9

Approximation in the 2-norm

9.1 Introduction

In Chapter 8 we discussed the idea of best approximation of a continuous real-valued function by polynomials of some fixed degree in the ∞-norm. Here we consider the analogous problem of best approximation in the 2-norm. Why, you might ask, is it necessary to consider best approximation in the 2-norm when we have already developed a perfectly adequate theory of best approximation in the ∞-norm? As our first example in Section 9.3 will demonstrate, the choice of norm can significantly influence the outcome of the problem of best approximation: the polynomial of best approximation of a certain fixed degree to a given continuous function in one norm need not bear any resemblance to the polynomial of best approximation of the same degree in another norm. Ultimately, in a practical situation, the choice of norm will be governed by the sense in which the given continuous function has to be well approximated.

As will become apparent, best approximation in the 2-norm is closely related to the notion of orthogonality and this in turn relies on the concept of *inner product*. Thus, we begin the chapter by recalling from linear algebra the definition of *inner product space*.

Throughout the chapter $[a, b]$ will denote a nonempty, bounded, closed interval of the real line, and (a, b) will signify a nonempty bounded open interval of the real line.

https://doi.org/10.1017/CBO9780511801181.010 Published online by Cambridge University Press

9.2 Inner product spaces

Definition 9.1 *Let $\mathcal{V}$ be a linear space over the field of real numbers. A real-valued function $\langle \cdot, \cdot \rangle$, defined on the Cartesian product $\mathcal{V} \times \mathcal{V}$, is called an* **inner product** *on $\mathcal{V}$ if it satisfies the following axioms:*

❶ $\langle f+g, h\rangle = \langle f, h\rangle + \langle g, h\rangle$ *for all f, g and h in $\mathcal{V}$;*
❷ $\langle \lambda f, g\rangle = \lambda\langle f, g\rangle$ *for all λ in $\mathbb{R}$, and all f, g in $\mathcal{V}$;*
❸ $\langle f, g\rangle = \langle g, f\rangle$ *for all f and g in $\mathcal{V}$;*
❹ $\langle f, f\rangle > 0$ *if $f \neq 0$, $f \in \mathcal{V}$.*

A linear space with an inner product is called an **inner product space**.

Example 9.1 *The n-dimensional Euclidean space $\mathbb{R}^n$ is an inner product space with*

$$\langle \boldsymbol{x}, \boldsymbol{y}\rangle = \sum_{i=1}^{n} x_i y_i\,, \qquad \boldsymbol{x},\, \boldsymbol{y} \in \mathbb{R}^n\,,$$

where $\boldsymbol{x} = (x_1, \dots, x_n)^{\mathrm{T}}$ and $\boldsymbol{y} = (y_1, \dots, y_n)^{\mathrm{T}}$. We can also write this in a more compact form as $\langle \boldsymbol{x}, \boldsymbol{y}\rangle = \boldsymbol{x}^{\mathrm{T}}\boldsymbol{y}$.

Definition 9.2 *Suppose that $\mathcal{V}$ is an inner product space, and f and g are two elements of $\mathcal{V}$ such that $\langle f, g\rangle = 0$; we shall then say that f is* **orthogonal** *to g.*

Due to the third axiom of inner product, if f is orthogonal to g, then g is orthogonal to f; therefore, if $\langle f, g\rangle = 0$, we shall simply say that f and g are orthogonal. Our next example shows that Definition 9.2 is a direct generalisation of the usual geometrical notion of orthogonality.

Example 9.2 *According to Example 9.1, with $n = 2$, the formula $\langle \boldsymbol{y}, \boldsymbol{z}\rangle = \boldsymbol{y}^{\mathrm{T}}\boldsymbol{z}$, where $\boldsymbol{y} = (y_1, y_2)^{\mathrm{T}}$ and $\boldsymbol{z} = (z_1, z_2)^{\mathrm{T}}$ are two-component vectors, defines an inner product in $\mathbb{R}^2$.*

The vectors $\boldsymbol{y}$ and $\boldsymbol{z}$ have respective lengths $\sqrt{y_1^2 + y_2^2} = \|\boldsymbol{y}\|_2$ and $\sqrt{z_1^2 + z_2^2} = \|\boldsymbol{z}\|_2$, where $\|\cdot\|_2$ denotes the 2-norm for vectors in $\mathbb{R}^2$. Let $\alpha \in [0, 2\pi)$ denote the angle, measured in an anticlockwise direction, between the positive x_1-coordinate direction and $\boldsymbol{y}$; similarly, let $\beta \in [0, 2\pi)$ be the angle between the positive x_1-coordinate direction and $\boldsymbol{z}$. Then,

$$\boldsymbol{y} = \|\boldsymbol{y}\|_2(\cos\alpha, \sin\alpha) \quad \textit{and} \quad \boldsymbol{z} = \|\boldsymbol{z}\|_2(\cos\beta, \sin\beta)\,.$$

https://doi.org/10.1017/CBO9780511801181.010 Published online by Cambridge University Press

Now,

$$\begin{aligned}\langle \boldsymbol{y}, \boldsymbol{z} \rangle &= \boldsymbol{y}^{\mathrm{T}} \boldsymbol{z} \\ &= \|\boldsymbol{y}\|_2 \|\boldsymbol{z}\|_2 \left(\cos\alpha \cos\beta + \sin\alpha \sin\beta\right) \\ &= \|\boldsymbol{y}\|_2 \|\boldsymbol{z}\|_2 \cos(\alpha - \beta) \\ &= \|\boldsymbol{y}\|_2 \|\boldsymbol{z}\|_2 \cos(\vartheta_{yz})\,,\end{aligned}$$

where $\vartheta_{yz} = |\alpha - \beta|$ *is the angle between the vectors* $\boldsymbol{y}$ *and* $\boldsymbol{z}$*. The vector* $\boldsymbol{y}$ *is orthogonal to* $\boldsymbol{z}$ *if, and only if,* ϑ_{yz} *is* $\pi/2$ *or* $3\pi/2$*; either way,* $\cos(\vartheta_{yz}) = 0$*, and hence* $\langle \boldsymbol{y}, \boldsymbol{z} \rangle = 0$*. We note in passing that if* $\boldsymbol{y} = \boldsymbol{z}$*, then* $\vartheta_{yz} = 0$ *and therefore*

$$\langle \boldsymbol{y}, \boldsymbol{y} \rangle = \|\boldsymbol{y}\|_2^2\,.$$

This last observation motivates our next definition.

Definition 9.3 *Suppose that* $\mathcal{V}$ *is an inner product space over the field of real numbers, with inner product* $\langle\,\cdot\,,\cdot\,\rangle$*. For* f *in* $\mathcal{V}$*, we define the* ***induced norm***

$$\|f\| = \langle f, f \rangle^{1/2}\,. \tag{9.1}$$

Although our terminology and our notation appear to imply that (9.1) defines a norm on $\mathcal{V}$, this is by no means obvious. In order to show that $f \mapsto \langle f, f\rangle^{1/2}$ is indeed a norm, we begin with the following result which is a direct generalisation of the Cauchy–Schwarz inequality (2.35) from Chapter 2.

Lemma 9.1 (Cauchy–Schwarz inequality)

$$|\langle f, g \rangle| \le \|f\|\,\|g\| \qquad \forall\, f,\, g \in \mathcal{V}\,. \tag{9.2}$$

Proof The proof is analogous to that of (2.35). Recalling the definition of $\|\cdot\|$ from (9.1) and noting the first three axioms of inner product, we find that, for $f, g \in \mathcal{V}$,

$$0 \le \|\lambda f + g\|^2 = \lambda^2 \|f\|^2 + 2\lambda\langle f, g\rangle + \|g\|^2 \qquad \forall\, \lambda \in \mathbb{R}\,. \tag{9.3}$$

Denoting, for $f, g \in \mathcal{V}$ fixed, the quadratic polynomial in λ on the right-hand side by $A(\lambda)$, the condition for $A(\lambda)$ to be nonnegative for all λ in $\mathbb{R}$ is that $[2\langle f, g\rangle]^2 - 4\|f\|^2\|g\|^2 \le 0$; this gives the inequality (9.2). $\square$

Now, putting $\lambda = 1$ in (9.3) and using (9.2) on the right yields

$$\|f + g\| \le \|f\| + \|g\| \qquad \forall\, f,\, g \in \mathcal{V}\,.$$

https://doi.org/10.1017/CBO9780511801181.010 Published online by Cambridge University Press

Consequently, $\|\cdot\|$ obeys the triangle inequality, the third axiom of norm. The first two axioms of norm, namely that

- $\|f\| \geq 0$ for all $f \in \mathcal{V}$, and $\|f\| = 0$ if, and only if, $f = 0$ in $\mathcal{V}$, and
- $\|\lambda f\| = |\lambda|\,\|f\|$ for all $\lambda \in \mathbb{R}$ and all $f \in \mathcal{V}$,

follow directly form (9.1) and from the last three axioms of inner product stated in Definition 9.1.

We have thus shown the following result.

Theorem 9.1 *An inner product space $\mathcal{V}$ over the field $\mathbb{R}$ of real numbers, equipped with the induced norm $\|\cdot\|$, is a normed linear space over $\mathbb{R}$.*

We conclude this section with a relevant example of an inner product space, whose induced norm is the 2-norm considered at the beginning of Chapter 8.

Example 9.3 *The set $\mathrm{C}[a,b]$ of continuous real-valued functions defined on the closed interval $[a,b]$ is an inner product space with*

$$\langle f, g\rangle = \int_a^b w(x) f(x)\, g(x) \mathrm{d}x\,, \tag{9.4}$$

where w is a **weight function***, defined, positive, continuous and integrable on the open interval (a,b). The norm $\|\cdot\|_2$, induced by this inner product and given by*

$$\|f\|_2 = \left(\int_a^b w(x)|f(x)|^2 \mathrm{d}x\right)^{1/2}, \tag{9.5}$$

is referred to as the 2-norm on $\mathrm{C}[a,b]$ (see Example 8.2). For the sake of simplicity, we have chosen not to distinguish in terms of our notation between the 2-norm on $\mathrm{C}[a,b]$ defined above and the 2-norm for vectors introduced in Chapter 2; it will always be clear from the context which of the two is intended.

Clearly, it is not necessary to demand the continuity of the function f on the closed interval $[a,b]$ to ensure that $\|f\|_2$ is finite. For example, $f\colon x \mapsto \operatorname{sgn}\left(x - \frac{1}{2}(a+b)\right)$, $x \in [a,b]$, has finite 2-norm, despite the fact that it has a jump discontinuity at $x = \frac{1}{2}(a+b)$.

Motivated by this observation, and the desire to develop a theory of approximation in the 2-norm whose range of applicability extends beyond the linear space of continuous functions on a bounded closed interval, we denote by $\mathrm{L}^2_w(a,b)$ the set of all real-valued functions f

defined on (a, b) such that $w(x)|f(x)|^2$ is integrable[1] on (a, b); the set $\mathrm{L}^2_w(a, b)$ is equipped with the inner product (9.4) and the induced 2-norm (9.5). Obviously, $\mathrm{C}[a, b]$ is a proper subset of $\mathrm{L}^2_w(a, b)$.

In this broader context, $\|\cdot\|_2$ is frequently referred to as the L^2-norm; for the sake of simplicity we shall continue to call it the 2-norm. As before, w is assumed to be a real-valued function, defined, positive, continuous and integrable on the open interval (a, b). When $w(x) \equiv 1$ on (a, b), we shall write $\mathrm{L}^2(a, b)$ instead of $\mathrm{L}^2_w(a, b)$.

We are now ready to consider best approximation in the 2-norm.

9.3 Best approximation in the 2-norm

The problem of best approximation in the 2-norm can be formulated as follows:

(B) Given that $f \in \mathrm{L}^2_w(a, b)$, find $p_n \in \mathcal{P}_n$ such that

$$\|f - p_n\|_2 = \inf_{q \in \mathcal{P}_n} \|f - q\|_2 \,;$$

such p_n is called a **polynomial of best approximation of degree n to the function f in the 2-norm on (a, b)**.

The existence and uniqueness of p_n will be shown in Theorem 9.2. However, we shall first consider some simple examples.

Example 9.4 *Suppose that $\varepsilon > 0$ and let $f(x) = 1 - \mathrm{e}^{-x/\varepsilon}$ with x in $[0, 1]$. For $\varepsilon = 10^{-2}$, the function f is depicted in Figure 9.1. We shall construct the polynomial of best approximation of degree 0 in the 2-norm, with weight function $w(x) \equiv 1$, for f on $(0, 1)$, and compare it with the minimax polynomial of degree 0 for f on $[0, 1]$.*

The best approximation to f by a polynomial of degree 0 in the 2-norm on the interval $(0, 1)$, with weight function $w(x) \equiv 1$, is determined by minimising $\|f - c\|_2$ over all $c \in \mathbb{R}$; equivalently, we need to minimise

$$\int_0^1 (f(x) - c)^2 \mathrm{d}x = \int_0^1 |f(x)|^2 \mathrm{d}x - 2c \int_0^1 f(x) \mathrm{d}x + c^2$$

[1] Strictly speaking, the integral in the definition of $\|\cdot\|_2$ should now be thought of as a Lebesgue integral, with the convention that any two functions in $\mathrm{L}^2_w(a, b)$ which differ only on a set of zero measure are identified. Readers who are unfamiliar with the concept of Lebesgue integral can safely ignore this footnote. For the definition of *set of measure zero* see Section 11.1 in Chapter 11.

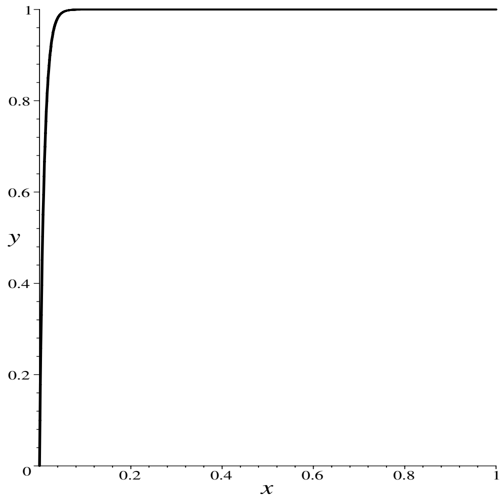

Fig. 9.1. Graph of the function $f\colon x \mapsto 1 - \mathrm{e}^{-x/\varepsilon}$ for $x \in [0,1]$ and $\varepsilon = 10^{-2}$.

over all $c \in \mathbb{R}$. The right-hand side is a quadratic polynomial in c; its minimum, as a function of c, is achieved for

$$c = \int_0^1 f(x)\,\mathrm{d}x = 1 - \varepsilon + \varepsilon\mathrm{e}^{-1/\varepsilon}\,.$$

Consequently, the polynomial of degree 0 of best approximation to f in the 2-norm on the interval $(0,1)$ with respect to the weight function $w(x) \equiv 1$ is

$$p_0^{(\text{2-norm})}(x) \equiv 1 - \varepsilon + \varepsilon\mathrm{e}^{-1/\varepsilon}\,, \qquad x \in [0,1]\,.$$

On the other hand, since $f \in \mathrm{C}[0,1]$ and f is strictly monotonic increasing on $[0,1]$, its minimax approximation of degree 0 on the interval $[0,1]$ is simply the arithmetic mean of $f(0)$ and $f(1)$:

$$p_0^{(\infty\text{-norm})}(x) \equiv \tfrac{1}{2}(1 - \mathrm{e}^{-1/\varepsilon})\,, \qquad x \in [0,1]\,.$$

Clearly, for $0 < \varepsilon \ll 1$, $p_0^{(\infty\text{-norm})}(x) \approx 1/2$, while $p_0^{(\text{2-norm})}(x) \approx 1$.

An even more dramatic discrepancy is observed between the polynomials of best approximation in the 2-norm and the ∞-norm when

$$f(x) = 1 - \varepsilon^{-1/2}\mathrm{e}^{-x/\varepsilon}\,, \qquad x \in [0,1]\,.$$

Here, for $0 < \varepsilon \ll 1$, $p_0^{(\text{2-norm})}(x) \approx 1$, as before. On the other hand,

$$p_0^{(\infty\text{-norm})}(x) \equiv 1 - \tfrac{1}{2}\varepsilon^{-1/2}(1 + \mathrm{e}^{-1/\varepsilon})\,, \qquad x \in [0,1]\,,$$

which tends to $-\infty$ as $\varepsilon \to 0+$. These examples indicate that the polynomial of best approximation from $\mathcal{P}_n$ to a function in the 2-norm can be vastly different from the minimax approximation from $\mathcal{P}_n$ to the same function. ◇

Given $f \in \mathrm{L}^2_w(a,b)$, we shall assume for the moment the existence of an associated polynomial of best approximation in the 2-norm; later on we shall prove that such a polynomial exists and is unique. In order to motivate the general discussion that will follow, it is helpful to begin with a straightforward approach to a simple example.

Let us suppose that we wish to construct the polynomial of best approximation $p_n \in \mathcal{P}_n$, $n \geq 0$, to a function $f \in \mathrm{L}^2_w(0,1)$ on the interval $(0,1)$ in the 2-norm; for simplicity, we shall assume that the weight function $w(x) \equiv 1$. Writing the polynomial p_n as

$$p_n(x) = c_0 + c_1 x + \cdots + c_n x^n \,,$$

we want to choose the coefficients c_j, $j = 0, \ldots, n$, so as to minimise the 2-norm of the error, $e_n = f - p_n$,

$$\|e_n\|_2 = \|f - p_n\|_2 = \left(\int_0^1 |f(x) - p_n(x)|^2 \mathrm{d}x \right)^{1/2} .$$

Since the 2-norm is nonnegative and the function $\xi \in \mathbb{R}_+ \mapsto \xi^{1/2}$ is monotonic increasing, this problem is equivalent to one of minimising the square of the norm; thus, instead, we shall minimise the expression

$$\begin{aligned} E(c_0, c_1, \ldots, c_n) &= \int_0^1 [f(x) - p_n(x)]^2 \mathrm{d}x \\ &= \int_0^1 [f(x)]^2 \mathrm{d}x - 2 \sum_{j=0}^n c_j \int_0^1 f(x) x^j \mathrm{d}x \\ &\quad + \sum_{j=0}^n \sum_{k=0}^n c_j c_k \int_0^1 x^{k+j} \mathrm{d}x \,, \end{aligned}$$

by treating it as a function of $(c_0, \ldots, c_n)$. At the minimum, the partial derivatives of E with respect to the c_j, $j = 0, \ldots, n$, are equal to zero. This leads to a system of $(n+1)$ linear equations for the coefficients $c_0, \ldots, c_n$:

$$\sum_{k=0}^n M_{jk} c_k = b_j \,, \qquad j = 0, \ldots, n \,, \tag{9.6}$$

https://doi.org/10.1017/CBO9780511801181.010 Published online by Cambridge University Press

where

$$M_{jk} = \int_0^1 x^{k+j}\mathrm{d}x = \frac{1}{k+j+1}\,,$$
$$b_j = \int_0^1 f(x)x^j\mathrm{d}x\,.$$

Equivalently, recalling that the inner product associated with the 2-norm (in the case of $w(x) \equiv 1$) is defined by

$$\langle g, h\rangle = \int_0^1 g(x)h(x)\mathrm{d}x\,,$$

M_{jk} and b_j can be written as

$$M_{jk} = \langle x^k, x^j\rangle\,, \qquad b_j = \langle f, x^j\rangle\,. \tag{9.7}$$

By solving the system of linear equations (9.6) for $c_0, \dots, c_n$, we obtain the coefficients of the polynomial of best approximation of degree n to the function f in the 2-norm on the interval $(0, 1)$. We can proceed in the same manner on any interval (a, b) with any positive, continuous and integrable weight function w defined on (a, b).

This approach is straightforward for small values of n, but soon becomes impractical as n increases. The source of the computational difficulties is the fact that the matrix M is the Hilbert matrix, discussed in Section 2.8. The Hilbert matrix is well known to be ill-conditioned for large n, so any solution to (9.6), computed with a fixed number of decimal digits, loses all accuracy due to accumulation of rounding errors. Fortunately, an alternative method is available, and is discussed in the next section.

9.4 Orthogonal polynomials

In the previous section we described a method for constructing the polynomial of best approximation $p_n \in \mathcal{P}_n$ to a function f in the 2-norm; it was based on seeking p_n as a linear combination of the polynomials x^j, $j = 0, \dots, n$, which form a basis for the linear space $\mathcal{P}_n$. The approach was not entirely satisfactory because it gave rise to a system of linear equations with a full matrix that was difficult to invert. The central idea of the alternative approach that will be described in this section is to expand p_n in terms of a different basis, chosen so that the resulting system of linear equations has a diagonal matrix; solving this linear system is then a trivial exercise. Of course, the nontrivial ingredient of this alternative approach is to find a suitable basis for $\mathcal{P}_n$ which achieves the objective that the matrix of the linear system is diagonal. The expression for M_{jk} in (9.7) gives us a clue how to proceed.

Suppose that φ_j, $j = 0, \ldots, n$, form a basis for $\mathcal{P}_n$, $n \geq 0$; let us seek the polynomial of best approximation as

$$p_n(x) = \gamma_0\varphi_0(x) + \cdots + \gamma_n\varphi_n(x)\,,$$

where $\gamma_0, \ldots, \gamma_n$ are real numbers to be determined. By the same process as in the previous section, we arrive at a system of linear equations of the form (9.6):

$$\sum_{k=0}^{n} M_{jk}\gamma_k = \beta_j\,, \qquad j = 0, \ldots, n\,,$$

where now

$$M_{jk} = \langle \varphi_k, \varphi_j \rangle \qquad \text{and} \qquad \beta_j = \langle f, \varphi_j \rangle\,,$$

with the inner product $\langle \cdot\,, \cdot \rangle$ defined by

$$\langle g, h \rangle = \int_a^b w(x)g(x)h(x)\,\mathrm{d}x\,,$$

and the weight function w assumed to be positive, continuous and integrable on the interval (a, b).

Thus, $M = (M_{jk})$ will be a diagonal matrix provided that the basis functions φ_j, $j = 0, \ldots, n$, for the linear space $\mathcal{P}_n$ are chosen so that $\langle \varphi_k, \varphi_j \rangle = 0$, for $j \neq k$; in other words, φ_k is required to be orthogonal to φ_j for $j \neq k$, in the sense of Definition 9.2. This observation motivates the following definition.

Definition 9.4 *Given a weight function w, defined, positive, continuous and integrable on the interval (a, b), we say that the sequence of polynomials φ_j, $j = 0, 1, \ldots$, is a* **system of orthogonal polynomials** *on the interval (a, b) with respect to w, if each φ_j is of exact degree j, and if*

$$\int_a^b w(x)\varphi_k(x)\varphi_j(x)\mathrm{d}x \quad \begin{cases} = 0 & \text{for all } k \neq j\,, \\ \neq 0 & \text{when } k = j\,. \end{cases}$$

Next, we show that a system of orthogonal polynomials exists on any interval (a, b) and for any weight function w which satisfies the conditions in Definition 9.4. We proceed inductively.

https://doi.org/10.1017/CBO9780511801181.010 Published online by Cambridge University Press

Let $\varphi_0(x) \equiv 1$, and suppose that φ_j has already been constructed for $j = 0, \ldots, n$, with $n \geq 0$. Then,

$$\int_a^b w(x)\varphi_k(x)\varphi_j(x)\mathrm{d}x = 0\,, \qquad k \in \{0, \ldots, n\} \setminus \{j\}\,.$$

Let us now define the polynomial

$$q(x) = x^{n+1} - a_0\varphi_0(x) - \cdots - a_n\varphi_n(x)\,,$$

where

$$a_j = \frac{\int_a^b w(x)\, x^{n+1}\varphi_j(x)\mathrm{d}x}{\int_a^b w(x)[\varphi_j(x)]^2\mathrm{d}x}\,, \qquad j = 0, \ldots, n\,.$$

It then follows that

$$\begin{aligned}\int_a^b w(x)q(x)\varphi_j(x)\mathrm{d}x &= \int_a^b w(x)x^{n+1}\varphi_j(x)\mathrm{d}x \\ &\quad -a_j \int_a^b w(x)[\varphi_j(x)]^2\mathrm{d}x \\ &= 0 \qquad \text{for } 0 \leq j \leq n\,,\end{aligned}$$

where we have used the orthogonality of the sequence φ_j, $j = 0, \ldots, n$. Thus, with this choice of the numbers a_j we have ensured that q is orthogonal to all the previous members of the sequence, and φ_{n+1} can now be defined as any nonzero-constant multiple of q. This procedure for constructing a system of orthogonal polynomials is usually referred to as **Gram–Schmidt orthogonalisation**.[1]

Example 9.5 *We shall construct a system of orthogonal polynomials* $\{\varphi_0, \varphi_1, \varphi_2\}$ *on the interval* $(0, 1)$ *with respect to the weight function* $w(x) \equiv 1$.

We put $\varphi_0(x) \equiv 1$, and we seek φ_1 in the form

$$\varphi_1(x) = x - c_0\varphi_0(x)$$

such that $\langle \varphi_1, \varphi_0 \rangle = 0$; that is,

$$\langle x, \varphi_0 \rangle - c_0 \langle \varphi_0, \varphi_0 \rangle = 0\,.$$

[1] Jørgen Pedersen Gram (27 June 1850, Nustrup, Denmark – 29 April 1916, Copenhagen, Denmark); Erhard Schmidt (13 January 1876, Dorpat, Russia (now Tartu, Estonia) – 6 December 1959, Berlin, Germany).

Hence,

$$c_0 = \frac{\langle x, \varphi_0 \rangle}{\langle \varphi_0, \varphi_0 \rangle} = \tfrac{1}{2}$$

and therefore,

$$\varphi_1(x) = x - \tfrac{1}{2}\varphi_0(x) = x - \tfrac{1}{2}\,.$$

By construction, $\langle \varphi_1, \varphi_0 \rangle = \langle \varphi_0, \varphi_1 \rangle = 0$.

We now seek φ_2 in the form

$$\varphi_2(x) = x^2 - (d_1\varphi_1(x) + d_0\varphi_0(x))$$

such that $\langle \varphi_2, \varphi_1 \rangle = 0$ and $\langle \varphi_2, \varphi_0 \rangle = 0$. Thus,

$$\begin{aligned} \langle x^2, \varphi_1 \rangle - d_1\langle \varphi_1, \varphi_1 \rangle - d_0\langle \varphi_0, \varphi_1 \rangle &= 0\,, \\ \langle x^2, \varphi_0 \rangle - d_1\langle \varphi_1, \varphi_0 \rangle - d_0\langle \varphi_0, \varphi_0 \rangle &= 0\,. \end{aligned}$$

As $\langle \varphi_0, \varphi_1 \rangle = 0$ and $\langle \varphi_1, \varphi_0 \rangle = 0$, we have that

$$\begin{aligned} d_1 &= \frac{\langle x^2, \varphi_1 \rangle}{\langle \varphi_1, \varphi_1 \rangle} = 1\,, \\ d_0 &= \frac{\langle x^2, \varphi_0 \rangle}{\langle \varphi_0, \varphi_0 \rangle} = \tfrac{1}{3}\,, \end{aligned}$$

and therefore

$$\varphi_2(x) = x^2 - x + \tfrac{1}{6}\,. \tag{9.8}$$

Clearly, $\langle \varphi_k, \varphi_j \rangle = 0$ for $j \neq k$, $j, k \in \{0, 1, 2\}$, and φ_j is of exact degree j, $j = 0, 1, 2$. Thus we have found the required system $\{\varphi_0, \varphi_1, \varphi_2\}$ of orthogonal polynomials on the interval $(0, 1)$ with respect to the given weight function w.

By continuing this procedure, we can construct a system of orthogonal polynomials $\{\varphi_0, \varphi_1, \ldots, \varphi_n\}$, with respect to the weight function $w(x) \equiv 1$ on the interval $(0, 1)$, for any $n \geq 1$. For example, when $n = 3$, we shall find $\{\varphi_0, \varphi_1, \varphi_2, \varphi_3\}$, with φ_0, φ_1, φ_2, as above, and

$$\varphi_3(x) = x^3 - \tfrac{3}{2}x^2 + \tfrac{3}{5}x - \tfrac{1}{20}\,.$$

◇

Having generated a system of orthogonal polynomials on the interval $(0, 1)$ with respect to the weight function $w(x) \equiv 1$, by performing the linear mapping $x \mapsto (b - a)x + a$ we may obtain a system of orthogonal polynomials on any open interval (a, b) with respect to the weight function $w(x) \equiv 1$. For example, when $(a, b) = (-1, 1)$, the mapping $x \mapsto 2x - 1$ leads to the system of Legendre polynomials on $(-1, 1)$.

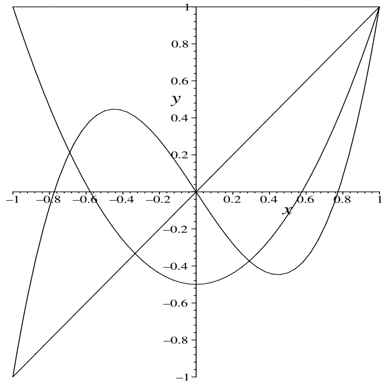

Fig. 9.2. The first four Legendre polynomials on the interval $(-1,1)$.

Example 9.6 (Legendre polynomials) *We wish to construct a system of orthogonal polynomials on* $(a,b) = (-1,1)$ *with respect to the weight function* $w(x) \equiv 1$.

On replacing x by

$$\frac{x-a}{b-a} = \tfrac{1}{2}(x+1)\,, \qquad x \in (a,b) = (-1,1)\,,$$

in $\varphi_0(x)$, $\varphi_1(x)$, $\varphi_2(x)$, $\varphi_3(x)$ from Example 9.5, we obtain, on normalising each of these polynomials so that its value at $x=1$ is equal to 1, the polynomials $\varphi_0, \varphi_1, \varphi_2, \varphi_3$, defined by

$$\begin{aligned}
\varphi_0(x) &= 1\,,\\
\varphi_1(x) &= x\,,\\
\varphi_2(x) &= \tfrac{3}{2}x^2 - \tfrac{1}{2}\,,\\
\varphi_3(x) &= \tfrac{5}{2}x^3 - \tfrac{3}{2}x\,.
\end{aligned}$$

These are the first four elements of the system of Legendre polynomials, orthogonal on the interval $(-1,1)$ with respect to the weight function $w(x) \equiv 1$. They are depicted in Figure 9.2. An alternative normalisation would have been to divide each φ_j by $\|\varphi_j\|_2$ so as to ensure that the 2-norm of the resulting scaled polynomial is equal to 1. ◇

Example 9.7 *The Chebyshev polynomials* T_n: $x \mapsto \cos(n\cos^{-1}x)$, $n = 0,1,\ldots$, *introduced in Section 8.4, form an orthogonal system on the interval* $(-1,1)$ *with respect to the positive, continuous and integrable weight function* $w(x) = (1-x^2)^{-1/2}$.

The proof of this is simple: let $\langle\,\cdot\,,\cdot\,\rangle$ denote the inner product in $\mathrm{L}^2_w(-1,1)$ with $w = (1-x^2)^{-1/2}$. By using the change of independent variable

$$t \in (0,\pi) \mapsto x = \cos t \in (-1,1)\,,$$

we have

$$\begin{aligned}\langle T_m, T_n\rangle &= \int_{-1}^{1} \frac{1}{\sqrt{1-x^2}}\,(\cos m\cos^{-1} x)\,(\cos n\cos^{-1} x)\,\mathrm{d}x \\ &= \int_0^\pi \cos mt\,\cos nt\,\mathrm{d}t \\ &= \tfrac{1}{2}\int_0^\pi \{\cos(m+n)t + \cos(m-n)t\}\,\mathrm{d}t \\ &= \begin{cases} 0 & \text{when } m \neq n\,, \\ \frac{\pi}{2} & \text{when } m = n\,, \end{cases}\end{aligned}$$

for any pair of nonnegative integers m and n. ◇

We are now ready to prove the existence and uniqueness of the polynomial of best approximation in the 2-norm. In particular, the next theorem shows that the infimum of $\|f-q\|_2$ over $q \in \mathcal{P}_n$ in problem (B) is attained and can be replaced by a minimum over $q \in \mathcal{P}_n$.

Theorem 9.2 *Given that $f \in \mathrm{L}^2_w(a,b)$, there exists a unique polynomial $p_n \in \mathcal{P}_n$ such that $\|f-p_n\|_2 = \min_{q\in\mathcal{P}_n}\|f-q\|_2$.*

Proof In order to simplify the notation, we recall the definition of the inner product $\langle\,\cdot\,,\cdot\,\rangle$:

$$\langle g, h\rangle = \int_a^b w(x)g(x)h(x)\mathrm{d}x\,,$$

and note that the induced 2-norm, $\|\cdot\|_2$, is defined by

$$\|g\|_2 = \langle g, g\rangle^{1/2}.$$

Suppose that φ_j, $j = 0,\ldots,n$, is a system of orthogonal polynomials with respect to the weight function w on (a,b). Let us normalise the polynomials φ_j by defining a new system of orthogonal polynomials,

$$\psi_j(x) = \frac{\varphi_j(x)}{\|\varphi_j\|_2}\,, \qquad j = 0,\ldots,n\,.$$

Then,

$$\langle \psi_k, \psi_j\rangle = \begin{cases} 1\,, & j = k\,, \\ 0\,, & j \neq k\,. \end{cases}$$

Such a system of polynomials is said to be **orthonormal**. The polynomials ψ_j, $j = 0, \ldots, n$, are linearly independent and form a basis for the linear space $\mathcal{P}_n$; therefore, each element $q \in \mathcal{P}_n$ can be expressed as a suitable linear combination,

$$q(x) = \beta_0 \psi_0(x) + \cdots + \beta_n \psi_n(x)\,.$$

We wish to choose β_j, $j = 0, \ldots, n$, so as to ensure that the corresponding polynomial q minimises $\|f - q\|_2^2$ over all $q \in \mathcal{P}_n$. Let us, therefore, consider the function E: $(\beta_0, \ldots, \beta_n) \in \mathbb{R}^{n+1} \mapsto E(\beta_0, \ldots, \beta_n)$ defined by $E(\beta_0, \ldots, \beta_n) = \|f - q\|_2^2$, where $q(x) = \beta_0\psi_0(x) + \cdots + \beta_n\psi_n(x)$. Then,

$$\begin{aligned}
E(\beta_0, \ldots, \beta_n) &= \langle f - q, f - q\rangle \\
&= \langle f, f\rangle - 2\langle f, q\rangle + \langle q, q\rangle \\
&= \|f\|_2^2 - 2\sum_{j=0}^{n} \beta_j \langle f, \psi_j\rangle + \sum_{j=0}^{n}\sum_{k=0}^{n} \beta_j\beta_k\langle \psi_k, \psi_j\rangle \\
&= \|f\|_2^2 - 2\sum_{j=0}^{n} \beta_j \langle f, \psi_j\rangle + \sum_{j=0}^{n} \beta_j^2 \\
&= \sum_{j=0}^{n} [\beta_j - \langle f, \psi_j\rangle]^2 + \|f\|_2^2 - \sum_{j=0}^{n} |\langle f, \psi_j\rangle|^2\,.
\end{aligned}$$

The function $(\beta_0, \ldots, \beta_n) \mapsto E(\beta_0, \ldots, \beta_n)$ achieves its minimum value at $(\beta_0^*, \ldots, \beta_n^*)$, where

$$\beta_j^* = \langle f, \psi_j\rangle\,, \qquad j = 0, \ldots, n\,.$$

Hence $p_n \in \mathcal{P}_n$ defined by

$$p_n(x) = \beta_0^*\psi_0(x) + \cdots + \beta_n^*\psi_n(x)$$

is the unique polynomial of best approximation of degree n to the function $f \in \mathrm{L}^2_w(a, b)$ in the 2-norm on the interval (a, b). $\square$

Remark 9.1 *As* $E(\beta_0^*, \ldots, \beta_n^*) = \|f - p_n\|_2^2 \geq 0$, *it follows from the proof of Theorem 9.2 that if* $f \in \mathrm{L}^2_w(a, b)$, *and* $\{\psi_0, \psi_1, \ldots\}$ *is an orthonormal system of polynomials in* $\mathrm{L}^2_w(a, b)$, *then*

$$\sum_{j=0}^{n} |\langle f, \psi_j\rangle|^2 \leq \|f\|_2^2$$

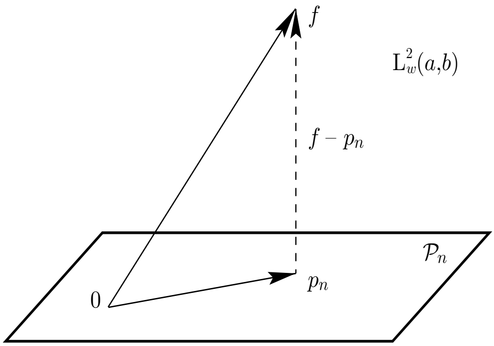

Fig. 9.3. Illustration of the orthogonality property $\langle f - p_n, q\rangle = 0$ for all q in $\mathcal{P}_n$, expressing the fact that if $p_n \in \mathcal{P}_n$ is a polynomial of best approximation to $f \in \mathrm{L}^2_w(a,b)$ in the 2-norm, then the error $f - p_n$ is orthogonal, in $\mathrm{L}^2_w(a,b)$, to all elements of the linear space $\mathcal{P}_n$. The 0 in the figure denotes the zero element of the linear space $\mathcal{P}_n$ (and, simultaneously, that of $\mathrm{L}^2_w(a,b)$), namely the function that is identically zero on the interval (a,b).

for each $n \geq 0$. This result is known as **Bessel's inequality**.[1]

The next theorem, in conjunction with the use of orthogonal polynomials, will be our key tool for constructing the polynomial of best approximation in the 2-norm.

Theorem 9.3 *A polynomial $p_n \in \mathcal{P}_n$ is the polynomial of best approximation of degree n to a function $f \in \mathrm{L}^2_w(a,b)$ in the 2-norm if, and only if, the difference $f - p_n$ is orthogonal to every element of $\mathcal{P}_n$, i.e.,*

$$\langle f - p_n, q\rangle = 0 \qquad \forall\, q \in \mathcal{P}_n\,. \tag{9.9}$$

A geometrical illustration of the property (9.9) is given in Figure 9.3.

Proof of theorem Suppose that (9.9) holds. Then,

$$\langle f - p_n, p_n - q\rangle = 0 \qquad \forall\, q \in \mathcal{P}_n,$$

given that $p_n - q \in \mathcal{P}_n$ for each q in $\mathcal{P}_n$. Therefore,

$$\begin{aligned}\|f - p_n\|_2^2 &= \langle f - p_n, f - p_n\rangle \\ &= \langle f - p_n, f - q\rangle + \langle f - p_n, q - p_n\rangle \\ &= \langle f - p_n, f - q\rangle \qquad\qquad \forall\, q \in \mathcal{P}_n\,.\end{aligned}$$

[1] Friedrich Wilhelm Bessel (22 July 1784, Minden, Westphalia, Holy Roman Empire (now Germany) – 17 March 1846, Königsberg, Prussia (now Kaliningrad, Russia)).

Hence, by the Cauchy–Schwarz inequality (9.2),

$$\|f - p_n\|_2^2 \le \|f - p_n\|_2 \, \|f - q\|_2 \qquad \forall\, q \in \mathcal{P}_n\,.$$

This implies that

$$\|f - p_n\|_2 \le \|f - q\|_2 \qquad \forall\, q \in \mathcal{P}_n\,.$$

On choosing $q = p_n$ on the right-hand side, equality will hold and therefore

$$\|f - p_n\|_2 = \min_{q \in \mathcal{P}_n} \|f - q\|_2\,.$$

Conversely, suppose that p_n is the polynomial of best approximation to $f \in \mathrm{L}^2_w(a,b)$. We have seen in the proof of Theorem 9.2 that p_n can be written in terms of the orthonormal polynomials ψ_k, $k = 0, \dots, n$, as

$$p_n(x) = \beta_0^* \psi_0(x) + \dots + \beta_n^* \psi_n(x)\,,$$

where

$$\beta_k^* = \langle f, \psi_k \rangle\,, \qquad k = 0, \dots, n\,. \tag{9.10}$$

On recalling that $\langle \psi_k, \psi_j \rangle = \delta_{jk}$, $j, k \in \{0, \dots, n\}$, where δ_{jk} is the Kronecker delta, we deduce from (9.10) that

$$\begin{aligned} \langle f - p_n, \psi_j \rangle &= \langle f, \psi_j \rangle - \sum_{k=0}^{n} \beta_k^* \langle \psi_k, \psi_j \rangle \\ &= \langle f, \psi_j \rangle - \sum_{k=0}^{n} \beta_k^* \delta_{jk} \\ &= \langle f, \psi_j \rangle - \beta_j^* = 0\,, \qquad j = 0, \dots, n\,. \end{aligned} \tag{9.11}$$

Since $\mathcal{P}_n = \mathrm{span}\{\psi_0, \dots, \psi_n\}$, it follows from (9.11) that $\langle f - p_n, q \rangle = 0$ for all $q \in \mathcal{P}_n$, as required. □

An equivalent, but slightly more explicit, form of writing (9.9) is

$$\int_a^b w(x)(f(x) - p_n(x))q(x)\,\mathrm{d}x = 0 \qquad \forall\, q \in \mathcal{P}_n\,.$$

Theorem 9.2 provides a simple method for determining the polynomial of best approximation $p_n \in \mathcal{P}_n$ to a function $f \in \mathrm{L}^2_w(a,b)$ in the 2-norm. First, proceeding as described in the discussion following Definition 9.4, we construct the system of orthogonal polynomials φ_j, $j = 0, \dots, n$, on the interval (a,b) with respect to the weight function w, if this system

https://doi.org/10.1017/CBO9780511801181.010 Published online by Cambridge University Press

is not already known. We normalise the polynomials φ_j, $j = 0, \dots, n$, by setting

$$\psi_j = \frac{\varphi_j}{\|\varphi_j\|_2}\,, \qquad j = 0, \dots, n\,,$$

to obtain the system of orthonormal polynomials ψ_j, $j = 0, \dots, n$, on (a, b). We then evaluate the coefficients $\beta_j^* = \langle f, \psi_j \rangle$, $j = 0, \dots, n$, and form $p_n(x) = \beta_0^* \psi_0(x) + \cdots + \beta_n^* \psi_n(x)$.

We may avoid the necessity of determining the normalised polynomials ψ_j by writing

$$\begin{aligned} p_n(x) &= \beta_0^* \psi_0(x) + \cdots + \beta_n^* \psi_n(x) \\ &= \beta_0^* \langle \varphi_0, \varphi_0 \rangle^{-1/2} \varphi_0(x) + \cdots + \beta_n^* \langle \varphi_n, \varphi_n \rangle^{-1/2} \varphi_n(x) \\ &= \gamma_0 \varphi_0(x) + \cdots + \gamma_n \varphi_n(x)\,, \end{aligned} \tag{9.12}$$

where

$$\gamma_j = \frac{\langle f, \varphi_j \rangle}{\langle \varphi_j, \varphi_j \rangle}\,, \qquad j = 0, \dots, n\,. \tag{9.13}$$

Thus, as indicated at the beginning of the section, with this approach to the construction of the polynomial of best approximation in the 2-norm, we obtain the coefficients γ_j explicitly and there is no need to solve a system of linear equations with a full matrix.

Example 9.8 *We shall construct the polynomial of best approximation of degree* 2 *in the* 2*-norm to the function* $f\colon x \mapsto \mathrm{e}^x$ *over* $(0, 1)$ *with weight function* $w(x) \equiv 1$.

We already know a system of orthogonal polynomials φ_0, φ_1, φ_2 on this interval from Example 9.5; thus, we seek $p_2 \in \mathcal{P}_2$ in the form

$$p_2(x) = \gamma_0 \varphi_0(x) + \gamma_1 \varphi_1(x) + \gamma_2 \varphi_2(x)\,, \tag{9.14}$$

where, according to (9.13),

$$\gamma_j = \frac{\int_0^1 \mathrm{e}^x \varphi_j(x) \mathrm{d}x}{\int_0^1 \varphi_j^2(x) \mathrm{d}x}\,, \qquad j = 0, 1, 2\,.$$

Recalling from Example 9.5 that

$$\varphi_0(x) \equiv 1\,, \quad \varphi_1(x) = x - \tfrac{1}{2}\,, \quad \varphi_2(x) = x^2 - x + \tfrac{1}{6}\,,$$

we then have that

https://doi.org/10.1017/CBO9780511801181.010 Published online by Cambridge University Press

$$\left.\begin{aligned}\gamma_0 &= \frac{\mathrm{e}-1}{1} = \mathrm{e}-1\,,\\ \gamma_1 &= \frac{3/2-\mathrm{e}/2}{1/12} = 18-6\mathrm{e}\,,\\ \gamma_2 &= \frac{7\mathrm{e}/6-19/6}{1/180} = 210\mathrm{e}-570\,.\end{aligned}\right\}\tag{9.15}$$

Substituting the values of γ_0, γ_1 and γ_2 into (9.14), we conclude that the polynomial of best approximation of degree 2 for the function $f\colon x \mapsto \mathrm{e}^x$ in the 2-norm is

$$p_2(x) = (210\mathrm{e}-570)x^2 + (588-216\mathrm{e})x + (39\mathrm{e}-105)\,.$$

The approximation error is

$$\|f - p_2\|_2 = 0.005431\,,$$

to six decimal digits. ◇

We conclude this section by giving a property of orthogonal polynomials that will be required in the next chapter.

Theorem 9.4 *Suppose that φ_j, $j = 0, 1, \ldots$, is a system of orthogonal polynomials on the interval (a, b) with respect to the positive, continuous and integrable weight function w on (a, b). It is understood that φ_j is a polynomial of exact degree j. Then, for $j \geq 1$, the zeros of the polynomial φ_j are real and distinct, and lie in the interval (a, b).*

Proof Suppose that ξ_i, $i = 1, \ldots, k$, are the points in the open interval (a, b) at which $\varphi_j(x)$ changes sign. Let us note that $k \geq 1$, because for $j \geq 1$, by orthogonality of $\varphi_j(x)$ to $\varphi_0(x) \equiv 1$, we have that

$$\int_a^b w(x)\varphi_j(x)\mathrm{d}x = 0\,.$$

Thus, the integrand, being a continuous function that is not identically zero on (a, b), must change sign on (a, b); however, w is positive on (a, b), so φ_j must change sign at least once on (a, b). Therefore $k \geq 1$.

Let us define

$$\pi_k(x) = (x - \xi_1)\ldots(x - \xi_k)\,.\tag{9.16}$$

Now the function $\varphi_j(x)\pi_k(x)$ does not change sign in the interval (a, b), since at each point where $\varphi_j(x)$ changes sign $\pi_k(x)$ changes sign also. Hence,

$$\int_a^b w(x)\varphi_j(x)\pi_k(x)\mathrm{d}x \neq 0\,.$$

However, φ_j is orthogonal to every polynomial of lower degree with respect to the weight function w, so the degree of the polynomial π_k must be at least j; thus, $k \geq j$. On the other hand, k cannot be greater than j, since a polynomial of exact degree j cannot change sign more than j times. Therefore $k = j$; *i.e.*, the points $\xi_i \in (a, b)$, $i = 1, \ldots, j$, are the zeros (and all the zeros) of $\varphi_j(x)$. □

9.5 Comparisons

We can show that the polynomial of best approximation in the 2-norm for a function $f \in \mathrm{C}[a, b]$ is also a near-best approximation in the ∞-norm for f on $[a, b]$ in the sense defined in Section 8.5.

Theorem 9.5 *Let $n \geq 0$ and assume that f is defined and continuous on the interval $[a, b]$, and $f \notin \mathcal{P}_n$. Let p_n be the polynomial of best approximation of degree n to f in the 2-norm on $[a, b]$, where the weight function w is positive, continuous and integrable on (a, b). Then, the difference $f - p_n$ changes sign at no less than $n + 1$ distinct points in the interval (a, b).*

Proof The proof is very similar to that of Theorem 9.4; we shall give an outline and leave the details as an exercise.

As $\langle f - p_n, 1 \rangle = 0$, *i.e.*,

$$\int_a^b w(x)(f(x) - p_n(x))\mathrm{d}x = 0\,,$$

and $w(x) > 0$ for all $x \in (a, b)$, it follows that $f - p_n$ changes sign in (a, b). Let ξ_j, $j = 1, \ldots, k$, denote distinct points in (a, b) where $f - p_n$ changes sign. We shall prove that $k \geq n + 1$.

Define the polynomial $\pi_k(x)$ as in (9.16); then, $w(x)[f(x)-p_n(x)]\pi_k(x)$ does not change sign in (a, b), and so its integral over (a, b) is not zero. Therefore, $\langle f - p_n, \pi_k \rangle \neq 0$. On the other hand, according to Theorem 9.3, $f - p_n$ is orthogonal to every polynomial of degree n or less. Hence the degree of $\pi_k(x)$ must be greater than n, and so $k \geq n + 1$. □

We return to the example illustrated by Figure 8.5, and consider the difference $f - p_n$ for the function $f\colon x \mapsto \mathrm{e}^{2x}$ on the interval $(0, 1)$. Figure 9.4 shows this difference for two polynomial approximations of degree 4: the minimax approximation of Section 8.5 and the best approximation in the 2-norm with weight function $w(x) \equiv 1$. It is clear that the

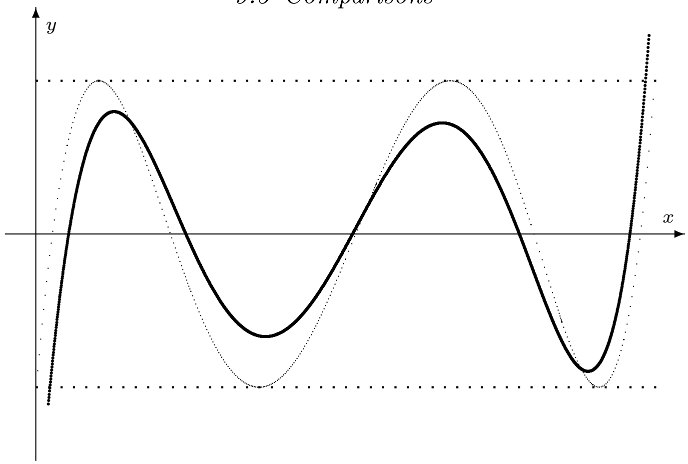

Fig. 9.4. The difference $e^{2x} - p_4(x)$ for two polynomial approximations of degree 4 on $[0, 1]$. Thin curve – minimax approximation; thick curve – best approximation in the 2-norm with weight function $w(x) \equiv 1$.

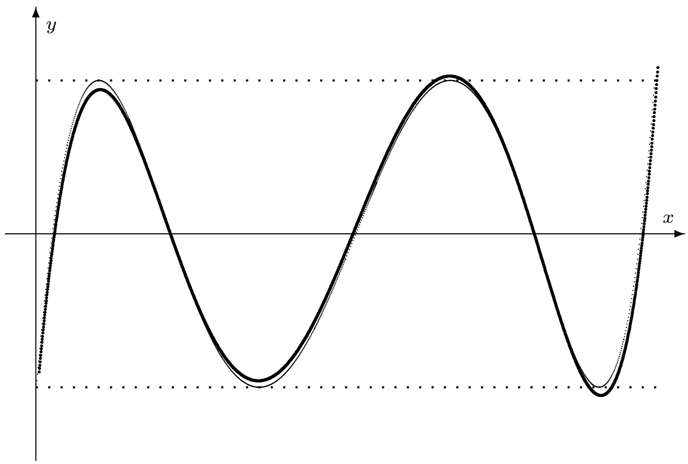

Fig. 9.5. The difference $e^{2x} - p_4(x)$ for two polynomial approximations of degree 4 on $[0, 1]$. Thin curve – minimax approximation; thick curve – best approximation in the 2-norm with weight function $w(x) = [x(1-x)]^{-1/2}$.

error of the 2-norm approximation has the right number of alternating local maxima and minima, and is a near-minimax approximation from $\mathcal{P}_4$ to f on $[0, 1]$; but the extrema at the ends of the interval are significantly larger than the internal extrema. If we use a weight function w

https://doi.org/10.1017/CBO9780511801181.010 Published online by Cambridge University Press

which gives greater weight near the ends of the interval, it seems likely that the extrema of the error might be more nearly equal. This can be achieved by using the weight function $w(x) = [x(1-x)]^{-1/2}$, so that the orthogonal polynomials are the Chebyshev polynomials adapted to the interval $(0, 1)$. Figure 9.5 shows the corresponding difference $f - p_n$, and we now see that the two best approximations, in the ∞-norm and the weighted 2-norm, are very close.

Polynomials of best approximation in the 2-norm have a special property which is often useful. Suppose that we have constructed the best polynomial approximation, p_n, of degree n, in the 2-norm, but that p_n does not achieve the required accuracy. To construct the best polynomial approximation of degree $n+1$ all we need is to calculate γ_{n+1} from

$$\gamma_{n+1} = \frac{\langle f - p_n, \varphi_{n+1} \rangle}{\|\varphi_{n+1}\|_2^2}$$

and then let $p_{n+1}(x) = p_n(x) + \gamma_{n+1}\varphi_{n+1}(x)$. By noting that

$$\langle f - p_{n+1}, \varphi_j \rangle = 0\,, \qquad j = 0, 1, \ldots, n+1\,,$$

it follows that p_{n+1} is best least squares approximation to f from $\mathcal{P}_{n+1}$. If we are constructing the minimax approximation of degree $n+1$, or using Lagrange interpolation with equally spaced points, the work involved in constructing p_n is lost, and the construction of p_{n+1} must begin completely afresh.

9.6 Notes

We give some pointers to the vast literature on orthogonal polynomials. The following are classical sources on the subject.

- Géza Freund, *Orthogonal Polynomials,* Pergamon Press, Oxford, New York, 1971.
- Paul Névai, *Orthogonal Polynomials,* Memoirs of the American Mathematical Society, no. 213, American Mathematical Society, Providence, RI, 1979.
- Gábor Szegő, *Orthogonal Polynomials*, Colloquium publications (American Mathematical Society), 23, American Mathematical Society, Providence, RI, 1959.

Tables of orthogonal polynomials are found in

➧ M. Abramowitz and I.A. Stegun (Editors), 'Orthogonal polynomials', Ch. 22 in *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, ninth printing, Dover, New York, pp. 771–802, 1972.

Computational aspects of the theory of orthogonal polynomials are discussed in the edited volume

➧ W. Gautschi, G.H. Golub, and G. Opfer (Editors), *Applications and Computation of Orthogonal Polynomials*, Conference at the Mathematical Research Institute, Oberwolfach, Germany, March 22–28, 1998, Birkhäuser, Basel, 1999.

A recent survey of the theory and application of orthogonal polynomials in numerical computations is contained in

➧ W. Gautschi, Orthogonal polynomials: applications and computation, *Acta Numerica* **5** (A. Iserles, ed.), Cambridge University Press, Cambridge, pp. 45–119, 1996.

Finally, we refer to the books of Powell and Cheney, cited in the Notes at the end of the previous chapter, concerning the application of orthogonal polynomials in the field of best least squares approximation.

Exercises

9.1 Construct orthogonal polynomials of degrees 0, 1 and 2 on the interval $(0, 1)$ with the weight function $w(x) = -\ln x$.

9.2 Let the polynomials φ_j, $j = 0, 1, \ldots$, form an orthogonal system on the interval $(-1, 1)$ with respect to the weight function $w(x) \equiv 1$. Show that the polynomials $\varphi_j((2x - a - b)/(b - a))$, $j = 0, 1, \ldots$, represent an orthogonal system for the interval (a, b) and the same weight function. Hence obtain the polynomials in Example 9.5 from the Legendre polynomials in Example 9.6.

9.3 Suppose that the polynomials φ_j, $j = 0, 1, \ldots$, form an orthogonal system on the interval $(0, 1)$ with respect to the weight function $w(x) = x^\alpha$, $\alpha > 0$. Find, in terms of φ_j, a system of orthogonal polynomials for the interval $(0, b)$ and the same weight function.

9.4 Show, by induction or otherwise, that, for $0 \le k \le n$,

$$\left(\frac{\mathrm{d}}{\mathrm{d}x}\right)^k (1-x^2)^n = (1-x^2)^{n-k} q_k(x)\,,$$

where q_k is a polynomial of degree k. Deduce that all the derivatives of the function $(1-x^2)^n$ of order less than n vanish at $x = \pm 1$.

Define $\varphi_j(x) = (\mathrm{d}/\mathrm{d}x)^j (1-x^2)^j$, and show by repeated integration by parts that

$$\int_{-1}^{1} \varphi_k(x)\varphi_j(x)\mathrm{d}x = 0\,, \qquad 0 \le k < j\,.$$

Hence verify the expressions in Example 9.6 for the Legendre polynomials of degrees 0, 1, 2 and 3.

9.5 Show, by induction or otherwise, that, for $0 \le k \le j$,

$$\left(\frac{\mathrm{d}}{\mathrm{d}x}\right)^k x^j \mathrm{e}^{-x} = x^{j-k} q_k(x)\mathrm{e}^{-x}\,,$$

where $q_k(x)$ is a polynomial of degree k.

The function φ_j is defined for $j \ge 0$ by

$$\varphi_j(x) = \mathrm{e}^x \frac{\mathrm{d}^j}{\mathrm{d}x^j}(x^j \mathrm{e}^{-x})\,.$$

Show that, for each $j \ge 0$, φ_j is a polynomial of degree j, and that these polynomials form an orthogonal system on the interval $(0, \infty)$ with respect to the weight function $w(x) = \mathrm{e}^{-x}$. Write down the polynomials with $j = 0, 1, 2$ and 3.

9.6 Suppose that φ_j, $j = 0, 1, \ldots$, form a system of orthogonal polynomials with weight function $w(x)$ on the interval (a, b). Show that, for some value of the constant C_j, $\varphi_{j+1}(x) - C_j x \varphi_j(x)$ is a polynomial of degree j, and hence that

$$\varphi_{j+1}(x) - C_j x\varphi_j(x) = \sum_{k=0}^{j} \alpha_{jk}\varphi_k(x)\,, \qquad \alpha_{jk} \in \mathbb{R}\,.$$

Use the orthogonality properties to show that $\alpha_{jk} = 0$ for $k < j - 1$, and deduce that the polynomials satisfy a recurrence relation of the form

$$\varphi_{j+1}(x) - (C_j x + D_j)\varphi_j(x) + E_j\varphi_{j-1}(x) = 0\,, \qquad j \ge 1\,.$$

https://doi.org/10.1017/CBO9780511801181.010 Published online by Cambridge University Press

9.7 In the notation of Exercise 6 suppose that the normalisation of the polynomials is so chosen that for each j the coefficient of x^j in $\varphi_j(x)$ is positive. Show that $C_j > 0$ for all j. By considering

$$\int_a^b w(x)\varphi_j(x)[\varphi_j(x) - C_{j-1}x\varphi_{j-1}(x)]\mathrm{d}x$$

show that

$$\int_a^b w(x)x\varphi_{j-1}(x)\varphi_j(x)\mathrm{d}x > 0\,,$$

and deduce that $E_j > 0$ for all j. Hence show that for all positive values of j the zeros of φ_j and φ_{j-1} interlace. (See the proof of Theorem 5.8.)

9.8 Using the weight function w on the interval (a, b) apply a similar argument to that for Theorem 8.6 to find the best polynomial approximation p_n of degree n in the 2-norm to the function x^{n+1}. Show that

$$\|x^{n+1} - p_n\|_2^2 = \int_a^b w(x)\varphi_{n+1}^2\mathrm{d}x/[c_{n+1}^{n+1}]^2\,,$$

where c_{n+1}^{n+1} is the coefficient of x^{n+1} in $\varphi_{n+1}(x)$.

Write down the best polynomial approximation of degree 2 to the function x^3 in the 2-norm with $w(x) \equiv 1$ on the interval $(-1, 1)$, and evaluate the 2-norm of the error.

9.9 Suppose that the weight w is an even function on the interval $(-a, a)$, and that a system of orthogonal polynomials φ_j, $j = 0, \ldots, n$, on the interval $(-a, a)$ is constructed by the Gram–Schmidt process. Show that, if j is even, then φ_j is an even function, and that, if j is odd, then φ_j is an odd function.

Now suppose that the best polynomial approximation of degree n in the 2-norm to the function f on the interval $(-a, a)$ is expressed in the form

$$p_n(x) = \gamma_0\varphi_0(x) + \cdots + \gamma_n\varphi_n(x)\,.$$

Show that if f is an even function, then all the odd coefficients γ_{2j-1} are zero, and that if f is an odd function, then all the even coefficients γ_{2j} are zero.

9.10 The function $H(x)$ is defined by $H(x) = 1$ if $x > 0$, and $H(-x) = -H(x)$. Construct the best polynomial approximations of degrees 0, 1 and 2 in the 2-norm to this function over the interval $(-1, 1)$ with weight function $w(x) \equiv 1$. (It may not appear very useful to consider a polynomial approximation to a discontinuous function, but representations of such functions by Fourier series will be familiar to most readers. Note that the function H belongs to $\mathrm{L}^2_w(-1,1)$.)