THE FUNCTIONS OF MATHEMATICAL PHYSICS

HARRY HOCHSTADT

Department of Mathematics
Polytechnic Institute of New York

Dover Publications, Inc.
New York

THE FUNCTIONS OF MATHEMATICAL PHYSICS

HARRY HOCHSTADT

Department of Mathematics Polytechnic Institute of New York

Dover Publications, Inc.
New York

This Dover edition, first published in 1986, is an unabridged, corrected republication of the work originally published by Wile Interscience, New York, 1971, as volume XXIII in the series Pure and Applied Mathematics, edited by R. Courant. L. Bers and J. Stoker.

Library of Congress Cataloging-in-Publication Data

Hochstadt, Harry.

The functions of mathematical physics.

"Unabridged, corrected republication . . . originally published by Wiley Interscience, New York, 1971, as volume XXIII in t series Pure and applied mathematics"—CIP t.p. verso.

Bibliography: p.

Includes index.

1. Functions, Special. I. Title.

QA351.H68 1986 530.1′5 86-6383

ISBN-13: 978-0-486-65214-6

ISBN-10: 0-486-65214-9

Manufactured in the United States by Courier Corporation

65214905

www.doverpublications.com

Foreword

The functions of mathematical physics show in a striking manner how a simple mathematical ide conceived for the solution of a very specific problem, can lead to a far-reaching theory. The olde example is the trigonometric functions which describe the uniform motion on a circle. Trigonomet deals effectively with some specific problems of astronomy and navigation. But the functions sin and cos x are also the basis for the theory of Fourier series and of the Fourier integral. And this theor apart from its eminent mathematical interest, has applications to many parts of physics whic emerged long after the original problem of describing the motion of the planets had been solved. B even all of these successes of the trigonometric functions do not indicate a surprising numbe theoretical property of the function $w = \exp 2i\pi z$. The values of z and w will both be algebra numbers if and only if z is rational. And the values of w for rational values of z provide us with th means of constructing exactly those finite algebraic extensions of the rational numbers which have a abelian Galois group.

The functions of mathematical physics, which have been studied since the end of the eighteen century, have an elementary theory that is much more involved than trigonometry. But they, too, ar part of, and frequently the motivation for, important general theories, filling the general framewo with substance. And they have many aspects, turning up for example as complete sets of orthogon polynomials and also in approximation and interpolation theory. Their applicability has kept up wi the times, and they are as important for the quantum theoretical model of the atom as for the vibratin membrane. Finally, some of them also enter into parts of mathematics far removed fro mathematical physics, such as the theory of discontinuous groups.

WILHELM MAGNU

New York University

Preface

The topics covered in this book were first studied by the outstanding mathematicians of the eighteen and nineteenth centuries. Among the many who devoted themselves to these studies are Gauss, Eule Fourier, Legendre, and Bessel. These men did not recognize the modern and somewhat artifici distinction between pure and applied mathematics. Much of their work was stimulated by physic problems that led to studies of differential equations. Frequently they developed generalizations obtain results having no immediate or obvious applications. As a consequence mathematics was ofte ahead of its time in having necessary tools ready before physicists and engineers felt the need f them. This book reflects this historic interplay by presenting topics of obvious interest to applie scientists as well as topics that are, for the present at least, of purely mathematical interest.

In order to cover all the topics of the nine chapters of this book a great deal of selectivity had to b used. Naturally my personal bias played a strong role in making these choices. Entire books have bee written on some of the chapters and no claim can be made of having provided an exhaustive treatmen It is hoped, however, that the selection of results and applications, and the methodology used wi enable the reader to go on to the more specialized treatises and handbooks with ease, and to deriv new results when necessary.

Chapters 1 and 2 are devoted to orthogonal polynomials. In addition to deriving the basic resul they cover various applications to numerical integration and approximation problems. Section 10 Chapter 1 discusses applications of orthogonal polynomials to the construction of conform mappings. Chapter 3 covers the principal properties of the gamma function. Most of this material standard, but necessary for later chapters. The final section of this chapter is devoted to the solution certain algebraic equations. This topic deserves to be better known, and hopefully readers will follo this up on their own.

The results of Chapter 4, on the hypergeometric equation, are needed for the later chapters. Sectio 4.6 through 4.9, however, have no obvious applications in mathematical physics and are not neede for later chapters. They were included only because they were mathematically pleasing. They show t relationship between the theory of conformal mapping of curvilinear polygons and the theory hypergeometric functions. Section 4.10, on nonlinear transformations, is useful in studying t analytic continuation of Legendre functions in the complex domain.

Chapter 5 is concerned with the Legendre functions and some of their applications to solutions Laplace's equation in spherical coordinates. Chapter 6 continues this theme in an n-dimension setting. The functions discussed in these chapters can be expressed in terms of hypergeometr functions.

Chapter 7 takes up the confluent hypergeometric functions that arise as limiting cases of t standard hypergeometric functions. Among the most important of the confluent functions are t Bessel functions found in Chapter 8. This is the longest chapter in the book. One might well argue th such a chapter is superfluous, since one can look up Watson's classic treatise on the subjec Surprisingly, however, there are gaps in Watson's book. For example, his proof of the Fourier-Besse series is valid only for $\nu > -\frac{1}{2}$. A complete proof for $\nu > -1$ is found in Section 8.13.

The ninth and final chapter deals with Hill's equation. In historical terms, this material is the mo ~~recent and so far the least adequately covered in the literature. The recent book by Magn~~us ar Winkler is most up to date, but contains no applications to physics. Section 9.8 demonstrates th connection between the study of energy bands in crystals and Hill's equation. Section 9.4 is devoted the expansion theorems. There are two types, namely, one for the interval $(-\pi, \pi)$ and one for th interval $(-\infty, \infty)$. The former corresponds to the study of a regular differential operator and the latt to that of a singular operator. The first is proved by a standard comparison technique that treats th eigenfunctions as perturbations of trigonometric functions. The singular case is treated as a limitin case of the regular case.

While little claim can be made to originality, it is hoped that there is enough distinction in th selection of material and the type of proof to throw new light on this classical subject. The aim was present a range of topics such that both mathematicians and applied scientists with a variety interests will find material that is useful, and mathematically and aesthetically pleasing.

HARRY HOCHSTAI

September 1970
Brooklyn, New York

Contents

Orthogonal Polynomials

1. Linear Spaces

Although this chapter concerns itself with orthogonal polynomials, which can be discussed witho reference to the subject of linear spaces, it will prove advantageous to frame much of the followi discussion in more abstract language. This will not only allow for greater economy of notation, b also will lead to a richer understanding of many of the notions of abstract spaces which pervade much of modern mathematics. To make the contents of this chapter self-contained all necessa definitions and basic proofs will be provided.

Definition. A linear space over a field F (in our work F will invariably be the field of real complex numbers) is a collection X of elements with two defined operations. The first of these addition of elements in X and the second multiplication of elements in X by scalars in F. In additi we stipulate the following conditions:

1. X forms a commutative group under the additive operation in X. That is if $f, g, h, \ldots \in X$ then
   (a) the operation is closed so that $f + g \in X$,
   (b) the operation is associative:
   $$(f + g) + h = f + (g + h),$$
   (c) there exists an identity 0 for which
   $$f + 0 = f \qquad \text{for} \quad \text{all}\, f \in X,$$
   (d) for every f there exists an inverse element denoted by $(-f)$ such that
   $$f + (-f) = 0,$$
   (e) the operation is commutative
   $$f + g = g + f \qquad \text{for} \quad \text{all}\, f, g \in X.$$
2. Multiplication by scalars is closed. That is,
   (a) $1 \cdot f = f$ for all $f \in X$,
   (b) $\alpha f \in X$ for $\alpha \in F$ and $f \in X$,
   (c) for all $\alpha, \beta \in F$, and $f \in X$ $\alpha(\beta f) = (\alpha\beta) f$.
3. The following distributive laws hold.
   (a) $\alpha(f + g) = \alpha f + \alpha g$ for all $\alpha \in F, f, g \in X$.
   (b) $(\alpha + \beta) f = \alpha f + \beta f$.

In all our applications we shall be concerned with inner product spaces. In such spaces an inn product is defined.

Definition. An inner product is a function that assigns to every pair of elements $f, g \in X$ a a compl number. We denote this function by (f, g). By definition it has the following properties:

1. $\overline{(f, g)} = (g, f)$.

2. $(\alpha f + \beta g, h) = \alpha(f, h) + \beta(g, h)$.

3. $(f, f) \geq 0$ and $(f, f) = 0$ if and only if $f \neq 0$.

The norm of a vector $\|f\|$ is defined by

$$\|f\| = (f, f)^{1/2}.$$

It has the following properties:

1. $\|f\| \geq 0$ and $\|f\| = 0$ if and only if $f = 0$.

2. $\|\alpha f\| = |\alpha|\,\|f\|$.

3. $\|f + g\| \leq \|f\| + \|g\|$.

(This inequality is known as the triangle inequality.) Property 1 is an immediate consequence of th definition. To verify property 2 we first note that

$$(f, \alpha g) = \overline{(\alpha g, f)} = \bar{\alpha}\overline{(g, f)} = \bar{\alpha}(f, g).$$

Then

$$\|\alpha f\| = (\alpha f, \alpha f)^{1/2} = (\alpha\bar{\alpha})^{1/2}\|f\| = |\alpha|^{1/2}\|f\|.$$

To verify property 3 we first prove the following lemma.

LEMMA For all $f, g \in X$ we have

$$|(f, g)| \leq \|f\|\,\|g\|.$$

Equality is achieved if and only if f and g are linearly dependent; that is, for suitable scalars α and $\alpha f + \beta g = 0$. The above is known as the Cauchy-Schwarz inequality. To prove it let

$$(f, g) = |(f, g)|e^{i\theta}, \qquad \alpha = \frac{|(f, g)|}{\|f\|^2}\,e^{-i\theta}$$

We assume that $f \neq 0$; in that eventuality the inequality is trivially true. Then

$$0 \leq (\alpha f - g, \alpha f - g)\|f\|^2 = \|f\|^2\|g\|^2 - |(f, g)|^2,$$

and the result follows immediately from the above.

To complete the proof we note that if f and g are linearly independent

$$0 < \|\alpha f - g\|$$

for all α and the inequality is strict. Otherwise if, for example,

$$f = \beta g, \qquad \text{where} \quad \beta = |\beta|e^{i\theta}$$

$$(f, g) = \beta\|g\|^2 = |\beta|\,\|g\|^2 e^{i\theta}$$

$$\alpha = \frac{|\beta|\,\|g\|^2}{|\beta|^2\|g\|^2}\,e^{-i\theta} = \beta^{-1}$$

so that

$$0 = \|f\|^2\|g\|^2 - |(f, g)|^2$$

and equality is achieved. ■

To prove the triangle inequality we note that, using the lemma, we have

$$\begin{aligned}\|f+g\| &= (f+g, f+g)^{1/2} \\ &= [\|f\|^2 + (f, g) + \overline{(f, g)} + \|g\|^2]^{1/2} \\ &\le [\|f\|^2 + 2|(f, g)| + \|g\|^2]^{1/2} \le \|f\| + \|g\|.\end{aligned}$$

Definition. If for two nonzero elements f, g we have $(f, g) = 0$ we say that f and g are orthogonal.

Definition. A set of vectors $f_1, f_2, \ldots, f_n$ is said to be linearly independent if no nontrivial line relationship between them can exist. In other words the relationship

$$\alpha_1 f_1 + \alpha_2 f_2 + \cdots + \alpha_n f_n = 0$$

can only hold if $\alpha_1 = \alpha_2 = \ldots = \alpha_n = 0$.

A linearly independent set of vectors $f_1, f_2, \ldots, f_n$ will be said to be orthogonal if

$$(f_i, f_j) = 0 \qquad \text{for all } i, j \text{ such that } i \neq j.$$

If in addition to the above we have

$$(f_i, f_i) = 1$$

we say that the set is orthonormal.

The set of all vectors of the form

$$f = \sum_{i=1}^{n} \alpha_i f_i, \qquad (1)$$

where the $\{f_i\}$ represents a fixed set of vectors, clearly forms a linear space. If the $\{f_i\}$ are linear independent they form a so-called basis for this space, and the dimension of the space is said to be Any element f can be expressed in the form (1). To find the coefficients we merely take the inn product with f_j and thus obtain a system of n equations in n unknowns for the α_i.

$$\sum_{i=1}^{n} \alpha_i (f_i, f_j) = (f, f_j), \qquad j = 1, 2, \ldots, n. \qquad (2)$$

That the above system has a unique solution is a consequence of the fact that the determinant

$$D = |(f_i, f_j)| \neq 0.$$

To see this we note that for $f = 0$, the homogeneous system (2) can have only the solution $\alpha_1 = a_2 = .$ $= \alpha_n = 0$, since the $\{f_i\}$ are linearly independent. As a result $D \neq 0$ necessarily.

If the system $\{f_i\}$ were orthonormal, the system (2) could be solved by inspection, since in that ca it reduces to

$$\alpha_i = (f, f_i)$$

It is evident that orthonormal systems are more convenient to work with than general systems. Giv any linearly independent system $\{f_i\}$ we can replace it by an equivalent orthonormal system. One wa of constructing such a system is by the Gram-Schmidt process.

We now let

$$e_1 = \frac{f_1}{\|f_1\|}$$

$$e_2 = \frac{f_2 - (f_2 \,.\, e_1)e_1}{\|f_2 - (f_2 \,,\, e_1)e_1\|}$$

and in general

$$e_k = \frac{f_k - \sum_{i=1}^{k-1}(f_k \,,\, e_i)e_i}{\|f_k - \sum_{i=1}^{k-1}(f_k \,,\, e_i)e_i\|}, \qquad k = 1, 2, \ldots, n. \tag{3}$$

Clearly

$$(e_k \,,\, e_k) = 1$$

and

$$(e_2 \,,\, e_1) = \frac{(f_2 \,,\, e_1) - (f_2 \,,\, e_1)(e_1 \,,\, e_1)}{\|f_2 - (f_2 \,,\, e_1)e_1\|} = 0.$$

By induction we can easily verify that for $k \neq j$

$$(e_k \,,\, e_j) = 0.$$

It is also evident that

$$\begin{aligned} f_k &= \sum_{i=1}^{k} \alpha_{i,k} e_i \\ e_k &= \sum_{i=1}^{k} \beta_{i,k} f_i \end{aligned} \tag{4}$$

and $(e_k, f_j) = 0$ for $j = 1, 2, \ldots, k - 1$.

One disadvantage of the Gram-Schmidt process is that it is recursive. However, an explic expression for the e_k can be obtained. Consider the determinant

$$d_k = \begin{vmatrix} (f_1, f_1) & (f_1, f_2) & \cdots & (f_1, f_k) \\ (f_2, f_1) & (f_2, f_2) & \cdots & (f_2, f_k) \\ \vdots & & & \\ (f_{k-1}, f_1) & (f_{k-1}, f_2) & \cdots & (f_{k-1}, f_k) \\ f_1 & f_2 & \cdots & f_k \end{vmatrix}$$

Using (4), for $k > j$, we see that

$$(e_j \,,\, d_k) = \sum_{i=1}^{j} \beta_{i,j}(f_i \,,\, d_k) = 0$$

since in (f_i, d_k) the kth and ith rows are identical. It follows therefore that

$$e_k = \frac{d_k}{\|d_k\|}.$$

2. Orthogonal Polynomials

The results and concepts of the preceding section can now be applied to a study of orthogon polynomials. For our field F we shall select the field of real numbers, and for our space X the set of a polynomials in the variable x with real coefficients. Using the standard operations of addition ar

multiplication we obtain a linear space. It is an elementary exercise to verify all axioms of th
preceding section.

By selecting an arbitrary interval (a, b) and restricting the values of x to that interval we obta another linear space. We do not preclude the possibility that $b = \infty$, or $a = -\infty$, or both. As yet th space is not an inner product space. To define such a product we select a positive function $w($ defined on (a, b) for which the integral

$$\int_a^b w(x)p(x)\,dx$$

exists for all polynomials $p(x)$.

More generally if the above integral is taken in the sense of Lebesgue $w(x)$ need be positi everywhere except on a set of measure zero. We can now introduce an inner product as follows. If $f($ and $g(x)$ are polynomials

$$(f, g) = \int_a^b w(x)f(x)g(x)\,dx. \tag{1}$$

The above can be shown to satisfy all properties of inner products, as long as $f(x)$ and $g(x)$ are rea More generally we could use

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

but we will restrict ourselves to real spaces. From (1) it is clear that

$$(f, g) = \overline{(g, f)} = (g, f)$$

in view of the fact that we are dealing with real quantities. Also

$$(\alpha f + \beta g, h) = \alpha(f, h) + \beta(g, h).$$

Lastly we note that

$$(f, f) = \int_a^b w(x)f^2(x)\,dx \geq 0$$

and $(f,f) = 0$ implies that

$$w(x)f^2(x) = 0, \text{ almost everywhere.}$$

Since $f(x)$ is a polynomial, the latter implies that $f(x) \equiv 0$.

To construct a family of orthonormal polynomials we can proceed as follows. First we select as linearly independent set the polynomials

$$1, x, x^2, \ldots, x^n.$$

That these are indeed linearly independent is a consequence of the fact that

$$\alpha_0 + \alpha_1 x + \cdots + \alpha_n x^n \equiv 0, \qquad x \in (a, b)$$

if and only if $\alpha_0 = \alpha_1 = \alpha_2 = \ldots = \alpha_n = 0$.

Using either the Gram-Schmidt process I.1.3 or the direct formula I.1.5 we can construct a orthonormal set $\varphi_k(x)$, where $\varphi_k(x)$ is a polynomial of precise degree k.

EXAMPLE Let $(a, b) = (-1, 1)$ and $w(x) = 1$. Then

$$\phi_0(x) = \tfrac{1}{2}^{1/2}$$

$$d_1(x) = \begin{vmatrix} (1,1) & (1,x) \\ 1 & x \end{vmatrix} = \begin{vmatrix} 2 & 0 \\ 1 & x \end{vmatrix} = 2x$$

$$\phi_1(x) = \frac{2x}{\|2x\|} = \tfrac{3}{2}^{1/2} x$$

$$d_2(x) = \begin{vmatrix} (1,1) & (1,x) & (1,x^2) \\ (x,1) & (x,x) & (x,x^2) \\ 1 & x & x^2 \end{vmatrix}$$

$$= \begin{vmatrix} 2 & 0 & 2/3 \\ 0 & 2/3 & 0 \\ 1 & x & x^2 \end{vmatrix} = 4/3x^2 - 4/9$$

$$\phi_2(x) = \frac{4/3x^2 - 4/9}{\|4/3x^2 - 4/9\|} = \tfrac{5}{8}^{1/2}(3x^2 - 1)$$

More generally, as will be shown in the next chapter, the classical Legendre polynomials $P_n(x)$ a related to the above by

$$P_n(x) = \sqrt{\frac{2n+1}{2}}\,\phi_n(x).$$

Note that the orthonormality requirement does not define these polynomials uniquely, but only u to the choice of a sign. But by also stipulating that the coefficient of the highest power of x be positi a unique polynomial is defined.

We also note for future reference that, using Eq. 1.4,

$$(\phi_n, x^k) = \int_a^b w(x)\phi_n(x)x^k\,dx = 0, \qquad k < n. \tag{2}$$

3. The Recurrence Formula

We let k_n denote the coefficients of x^n in $\varphi_n(x)$, so that

$$\phi_n(x) = k_n x^n + \cdots. \tag{1}$$

THEOREM The orthonormal polynomials $\{\varphi_n\}$ satisfy the recurrence formula

$$\phi_{n+1} - (A_n x + B_n)\phi_n + C_n\phi_{n-1} = 0 \qquad n = 0, 1, 2, \ldots, \tag{2}$$

where

$$A_n = \frac{k_{n+1}}{k_n}, \qquad C_n = \frac{A_n}{A_{n-1}}, \qquad C_0 = 0. \tag{3}$$

Proof. From (1) and (3) it is evident that

$$\phi_{n+1} - A_n x\phi_n = \sum_{k=0}^{n} \alpha_k x^k = \sum_{k=0}^{n} \beta_k \phi_k(x)$$

for suitable α_k and β_k, since the x^{n+1} term on the left has been cancelled out. Using the orthonormali we obtain from the above

$$-A_n(x\phi_n, \phi_j) = \beta_j \qquad j \leq n.$$

Reference to I.2.2. shows that

$$(x\phi_n, \phi_j) = \int_a^b w(x)x\phi_n(x)\phi_j(x)\, dx = (\phi_n, x\phi_j) = 0 \qquad j \leq n-2$$

since $x\varphi_j(x)$ is at most of degree $n-1$. It follows that

$$\beta_0 = \beta_1 = \cdots = \beta_{n-2} = 0,$$

and we let $\beta_n = B_n$, $\beta_{n-1} = -C_n$, which proves the validity of (2). To obtain the explicit value of C_n w
note that

$$C_n = A_n(x\phi_n, \phi_{n-1}) = A_n(\phi_n, x\phi_{n-1})$$

but

$$x\phi_{n-1} = k_{n-1}x^n + \cdots = \frac{k_{n-1}}{k_n}[k_n x^n + \cdots]$$

and the bracketed expression must be of the form

$$x\phi_{n-1} = \frac{1}{A_{n-1}}\left[\phi_n(x) + \sum_{j=0}^{n-1}\gamma_j\phi_j(x)\right].$$

It follows that

$$C_n = \frac{A_n}{A_{n-1}}\left(\phi_n, \phi_n + \sum_{j=0}^{n-1}\gamma_j\phi_j\right) = \frac{A_n}{A_{n-1}}. \qquad \blacksquare$$

4. The Christoffel-Darboux Formula

A function that plays a significant role in many aspects of this subject is

$$K_n(x, y) = \sum_{k=0}^{n}\phi_k(x)\phi_k(y).$$

It is possible to find a closed form for the above sum.

LEMMA

$$K_n(x, y) = \sum_{k=0}^{n}\phi_k(x)\phi_k(y) = \frac{k_n}{k_{n+1}}\left[\frac{\phi_n(y)\phi_{n+1}(x) - \phi_n(x)\phi_{n+1}(y)}{x-y}\right] \qquad (1)$$

Formula (1) is known as the Christoffel-Darboux formula.

Proof. Since

$$\phi_0(x) = k_0$$

$$\phi_1(x) = k_1 x - k_1 k_0^2 \int_a^b xw(x)\, dx$$

we have trivially

$$K_0(x, y) = \phi_0^2(x) = k_0^2 = \frac{k_0}{k_1}\left[\frac{\phi_0(y)\phi_1(x) - \phi_0(x)\phi_1(y)}{x-y}\right].$$

To prove the general case, we use the recurrence formula and note that

$$\begin{aligned}\frac{k_n}{k_{n+1}}&\left[\frac{\phi_n(y)\phi_{n+1}(x)-\phi_n(x)\phi_{n+1}(y)}{x-y}\right]\\&=\frac{k_n}{k_{n+1}}A_n\phi_n(x)\phi_n(y)+\frac{k_n}{k_{n+1}}C_n\left[\frac{\phi_{n-1}(y)\phi_n(x)-\phi_{n-1}(x)\phi_n(y)}{x-y}\right]\\&=\phi_n(x)\phi_n(y)+K_{n-1}(x,y)\end{aligned}$$

so that

$$K_n(x, y)=\phi_n(x)\phi_n(y)+K_{n-1}(x, y).$$

From the latter, by iteration we find

$$K_n(x, y)=\sum_{k=1}^{n}\phi_k(x)\phi_k(y)+K_0(x, y)=\sum_{k=0}^{n}\phi_k(x)\phi_k(y). \qquad ■$$

By letting y approach x in the limit we find

$$\begin{aligned}K_n(x, x)&=\lim_{y\to x}\frac{k_n}{k_{n+1}}\left[\frac{\phi_n(y)\phi_{n+1}(x)-\phi_n(x)\phi_{n+1}(y)}{x-y}\right]\\&=\frac{k_n}{k_{n+1}}[\phi_n(x)\phi'_{n+1}(x)-\phi'_n(x)\phi_{n+1}(x)]\\&=\sum_{k=0}^{n}\phi_k{}^2(x)\geq 0 \qquad (2)\end{aligned}$$

The function $K_n(x, y)$ can be characterized in terms of an extremal property independently of th polynomials $\varphi_n(x)$.

THEOREM Among all polynomials $\rho(x)$ of degree n and unit norm, the ones that maximize $|\rho(y$ where y is a prescribed point in (a, b), are given by

$$\rho(x)=\frac{\pm K_n(x, y)}{K_n{}^{1/2}(y, y)}. \qquad (3)$$

Proof. Since $\rho(x)$ is of degree n it can be represented in the form

$$\rho(x)=\sum_{k=0}^{n}\alpha_k\,\phi_k(x)$$

and since $\|\rho\|=1$ we have

$$\sum_{0}^{n}\alpha_k{}^2=1.$$

The quantity $\rho(y)$ can be construed as an inner product of the two vectors in $n+1$ dimension Euclidean space with components $\{\alpha_k\}$ and $\{\varphi_k(y)\}$. By means of the Cauchy-Schwarz inequality w see that

$$\rho^2(y)\leq\sum_{k=0}^{n}\alpha_k{}^2\sum_{k=0}^{n}\phi_k{}^2(y)=\sum_{k=0}^{n}\phi_k{}^2(y).$$

$\rho^2(y)$ will be maximized when equality is achieved. In that case the vectors $\{\alpha_k\}$ and $\{\varphi_k(y)\}$ must b linearly dependent, that is, for some scalar λ

$$\alpha_k = \lambda\phi_k(y) \qquad k = 0, 1, 2, \ldots n.$$

We see therefore that

$$\rho(x) = \lambda \sum_{k=0}^{n} \phi_k(y)\phi_k(x) = \lambda K_n(x, y).$$

To determine λ we still require that $\|\rho\| = 1$ and

$$\lambda^2 \sum_{k=0}^{n} \phi_k^{\,2}(y) = \lambda^2 K_n(y, y) = 1,$$

which shows that

$$\rho(x) = \pm \frac{K_n(x, y)}{K_n^{1/2}(y, y)}. \qquad \blacksquare$$

N.B. By (2) the denominator is real.

If $p(x)$ is a polynomial of degree n or less we have

$$p(x) = \sum_{k=0}^{n} (p, \phi_k)\phi_k(x).$$

It is also evident that

$$(p(x), K_n(x, y)) = \sum_{k=0}^{n} (p, \phi_k)\phi_k(y)) = p(y).$$

$K_n(x, y)$ is sometimes known as a reproducing kernel, as a result of the above property. In particul we note that for $p(x) = 1$ we obtain

$$(1, K_n(x, y)) = \int_a^b w(x)K_n(x, y)\, dx = 1, \qquad n \geq 0. \qquad (4)$$

5. The Weierstrass Approximation Theorem

A result that will prove to be of fundamental importance in the sequel is the Weierstras Approximation Theorem.

THEOREM Let $f(x)$ be a continuous function defined on the closed and bounded interval $[a,$ Given any $\varepsilon > 0$, it is possible to find a polynomial $p(x)$ such that

$$|f(x) - p(x)| < \varepsilon \qquad \text{for all } x \in [a, b].$$

Proof. The proof to be given will be a constructive one. By means of a sequence of so-call Bernstein polynomials we shall construct an explicit $p(x)$ satisfying the theorem.

By means of a linear transformation the finite interval $[a, b]$ into the interval $[0, 1]$. Now let

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

where $\binom{n}{k}$ is the binomial coefficient

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

We shall show that for every $\varepsilon > 0$ we can find $N(\varepsilon)$ so that

$$|f(x) - B_n(x; f)| < \varepsilon \qquad \text{for} \quad n > N(\varepsilon).$$

We shall first prove the theorem for three special cases required for the general proof.

1. By the binomial theorem we have

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

By differentiating the above with respect to x and multiplying by x/n we obtain

$$x(x + y)^{n-1} = \sum_{k=0}^{n} \binom{n}{k} \frac{k}{n} x^k y^{n-k}. \tag{2}$$

Repeating the above operation we have

$$x^2(x + y)^{n-2} + \frac{xy}{n}(x + y)^{n-2} = \sum_{k=0}^{n} \binom{n}{k} \left(\frac{k}{n}\right)^2 x^k y^{n-k}. \tag{3}$$

For $y = 1 - x$ we have from (1)

$$1 = \sum_{k=0}^{n} \binom{n}{k} x^k (1 - x)^{n-k}. \tag{4}$$

(4) shows the theorem to be correct for $f(x) = 1$, that is,

$$B_n(x; 1) = 1 \qquad \text{for} \quad \text{all } n.$$

2. For $f(x) = x$ we have, letting $y = 1 - x$ in (2)

$$B_n(x; x) = x \qquad \text{for} \quad \text{all } n. \tag{5}$$

3. For $f(x) = x^2$ we have, letting $y = 1 - x$ in (3)

$$|x^2 - B_n(x; x^2)| \le \frac{x(1-x)}{n} \le \frac{1}{4n} \qquad \text{for} \quad \text{all } n.$$

In 1., 2., 3., we proved the theorem for the cases $f(x) = 1$, $f(x) = x$, $f(x) = x^2$.
To prove the general case we note that if $|f(x)| < M$ then

$$\left|f(x) - f\left(\frac{k}{n}\right)\right| < 2M, \qquad x \in [0, 1].$$

Also by the uniform continuity of $f(x)$, given ε we can find $\delta(\varepsilon)$ so that

$$\left|f(x) - f\left(\frac{k}{n}\right)\right| < \frac{\varepsilon}{2} \qquad \text{for} \quad \left|x - \frac{k}{n}\right| < \delta(\varepsilon).$$

We now see that by (4)

$$
\begin{aligned}
|f(x) - B_n(x; f)| &= \left| f(x) - \sum_{k=0}^{n} \binom{n}{k} f\left(\frac{k}{n}\right) x^k (1-x)^{n-k} \right| \\
&= \left| \sum_{k=0}^{n} \binom{n}{k} \left[f(x) - f\left(\frac{k}{n}\right) \right] x^k (1-x)^{n-k} \right| \\
&\le \sum_{|x-k/n| < \delta(\varepsilon)} \binom{n}{k} \left| f(x) - f\left(\frac{k}{n}\right) \right| x^k (1-x)^{n-k} \\
&\quad + \sum_{|x-k/n| \ge \delta(\varepsilon)} \binom{n}{k} \left| f(x) - f\left(\frac{k}{n}\right) \right| x^k (1-x)^{n-k} \\
&= S_1 + S_2.
\end{aligned}
$$

For the first sum on the right we have evidently

$$
S_1 < \frac{\varepsilon}{2} \sum_{|x-k/n| < \delta(\varepsilon)} \binom{n}{k} x^k (1-x)^{n-k} \le \frac{\varepsilon}{2} \sum_{k=0}^{n} \binom{n}{k} x^k (1-x)^{n-k} = \frac{\varepsilon}{2}.
$$

To estimate S_2 we have, using the fact that $|x - k/n| \ge \delta(\varepsilon)$,

$$
\begin{aligned}
S_2 &< 2M \sum_{|x-k/n| \ge \delta(\varepsilon)} \binom{n}{k} x^k (1-x)^{n-k} \\
&\le \frac{2M}{\delta^2(\varepsilon)} \sum_{k=0}^{n} \binom{n}{k} \left(x - \frac{k}{n} \right)^2 x^k (1-x)^{n-k} \\
&= \frac{2M}{\delta^2(\varepsilon)} [x^2 B_n(x; 1) - 2x B_n(x; x) + B_n(x; x^2)] \\
&= \frac{2M}{\delta^2(\varepsilon)} \left[x^2 - 2x^2 + x^2 + \frac{x(1-x)}{n} \right] \le \frac{2M}{4\delta^2(\varepsilon) n}.
\end{aligned}
$$

We summarize these results in the statement

$$
|f(x) - B_n(x; f)| < \frac{\varepsilon}{2} + \frac{2M}{4\delta^2(\varepsilon) n} < \varepsilon \qquad \text{for} \quad n > \frac{M}{\delta^2(\varepsilon)\varepsilon},
$$

thus proving the theorem. ■

This proof can be given a probabilistic interpretation. We conceive the following game. n numbe are to be selected at random from the interval $[0, 1]$. Let k denote the number falling into $[0, x]$ and $-\, k$ the number falling into $(x, 1]$, where x is a preassigned quantity. The payoff is to be $f(k/n)$. T probability of k coming from the interval $[0, x]$ is clearly

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

since x is the probability of a single selection falling into $[0, x]$. The expectation of this game clearly

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

For n large the expected value of k will be nx and by the law of large numbers we expect that

$$
B_n(x; f) \approx f\left(\frac{nx}{n}\right) = f(x).
$$

In other words the Weierstrass approximation theorem is nothing but the law of large numbers for binomial distribution.

6. The Zeros of the Orthogonal Polynomials

In view of the fact that $\varphi_n(x)$ is a polynomial of degree n, we know from the fundamental theore of algebra that it must have precisely n zeros in the field of complex numbers. But we can say eve more about it, as a result of its special properties.

THEOREM $\varphi_n(x)$ has n real, simple zeros, all in the interval (a, b).

Proof. We suppose that $\varphi_n(x)$ has k zeros in the interval (a, b), where it changes sign. The latter equivalent to saying that we are examining zeros of odd multiplicity. If we can succeed in showin that $k = n$, then $\varphi_n(x)$ must have n zeros in (a, b), where it changes sign. But since $\varphi_n(x)$ has precisely zeros they must all be simple and in (a, b).

We now suppose that $k < n$. Then

$$p_k(x) = (x - x_1)(x - x_2)\cdots(x - x_k) = \sum_{j=0}^{k} \alpha_j \phi_j(x)$$

is a polynomial of degree k so that

$$(\phi_n, p_k) = 0.$$

But clearly $\varphi_n(x)p_k(x)$ cannot change sign, since it has only zeros of even multiplicity. Therefore i without loss of generality,

$$\phi_n(x)p_k(x) \geq 0$$

we have

$$(\phi_n, p_k) = \int_a^b w(x)\phi_n(x)p_k(x)\,dx > 0.$$

The contradiction can only be resolved by having $k = n$. ■

We have another result showing how zeros of successive orthonormal polynomials are related.

THEOREM The zeros of $\varphi_n(x)$ and $\varphi_{n+1}(x)$ alternate on the interval (a, b), and $\varphi_n(x)$ and $\varphi_{n+1}(x)$ do n vanish simultaneously.

Proof. Let x_r and x_{r+1} denote two successive zeros of $\varphi_n(x)$. Then by 1.4.2

$$\sum_{k=0}^{n} \phi_k^2(x_r) = \frac{k_n}{k_{n+1}}\left[-\phi_n'(x_r)\phi_{n+1}(x_r)\right] \geq 0$$

so that

$$\phi_n'(x_r)\phi_{n+1}(x_r) \leq 0$$

and also

$$\phi_n'(x_{r+1})\phi_{n+1}(x_{r+1}) \leq 0.$$

In view of the fact that x_r and x_{r+1} are successive zeros of $\varphi_n(x)$ it follows that $\varphi'_n(x_r)$ and $\varphi'_n(x_{r+}$ are of opposite sign. Therefore $\varphi_{n+1}(x_r)$ and $\varphi_{n+1}(x_{r+1})$ also have opposite sign, and being continuous function of x, $\varphi_{n+1}(x)$ must vanish at least once in $[x_r, x_{r+1}]$. Similarly between tw successive zeros of $\varphi_{n+1}(x)$ there must be at least one zero of $\varphi_n(x)$ and the statement of the theore follows.

To show that $\varphi_n(x)$ and $\varphi_{n+1}(x)$ do not vanish simultaneously, we see from the recurrence formu that if

$$\phi_{n+1}(x_r) = \phi_n(x_r) = 0$$

then also

$$\phi_{n-1}(x_r) = 0.$$

By repeating this we have finally

$$\phi_{n-2}(x_r) = \cdots = \phi_1(x_r) = \phi_0(x_r) = 0.$$

But $\varphi_0 > 0$ so that φ_n and φ_{n+1} do not vanish simultaneously. ■

7. Approximation Theory

In many aspects of pure as well as applied mathematics it is often important to investigate ho closely one member of a set of functions can be approximated by a member of another class functions. Suppose, for example, we are given a function $f(x)$ for which

$$\int_a^b w(x)f^2(x)\,dx < \infty.$$

Is it possible to approximate $f(x)$ by a finite linear combination of ortho-normal polynomials (over t interval (a, b) with weight function $w(x)$) in some optimum fashion? To decide the answer we mu select some criterion by which we can measure the quality of the approximation. One such criterion given by the norm defined in Section I.1.

We shall say that a polynomial $p_n(x)$ of degree n yields a best approximation to some $f(x)$ over t linear, inner product space X, which consists of functions $g(x)$ for which

$$\|g(x)\| = \left[\int_a^b w(x)g^2(x)\,dx\right]^{1/2}$$

if

$$\|f(x) - p_n(x)\| = \min_{\text{all } q_n(x)} \|f(x) - q_n(x)\|.$$

In the above $q_n(x)$ is an arbitrary polynomial of degree n.

THEOREM Among all polynomials of degree n, there is precisely one for which

$$\|f(x) - p_n(x)\|$$

is minimized. It is given by

$$p_n(x) = \sum_{k=0}^{n} \alpha_k \phi_k(x), \qquad \alpha_k = (f, \phi_k) = \int_a^b w(x)f(x)\phi_k(x)\,dx.$$

Proof. Let $p_n(x)$ be as defined in the theorem and $q_n(x)$ any other polynomial of degree n, say

$$q_n(x) = \sum_{k=0}^{n} \beta_k \phi_k(x).$$

Let

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

and clearly, by construction

$$(g, \phi_k) = 0, \qquad k \leq n,$$

so that

$$(g, q_n) = 0.$$

Also

$$\|g - q_n\|^2 = (g - q_n, g - q_n) = \|g\|^2 + \|q_n\|^2$$

so that

$$\min_{\text{all } q_n(x)} \|g - q_n\| = \|g\|$$

and the minimum is attained if and only if $q_n(x) \equiv 0$.

If $f(x)$ is a continuous function we can say even more. ■

THEOREM Let $f(x)$ be continuous and $p_n(x)$ as in the preceding theorem. Then $f(x) - p_n(x)$ chang sign at least $n + 1$ times in (a, b), or else vanishes identically.

Proof. Denote the points where $f(x) - p_n(x)$ changes sign by $x_1, x_2, \ldots x_k$, where we suppose that k n. Then

$$q(x) = (x - x_1)\cdots(x - x_k) = \sum_{j=0}^{k} \beta_j \phi_j(x)$$

and since

$$(f - p_n, \phi_j) = 0, \qquad j \leq n$$

by construction

$$(f - p_n, q) = 0.$$

The function $(f - p_n)q$ must be of constant sign so that either

$$(f - p_n, q) = \int_a^b w(x)(f(x) - p_n(x))q(x)\,dx$$

does not vanish, or else $f(x) - p_n(x) \equiv 0$. The former contradicts the statement $k \leq n$, proving t theorem. ■

It is often of value to find a polynomial that agrees with a given continuous function at n prescrib points. To construct such a polynomial we can make use of the Lagrange Interpolation Formula. L $f(x)$ be some continuous function on (a, b) and $\varphi_n(x)$ an orthonormal polynomial. Consider

$$F(x) = \sum_{k=1}^{n} f(x_k) \frac{\phi_n(x)}{(x - x_k)\phi_n'(x_k)}. \qquad (1)$$

Clearly $F(x)$ is a polynomial of degree $n - 1$ and also

$$F(x_k) = \lim_{x \to x_k} F(x) = f(x_k), \qquad k \leq n.$$

N.B. For this construction $\varphi_n(x)$ did not have to be an orthonormal polynomial. Any polynomial of degree n, with simple zeros would have served.

If $f(x)$ were a polynomial of degree $n - 1$, then clearly $f(x) = F(x)$ in (1).

We shall now derive the Gauss Quadrature Formula. To do so we shall assume that $f(x)$ is a polynomial of degree $2n - 1$. Then $F(x) - f(x)$ is a polynomial of degree $2n - 1$ and it vanishes at x_1, $x_2, \ldots x_n$. Therefore $r(x)$, given by

$$r(x) = \frac{F(x) - f(x)}{\phi_n(x)}$$

must be a polynomial of degree $n - 1$. We can now write

$$f(x) = \sum_{k=1}^{n} f(x_k) \frac{\phi_n(x)}{(x - x_k)\phi_n'(x_k)} - r(x)\phi_n(x).$$

By integration we find

$$\int_a^b w(x)f(x)\,dx = \sum_{k=1}^{n} f(x_k) \int_a^b w(x) \frac{\phi_n(x)}{(x - x_k)\phi_n'(x_k)} - \int_a^b w(x)r(x)\phi_n(x)\,dx.$$

We let

$$\int_a^b w(x) \frac{\phi_n(x)}{(x - x_k)\phi_n'(x_k)}\,dx = \lambda_{k,n}$$

and clearly

$$\int_a^b w(x)r(x)\phi_n(x)\,dx = 0$$

since $r(x)$ is of degree $n - 1$. The numbers $\lambda_{k,n}$ are known as the Christoffel numbers, and they are independent of $f(x)$.

From the above we have

$$\int_a^b w(x)f(x)\,dx = \sum_{k=1}^{n} f(x_k)\lambda_{k,n}. \qquad (2)$$

(2) is known as the Gauss Quadrature Formula. Evidently to find the value of the integral on the left, if $f(x)$ is a polynomial of $2n - 1$, we merely need to know the value of $f(x)$ at the n points $x_1, x_2, \ldots, x_n$. If $f(x)$ is not such a polynomial (2) may still furnish a useful approximation for the integral.

For $f(x) = x^j$ where $j \leq 2n - 1$ we have the result

$$\int_a^b w(x)x^j\,dx = \sum_{k=1}^{n} x_k{}^j\lambda_{k,n} \qquad j \leq 2n-1. \qquad (3)$$

EXAMPLE Consider