A BLAISDELL BOOK IN NUMERICAL ANALYSIS AND COMPUTER SCIENCE

CONSULTING EDITOR

Philip J. Davis, Brown University

AN INTRODUCTION TO THE APPROXIMATION OF FUNCTIONS

Theodore J. Rivlin

THOMAS J. WATSON RESEARCH CENTER
IBM CORPORATION

BLAISDELL PUBLISHING COMPANY
A Division of Ginn and Company
WALTHAM, MASSACHUSETTS · TORONTO · LONDON

For
Madeline
and
Jean

LIBRARY OF CONGRESS CATALOG CARD NUMBER : 69-10378
PRINTED IN THE UNITED STATES OF AMERICA.

PREFACE

It is by now a commonplace observation that the needs of automatic digital computation have spurred an enormous revival of interest in methods of approximating continuous functions by functions which depend only on a finite number of parameters. The purpose of this book is to provide an introduction to some of the most significant of these methods with particular emphasis on approximation by polynomials, although approximation by piecewise polynomial functions and rational functions is also discussed.

The author views approximation theory as an area of mathematics with important practical applications in computation, and intends to provide here an introduction to the theoretical foundations which underlie many of the algorithms of everyday use. For this reason, for each method of approximation studied at least one algorithm leading to actual numerical approximations is described and, indeed, traced from its theoretical origins to its present formulation. There are, however, no flow charts or actual programs in this book, and the algorithms that are described in detail are intended more to illustrate one possibility than to suggest that they are the "best" available.

Apart from its applications approximation theory is a lively branch of mathematical analysis. The material in this book can be used as additional reading in introductory courses in mathematical analysis as well as numerical analysis. A reader who has studied advanced calculus and the rudiments of linear algebra is amply prepared to understand this book. An effort has been made to avoid more sophisticated prerequisites even at the cost of making the presentation clumsier, as, for example, is the case in Chapter 3 where measure theory, which is the natural language of the material, is not mentioned in order to keep the topic within the grasp of the uninitiated. In keeping with this philosophy and in the interests of good pedagogy the author has not hesitated, at times, to repeat similar arguments, to prove a weaker and then a stronger form of the same theorem, to prove a special case when essentially the same proof gives the general result, to stick to garden variety polynomials, when arbitrary Chebyshev Systems cost very little more and to provide a proof of "something everybody already knows."

Some of the exercises at the end of each chapter verge on the trivial, others are details needed in the text whose inclusion there would needlessly delay the exposition, while still other connected sets of exercises are intended to entice the reader into some interesting side excursions.

References to the literature are given in the usual fashion. A superscripted number in the form $^{(n)}$ refers to item (n) in the notes at the end of the chapter in question.

The author has profited from discussions with many of his colleagues in the Mathematical Sciences Department of the IBM Research Center and welcomes this opportunity to express his thanks to them, as well as to Mrs. Joyce Abish who typed the manuscript in her usual impeccable fashion.

T. J. R.

Bronx, New York

CONTENTS

AN INTRODUCTION TO THE APPROXIMATION OF FUNCTIONS

INTRODUCTION

There are certain generalities about approximation theory that will be useful in our later, more detailed study of specific approximation techniques. The natural setting for these general results is a *normed linear space*. Linear spaces have become familiar objects in mathematics, and so we assume that the reader is familiar with their definition and most elementary properties. We shall take the scalars to be the real numbers unless some other field is specified.

Let V be a linear space. We recall that a *norm* is a function from V into the nonnegative real numbers. This function is written $\|\cdot\|$ and satisfies the following three properties:

$$\begin{array}{lll} \text{(i)} & \|v\| \geq 0 \text{ with equality if and only if } v = 0. & \\ \text{(ii)} & \|\lambda v\| = |\lambda| \, \|v\| \qquad \text{for any scalar } \lambda. & \qquad \text{(I.1)} \\ \text{(iii)} & \|v + w\| \leq \|v\| + \|w\| \qquad \text{(the Triangle Inequality).} & \end{array}$$

The norm gives us a notion of *distance* in V. If $w, v \in V$, then the distance from w to v (or v to w) is $\|v - w\|$.

We are now in a position to present the general setting for much of approximation theory. Let W be a subset of V, then, given $v \in V$, the approximation problem, baldly stated, is: Find a $w \in W$ whose distance from v is least; that is, find $w^* \in W$ such that $\|v - w\|$ is least for $w = w^*$. Such a w^* we call a best approximation to v out of W. Problems arise immediately. Is there such a w^*? If there is, is there only one? Since, as we shall see, many of the most widely studied and used methods of approximation are instances of this general approximation problem, we shall save much duplication of effort by obtaining some results in the general situation.

We turn first to the existence question. We have

THEOREM I.1. *If V is a normed linear space and W a finite-dimensional subspace of V, then, given $v \in V$, there exists $w^* \in W$ such that*

$$\|v - w^*\| \leq \|v - w\|$$

for all $w \in W$.

Proof. Since $0 \in W$, it is a competitor for best approximation to v out of W. Its distance from v is $\|v - 0\| = \|v\|$. If $\|v - w\| > \|v\|$, we are, therefore, sure that w cannot possibly be a best approximation to v, and hence we restrict our attention to $w \in W$ which satisfy

$$\|v - w\| \leq \|v\| = M.$$

If $\|v - w\| \leq M$, then

$$\|w\| = \|-w\| = \|(v - w) + (-v)\| \leq \|v - w\| + \|v\| \leq 2M.$$

Suppose W is k-dimensional and $w_1, \ldots, w_k$ is a basis for W; then we are trying to prove that the function

$$f(\lambda_1, \ldots, \lambda_k) = \|v - (\lambda_1 w_1 + \lambda_2 w_2 + \cdots + \lambda_k w_k)\| \tag{I.2}$$

takes on a minimum value as the point $(\lambda_1, \ldots, \lambda_k)$ varies in k-space in such a way that

$$\|\lambda_1 w_1 + \cdots + \lambda_k w_k\| \leq 2M. \tag{I.3}$$

We want to see what restriction (I.3) implies for the position of the point $(\lambda_1, \ldots, \lambda_k)$. The function

$$g(\lambda_1, \ldots, \lambda_k) = \|\lambda_1 w_1 + \cdots + \lambda_k w_k\| \tag{I.4}$$

is a continuous function of $(\lambda_1, \ldots, \lambda_k)$ (see Exercise I.1) and so assumes its minimum value on the compact set

$$|\lambda_1| + |\lambda_2| + \cdots + |\lambda_k| = 1. \tag{I.5}$$

This minimum value, m, is *positive*. To see this, note that if

$$g(\lambda_1^*, \lambda_2^*, \ldots, \lambda_k^*) = m,$$

then $m \geq 0$. If $m = 0$, then $\|\lambda_1^* w_1 + \cdots + \lambda_k^* w_k\| = 0$, which by (I.1) (i) implies that $\lambda_1^* w_1 + \cdots + \lambda_k^* w_k = 0$. Since $w_1, \ldots, w_k$ are linear independent, we conclude that $\lambda_1^* = \lambda_2^* = \cdots = \lambda_k^* = 0$, thus violating (I.5).

Now if we relax (I.5) and require only that

$$\sum_{i=1}^{k} |\lambda_i| \neq 0,$$

then

$$g\left(\frac{\lambda_1}{\sum_{i=1}^k |\lambda_i|}, \frac{\lambda_2}{\sum_{i=1}^k |\lambda_i|}, \ldots, \frac{\lambda_k}{\sum_{i=1}^k |\lambda_i|}\right) \geq m > 0,$$

and so, in view of (I.1) (ii) we obtain

$$g(\lambda_1, \ldots, \lambda_k) \geq m \sum_{i=1}^{k} |\lambda_i|. \tag{I.6}$$

But (I.6) remains true if $\sum_{i=1}^{k} |\lambda_i| = 0$ and, therefore, is true for all $(\lambda_1, \ldots, \lambda_k)$. Hence, we conclude that (I.3) implies the restriction

$$|\lambda_1| + \cdots + |\lambda_k| \leq \frac{2M}{m}, \tag{I.7}$$

which in turns implies that

$$|\lambda_i| \leq 2\frac{M}{m}, \qquad i = 1, \ldots, k. \tag{I.8}$$

(I.8) defines a hypercube in k-space, which is a compact set, and $f(\lambda_1, \ldots, \lambda_k)$ is a continuous function of $(\lambda_1, \ldots, \lambda_k)$; therefore f assumes its minimum value on (I.8). (As a matter of fact, the point $(\lambda_1, \ldots, \lambda_k)$ at which the minimum value of f is assumed which *might* satisfy (I.3) *must* satisfy (I.3) since all other points have been ruled out as competitors.) The theorem is proved. ∎

Remark. The field of scalars of V could just as well be the complex numbers. The proof is the same.

Let us look at some examples before continuing with the general theory.

Example I.1. The set of functions continuous on a given closed interval $[a, b]$, which we denote by $C[a, b]$, is a linear space. If $f \in C[a, b]$, we can define a norm in $C[a, b]$ by

$$\|f\| = \max_{a \leq x \leq b} |f(x)|. \tag{I.9}$$

The norm is called the *uniform* or *Chebyshev* norm. It is easy to check (and the reader should do so) that (I.9) does, indeed, define a norm. As an example of Theorem I.1 take V to be $C[a, b]$ and let W be the $(n + 1)$-dimensional subspace of $C[a, b]$ spanned by the functions $1, x, \ldots, x^n$. That is, W consists of all polynomials of degree at most n. We call this particular subspace, which plays an important role in our book, P_n. Theorem I.1 now informs us that every continuous function, $f(x)$, on $[a, b]$ has a best approximation out of the polynomials of degree at most n in the uniform norm. That is, given $f \in C[a, b]$, there exists $p^* \in P_n$ such that

$$\max_{a \leq x \leq b} |f(x) - p^*(x)| \leq \max_{a \leq x \leq b} |f(x) - p(x)| \tag{I.10}$$

for all $p \in P_n$.

Notice that the uniform norm singles out x values at which the approximation is worst (that is, where the absolute error is greatest) and assigns as a measure of approximation these worst possibilities. It thus provides absolutely certain bounds on the error at the expense of these bounds having to be large enough to be valid at every point, no matter how exceptional. Another way of expressing (I.10) is

$$\min_{p \in P_n} \max_{a \leq x \leq b} |f(x) - p(x)| = \max_{a \leq x \leq b} |f(x) - p^*(x)| ;$$

hence the name "min-max" is sometimes used for approximation using the uniform norm.

Example I.2. Another instance in which the uniform norm is widely used is the case of functions defined on a finite point set. Given m distinct real points $x_1 < x_2 < \cdots < x_m$, the set of functions defined on $x_1, \ldots, x_m$ is precisely E_m, the m-dimensional linear space of numerical vectors f: $(f_1, \ldots, f_m)$; f_j may be thought of as the value of $f(x)$ at $x = x_j, j = 1, \ldots, m$. The uniform norm on E_m is defined by

$$\|f\| = \max_{i=1,\ldots,m} |f_i| .$$

(Note that while this defines a norm on E_m, it does not define a norm on, say, $C[x_1, x_m]$ since $\|f\|$ may equal 0 without f being the zero function.) As an application of Theorem I.1 let us take $V = E_m$ and let W be the $(n + 1)$-dimensional subspace of E_m consisting of all vectors p: $(p(x_1), \ldots, p(x_m))$, where $p \in P_n$ and $n < m - 1$. Theorem I.1 tells us that there exists $p^* \in P_n$ such that

$$\max_{i=1,\ldots,m} |f_i - p^*(x_i)| \leq \max_{i=1,\ldots,m} |f_i - p(x_i)|$$

for all $p \in P_n$.

Example I.3. Instead of the uniform norm in $C[a, b]$, we often consider the norm defined by

$$\|f\| = \left[\int_a^b |f(x)|^p \, dx\right]^{1/p}, \tag{I.11}$$

where p is a real number, $p \geq 1$. Here, also, there is a finite point set analogue. Given m real distinct points $x_1 < x_2 < \cdots < x_m$, we can introduce as norm in E_m

$$\|f\| = \left[\sum_{i=1}^{m} |f_i|^p\right]^{1/p}, \quad p \geq 1. \tag{I.12}$$

In the case $p = 2$, we recover the usual Euclidean distance in E_m. In both cases of the "p-norm," we conclude from Theorem I.1 that there exists a best approximation to a given f out of P_n.

Example I.4. The requirement that W be *finite*-dimensional in Theorem I.1 is essential. For suppose W is the subspace of $V = C[0, \frac{1}{2}]$ consisting of all polynomials (of any degree). Clearly, W is not finite-dimensional (if it were, what could its dimension be?). We wish to show that $f(x) = 1/(1 - x)$ has no best approximation in the uniform sense on $[0, \frac{1}{2}]$ out of W. Note that, given $\varepsilon > 0$, there exists N such that

$$|f(x) - (1 + x + x^2 + \dots + x^N)| < \varepsilon, \qquad 0 \leq x \leq \tfrac{1}{2}.$$

Hence, if there were a best uniform approximation to $f(x)$ out of W, say p^*, it would have to satisfy

$$\|f - p^*\| = 0,$$

which implies that

$$\frac{1}{1 - x} \equiv p^*,$$

an impossibility.

Suppose now that W is a subspace of V and let W^* be the set of best approximations to a given $v \in V$ out of W. (Theorem I.1 gives us a condition under which W^* is not empty.) We wish to prove that W^* is a *convex* set. We recall that a set, S, in a linear space is convex if $s_1, s_2 \in S$ implies that

$$\lambda_1 s_1 + \lambda_2 s_2 \in S$$

if λ_1 and λ_2 are nonnegative and

$$\lambda_1 + \lambda_2 = 1.$$

If S is empty or consists of one point, then it is clearly convex.

THEOREM I.2. *If $v \in V$ and W is a subspace of V, the set of best approximations to v out of W, call it W^*, is convex.*

Proof. If W^* is empty, the theorem is true. Suppose that $w_1^*, w_2^* \in W^*$; then

$$\|v - w_1^*\| = \|v - w_2^*\| = \rho.$$

Suppose $\lambda_1, \lambda_2 \geq 0$ and $\lambda_1 + \lambda_2 = 1$; then

$$\begin{aligned}\|v - (\lambda_1 w_1^* + \lambda_2 w_2^*)\| &= \|\lambda_1(v - w_1^*) + \lambda_2(v - w_2^*)\| \\ &\leq \lambda_1\|v - w_1^*\| + \lambda_2\|v - w_2^*\| = (\lambda_1 + \lambda_2)\rho = \rho.\end{aligned}$$

Thus, $\lambda_1 w_1^* + \lambda_2 w_2^* \in W^*$, and so W^* is convex. ▮

Theorem I.2 has the consequence that, if there are two distinct best approximations out of W to v, there are infinitely many (in fact, uncountably many) best approximations.

A final general result gives a criterion that insures that, if there is a best approximation, there is only one. The normed linear space V is said to have a *strictly convex norm* if the set B: $\{v/\|v\| \leq 1\}$, called the *unit ball* in V, is *strictly convex* (sometimes called rotund). B is certainly convex. For it to be *strictly* convex, we require that, if $v_1 \neq v_2$, $\|v_1\| = 1$ and $\|v_2\| = 1$, then $\|\lambda_1 v_1 + \lambda_2 v_2\| < 1$ if $\lambda_1, \lambda_2 > 0$ and $\lambda_1 + \lambda_2 = 1$. (That is, the boundary of B contains no open line segment.)

We can now state a uniqueness theorem.

THEOREM I.3. *If V has a strictly convex norm, then a given $v \in V$ has at most one best approximation out of a subspace, W, of V.*

Proof. Suppose w_1^* and w_2^* are two distinct best approximations to v out of W. By Theorem I.2, $(w_1^* + w_2^*)/2$ is also a best approximation to V out of W. Suppose

$$\|v - w_1^*\| = \|v - w_2^*\| = \rho.$$

Put

$$v_1 = (v - w_1^*)/\rho, \qquad v_2 = (v - w_2^*)/\rho.$$

Then $v_1 \neq v_2$, $\|v_1\| = 1$, $\|v_2\| = 1$ and, since the norm in V is *strictly convex*,

$$\left\|\frac{1}{2} v_1 + \frac{1}{2} v_2\right\| = \left\|\frac{1}{2\rho}(v - w_1^*) + \frac{1}{2\rho}(v - w_2^*)\right\| = \frac{1}{\rho}\left\|v - \frac{w_1^* + w_2^*}{2}\right\| < 1$$

or

$$\left\|v - \frac{w_1^* + w_2^*}{2}\right\| < \rho.$$

This contradicts the definition of ρ; therefore, the theorem is proved. ∎

It now becomes of interest to determine which spaces have strictly convex norms. Let us examine the spaces of Examples I.1–I.3. Turning first to Example I.3 with the norm defined by (I.11), we suppose that $f_1, f_2 \in C[a, b]$, $f_1 \neq f_2$,

$$\|f_1\| = \left[\int_a^b |f_1(x)|^p \, dx\right]^{1/p} = 1, \qquad \text{and} \qquad \|f_2\| = \left[\int_a^b |f_2(x)|^p \, dx\right]^{1/p} = 1.$$

The triangle inequality implies that $\|\lambda_1 f_1 + \lambda_2 f_2\| \leq 1$ if $\lambda_1, \lambda_2 > 0$, $\lambda_1 + \lambda_2 = 1$. Suppose $\|\lambda_1 f_1 + \lambda_2 f_2\| = 1$ for some $\lambda_1, \lambda_2 > 0$ satisfying $\lambda_1 + \lambda_2 = 1$. We shall show that this is impossible if $p > 1$. To this end, we need the following lemma.

LEMMA I.1. *If $A > 0$ and $B > 0$ and $0 \leq t \leq 1$, then*

$$A^t B^{1-t} \leq tA + (1 - t)B, \tag{I.13}$$

and equality holds in (I.13) *only if $t = 0$ or* 1, *or $A = B$.*

Proof. The second derivative of the function $\log(1/x)$ is $1/x^2$ which is positive for positive x. Hence, $\log(1/x)$ is a *convex* function of x for positive

x; that is, the chord joining two points of the curve $y = \log(1/x)$ lies above the curve between the two points. This means that, for $0 \le t \le 1$ and $A, B > 0$,

$$\log \frac{1}{tA + (1-t)B} \le t \log \frac{1}{A} + (1-t) \log \frac{1}{B}$$

with equality possible only if $t = 0, 1$ or $A = B$. The lemma now follows by exponentiation.

We return to the proof of the theorem.

$$\begin{aligned} |\lambda_1 f_1 + \lambda_2 f_2|^p &= |\lambda_1 f_1 + \lambda_2 f_2|\, |\lambda_1 f_1 + \lambda_2 f_2|^{p-1} \\ &\le \lambda_1 |f_1|\, |\lambda_1 f_1 + \lambda_2 f_2|^{p-1} + \lambda_2 |f_2|\, |\lambda_1 f_1 + \lambda_2 f_2|^{p-1}. \end{aligned} \tag{I.14}$$

Equality is possible in (I.14) only if $f_1(x) f_2(x) \ge 0$ for $a \le x \le b$. Let us take

$$A = |f_1|^p, \qquad B = |\lambda_1 f_1 + \lambda_2 f_2|^p,$$

and

$$t = \frac{1}{p}$$

in (I.13); then

$$|f_1|\, |\lambda_1 f_1 + \lambda_2 f_2|^{p-1} \le \frac{|f_1|^p}{p} + \left(1 - \frac{1}{p}\right) |\lambda_1 f_1 + \lambda_2 f_2|^p \tag{I.15}$$

with equality only if $|f_1| = |\lambda_1 f_1 + \lambda_2 f_2|$. Similarly,

$$|f_2|\, |\lambda_1 f_1 + \lambda_2 f_2|^{p-1} \le \frac{|f_2|^p}{p} + \left(1 - \frac{1}{p}\right) |\lambda_1 f_1 + \lambda_2 f_2|^p \tag{I.16}$$

with equality only if $|f_2| = |\lambda_1 f_1 + \lambda_2 f_2|$. Using these inequalities in (I.14) and integrating, we obtain

$$\begin{aligned} 1 = \int_a^b |\lambda_1 f_1 + \lambda_2 f_2|^p \, dx &\le \frac{\lambda_1}{p} \int_a^b |f_1|^p \, dx + \frac{\lambda_2}{p} \int_a^b |f_2|^p \, dx \\ &\quad + \left(1 - \frac{1}{p}\right) \int_a^b |\lambda_1 f_1 + \lambda_2 f_2|^p \, dx = 1; \end{aligned}$$

hence, equality must hold in (I.14), (I.15), and (I.16). But the equality in (I.15) and (I.16) means that $|f_1| = |f_2|$ and equality in (I.14) that $f_1 f_2 \ge 0$; it follows that $f_1 = f_2$, contrary to our assumption. The proof is complete. ∎

In the case $p = 1$, the norm is no longer strictly convex. For example, let $f_1(x) = \frac{3}{2}x^2$, $f_2(x) = \frac{3}{4}(1 - x^2)$; then

$$\int_{-1}^{1} |f_1(x)| \, dx = \int_{-1}^{1} |f_2(x)| \, dx = 1,$$

and

$$\int_{-1}^{1} \left| \frac{f_1(x) + f_2(x)}{2} \right| dx = \frac{3}{8} \int_{-1}^{1} (1 + x^2) \, dx = 1.$$

The norm (I.12) is also strictly convex for $p > 1$. The proof proceeds exactly as in the case of the integral p-norm except that integration is replaced by summation over the points $x_1, \ldots, x_m$. When $p = 1$, the norm is not strictly convex and the uniqueness problem remains open.

The uniform norm (Examples I.1, I.2) is not strictly convex. For example, let $f_1(x) = x, f_2(x) = x^2$; then

$$\max_{0 \le x \le 1} |f_1(x)| = \max_{0 \le x \le 1} |f_2(x)| = 1$$

and

$$\max_{0 \le x \le 1} \left| \frac{f_1(x) + f_2(x)}{2} \right| = \frac{1}{2} \max_{0 \le x \le 1} |x + x^2| = 1.$$

The same is true for the discrete case (Example I.2). Thus, Theorem I.3 is uninformative in these cases, and the uniqueness question will have to receive special consideration.

Exercises

I.1 Show that $\|\cdot\|$ is a continuous function at each point v_0 of V in the sense that, given $\varepsilon > 0$, there exists $\delta > 0$ such that $|\,\|v\| - \|v_0\|\,| < \varepsilon$ whenever $\|v - v_0\| < \delta$. (In fact, $\delta = \varepsilon$.)

I.2 Show that

$$\int_a^b |f(x)g(x)|\, dx \le \left[\int_a^b |f(x)|^p\, dx\right]^{1/p} \cdot \left[\int_a^b |g(x)|^q\, dx\right]^{1/q},$$

where $p > 1$ and

$$(1/p) + (1/q) = 1.$$

Similarly, show that

$$\sum_{i=1}^n |a_i b_i| \le \left[\sum_{i=1}^n |a_i|^p\right]^{1/p} \cdot \left[\sum_{i=1}^n |b_i|^q\right]^{1/q}.$$

These inequalities are called Hölder's inequalities. The case $p = q = 2$ is usually called Schwarz's inequality.

[*Hint:* Use Lemma I.1.]

I.3 Show that if $p \ge 1$

$$\left[\int_a^b |f(x) + g(x)|^p\, dx\right]^{1/p} \le \left[\int_a^b |f(x)|^p\, dx\right]^{1/p} + \left[\int_a^b |g(x)|^p\, dx\right]^{1/p},$$

and

$$\left[\sum_{i=1}^n |a_i + b_i|^p\right]^{1/p} \le \left[\sum_{i=1}^n |a_i|^p\right]^{1/p} + \left[\sum_{i=1}^n |b_i|^p\right]^{1/p}.$$

These inequalities are called Minkowski's inequalities.

I.4 Show that (I.11) is, indeed, a norm in $C[a, b]$.

I.5 Show that (I.12) is not a strictly convex norm for $p = 1$.

I.6 Show that

$$\|f\| = \max_{i=1,\ldots,m} |f_i|$$

is not a strictly convex norm.

I.7 Prove: If V is a normed linear space, W a finite-dimensional subspace of V, and U a *closed* subset of W, then, given $v \in V$, there exists $u^* \in U$ such that $\|v - u^*\| \leq \|v - u\|$ for all $u \in U$.

I.8 A trigonometric polynomial of degree, at most, n, is an expression of the form

$$t_n(\theta) = \sum_{k=0}^{n} (\alpha_k \cos k\theta + \beta_k \sin k\theta). \tag{I.17}$$

Show that, if $f(\theta)$ is continuous for $0 \leq \theta \leq 2\pi$, there exists a trigonometric polynomial of degree at most n, $t_n^*(\theta)$, such that

$$\max_{0 \leq \theta \leq 2\pi} |f(\theta) - t_n^*(\theta)| \leq \max_{0 \leq \theta \leq 2\pi} |f(\theta) - t_n(\theta)|$$

for all t_n of the form (I.17).

I.9 The norm defined in (I.11) can be generalized by using a weight function. Namely, if $w(x)$ is a continuous nonnegative function on $[a, b]$, take

$$\|f\| = \left[\int_a^b |f(x)|^p w(x)\, dx\right]^{1/p}.$$

Show that this norm has all the properties we proved (I.11) to have.

I.10 Show that, if W is closed, so is W^* in Theorem I.2.

I.11 Show that in Theorem I.2 W^* is a bounded set. Indeed, if $w \in W^*$, then $\|w\| \leq 2\|v\|$.

CHAPTER 1

UNIFORM APPROXIMATION

This chapter is devoted to the study of best approximations in the uniform norm. We first discuss how well continuous functions can be approximated by polynomials. Then we investigate the properties that characterize a best approximating polynomial. The scene next shifts to approximation on finite point sets and its relationship to approximation on an interval. This leads us into a discussion of computational methods for obtaining best uniform approximations numerically.

1.1 Uniform Approximation by Polynomials

1.1.1 The Weierstrass Theorem and Bernstein Polynomial Approximation

If P_n denotes the space of polynomials of degree at most n, then Theorem I.1 assures us that, given $f \in C[a, b]$, there exists a polynomial $p_n^* \in P_n$ such that

$$\|f - p_n^*\| \leq \|f - p\|, \qquad \text{all} \qquad p \in P_n,$$

where $\|\cdot\|$ is the uniform norm over the interval $[a, b]$, that is,

$$\|g\| = \max_{a \leq x \leq b} |g(x)|$$

for any $g \in C[a, b]$.

Let us put

$$E_n(f; [a, b]) = E_n(f) = \|f - p_n^*\|.$$

The first question we consider is: What is the behavior of $E_n(f)$ as $n \to \infty$? We shall show that $E_n(f) \to 0$ as $n \to \infty$ for each function $f(x)$ continuous on $[a, b]$. That is, a continuous function on a finite interval can be approximated uniformly within any preassigned error by polynomials. This result is the famous Weierstrass approximation theorem, which we state as follows.

THEOREM 1.1.[(1)]† *Given $f(x)$ continuous on $[a, b]$ and $\varepsilon > 0$, there exists a polynomial, $p(x)$, such that*

$$\|f(x) - p(x)\| < \varepsilon.$$

† Exponent numbers in parentheses refer to notes at the end of the chapters.