approx. Theory, 24(1978), 289-303.

- [B-R] C. de Boor - J. Rice, An adaptive algorithm for multivariate approximation giving optimal convergence rates, *JAT* 25(1979), 337-359.
- [D₁] R. DeVore, Degree of approximation, in: Approximation II, Academic Press, New York, 1976, pp. 117-161.
- [D₂] R. DeVore, Maximal functions and their application to rational approximation, in Approximation Theory, CMS Conference Proceedings Vol. 3, 1983, Amer. Math. Soc., p. 143-155.
- [F] J. Favard, Sur les meilleures procédés d'approximation de certaines classes des fonctions par des polynômes trigonométriques, *Bull. Sci. Math.* 61(1937) 209-224, 243-256.
- [J] D. Jackson, On the approximation by trigonometric sums and polynomials, *TAMS*, 13(1912), 491-515.
- [K-S] S. Karlin - W. Studden, Tchebycheff Systems: with Applications in Analysis and Statistics, Interscience, Wiley, New York, 1966.
- [Kr] M.G. Krein, The ideas of P.L. Chebyshev and A.A. Markov in the theory of limiting values of integrals and their further developments, *Amer. Math. Soc. Translations*, Ser. 2, 12, 1-122.
- [Ki] T.A. Kilgore, A characterization of Lagrange interpolating projections with minimal Tchebycheff norm, *J. Approx. Theory* 24(1978), 273-288.
- [L] G.G. Lorentz, Approximation of Functions, Holt, Rinehart and Winston, New York, 1966.
- [N] D.J. Newman, Rational approximation of $|x|$. *Michigan Math. J.* 11(1964), 11-14.
- [P] V. Popov, Uniform rational approximation of the class V_r and its applications, *Acta Math. Acad. Sci. Hung.* 29(1977), 119-129.
- [R] T. J. Rivlin, The Chebyshev Polynomials, Interscience, Wiley, New York, 1974.
- [S] L. Schumaker, Spline Functions: Basic Theory, Interscience, Wiley, New York, 1981.
- [W] K. Weierstrass, Über die analytische Darstellbarkeit sogenannter willkürlicher Functionen reeller Argumente, *Sitzungsberichte der Acad. Berlin* (1885), 633-639, 789-805.