Appendix B

Polynomial Interpolation and Orthogonal Polynomials

B.1 The general interpolation problem

Given a set of discrete points x_i for $i = 0, 1, \ldots, n$ and function values F_i, the interpolation problem is to determine a function $\phi(x)$ of some specified form passing through these points,

$$\phi(x_i) = F_i \quad \text{for } i = 0, 1, \ldots, n. \tag{B.1}$$

We use the notation $\text{Int}(x_0, \ldots, x_n)$ to denote the smallest interval containing all these points (which need not be in increasing order but which are assumed to be distinct).

Interpolation has many uses; for example,

- we may have only discrete data values and want to estimate values in between, $x \in \text{Int}(x_0, \ldots, x_n)$. This is the origin of the term *interpolation*. We might also use this function to *extrapolate* if we evaluate it outside the interval where data are given.
- we may know the true function $F(x)$ but want to approximate it by a function $\phi(x)$ that is cheaper to evaluate, or easier to work with symbolically (to differentiate or integrate, for example).
- we may use it as a starting point for deriving numerical methods for differential equations (or for integral equations or numerical integration).

There are infinitely many possible functions ϕ. Typically ϕ is chosen to be a linear combination of some $n + 1$ given *basis functions* $\phi_0(x), \ldots, \phi_n(x)$,

$$\phi(x) = c_0\phi_0(x) + \cdots + c_n\phi_n(x). \tag{B.2}$$

Then condition (B.1) gives a linear system of $n + 1$ equations to solve for the coefficients $c_0, \ldots, c_n$,

$$\begin{bmatrix} \phi_0(x_0) & \phi_1(x_0) & \cdots & \phi_n(x_0) \\ \phi_0(x_1) & \phi_1(x_1) & \cdots & \phi_n(x_1) \\ \vdots & & & \vdots \\ \phi_0(x_n) & \phi_1(x_n) & \cdots & \phi_n(x_n) \end{bmatrix} \begin{bmatrix} c_0 \\ c_1 \\ \vdots \\ c_n \end{bmatrix} = \begin{bmatrix} F_0 \\ F_1 \\ \vdots \\ F_n \end{bmatrix}. \tag{B.3}$$

This system can we written as $\Phi c = F$. Different choices of basis functions lead to different types of interpolation. Using trigonometric functions gives Fourier series, for example, the basis of Fourier spectral methods.

In this appendix we consider interpolation by polynomials, the basis of many finite difference and spectral methods.

B.2 Polynomial interpolation

Through any $n + 1$ points there is a unique interpolating polynomial $p(x)$ of degree n. There are many ways to represent this function depending on what basis is chosen for $\mathcal{P}_n$, the set of all polynomials of degree n.

B.2.1 Monomial basis

The monomial functions are

$$\phi_0(x) = 1, \quad \phi_1(x) = x, \quad \phi_2(x) = x^2, \quad \ldots, \quad \phi_n(x) = x^n. \tag{B.4}$$

The matrix Φ appearing in (B.3) is then the *Vandermonde matrix*. This matrix may be quite ill-conditioned for larger values of n.

B.2.2 Lagrange basis

The jth Lagrange basis function (based on a given set of interpolation points x_i) is given by

$$\phi_j(x) = \prod_{\substack{i=0 \\ i \neq j}}^{n} \frac{(x - x_i)}{(x_j - x_i)}. \tag{B.5}$$

This is a polynomial of degree n. Note that

$$\phi_j(x_i) = \delta_{ij} = \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{if } i \neq j. \end{cases}$$

Then the matrix in (B.3) is the identity matrix and $c_i = F_i$. The coefficients are easy to determine in this form but the basis functions are a bit cumbersome.

B.2.3 Newton form

The Newton form of the interpolating polynomial is

$$p(x) = c_0 + c_1(x - x_0) + c_2(x - x_0)(x - x_1) + \cdots + c_n(x - x_0)(x - x_1)\cdots(x - x_{n-1}). \tag{B.6}$$

For these basis functions the matrix Φ is *lower triangular* and the c_i may be found by forward substitution. Alternatively they are most easily computed using *divided differences*, $c_i = F[x_0, \ldots, x_i]$. These can be computed from a tableau of the form

$$\begin{array}{llll} x_0 & F[x_0] & & \\ & & F[x_0,x_1] & \\ x_1 & F[x_1] & & F[x_0,x_1,x_2] \\ & & F[x_1,x_2] & \\ x_2 & F[x_2], & & \end{array} \tag{B.7}$$

where

$$F[x_j] = F_j$$

and for $k > 0$,

$$F[x_j,\dots,x_{j+k}] = \frac{F[x_{j+1},\dots,x_{j+k}] - F[x_j,\dots,x_{j+k-1}]}{x_{j+k} - x_j}. \tag{B.8}$$

Then the Newton form can be built up as follows:

$$\begin{aligned} p_0(x) &= F[x_0] \\ &\qquad \text{is the polynomial of degree 0 interpolating at } x_0, \\ p_1(x) &= F[x_0] + F[x_0,x_1](x-x_0) \\ &\qquad \text{is the polynomial of degree 1 interpolating at } x_0, x_1, \\ p_2(x) &= F[x_0] + F[x_0,x_1](x-x_0) + F[x_0,x_1,x_2](x-x_0)(x-x_1) \\ &\qquad \text{is the polynomial of degree 2 interpolating at } x_0, x_1, x_2, \end{aligned}$$

etc.

In each step we add a term that vanishes at all the preceding interpolation points and makes the function also interpolate at one new point. Note that the coefficients of previous basis functions do not change.

Relation to Taylor series. Note that

$$F[x_j, x_{j+1}] = \frac{F_{j+1} - F_j}{x_{j+1} - x_j}. \tag{B.9}$$

Suppose the data values F_i come from some underlying smooth function $F(x)$, so $F_i = F(x_i)$. Then (B.9) approximates the derivative $F'(x_j)$. Similarly, if $x_j,\dots,x_{j+k}$ are close together, then

$$F[x_j,\dots,x_{j+k}] \approx \frac{1}{k!}F^{(k)}(x_j), \tag{B.10}$$

where $F^{(k)}(x)$ is the kth derivative. In fact, one can show that for sufficiently smooth F,

$$F[x_j,\dots,x_{j+k}] = \frac{1}{k!}F^{(k)}(\xi) \tag{B.11}$$

for some ξ lying in the interval $\text{Int}(x_j,\dots,x_{j+k})$. This is true provided that F is k times continuously differentiable on this interval. The Newton form (B.6) thus is similar to the Taylor series

$$F(x) = F(x_0) + F'(x_0)(x-x_0) + \frac{1}{2!}F''(x_0)(x-x_0)^2 + \cdots \tag{B.12}$$

and reduces to this in the limit as $x_j \to x_0$ for all j.

B.2.4 Error in polynomial interpolation

Suppose $F(x)$ is a smooth function, we evaluate $F_i = F(x_i)$ ($i = 0, 1, \ldots, n$), and we now fit a polynomial $p(x)$ of degree n through these points. How well does $p(\bar{x})$ approximate $F(\bar{x})$ at some other point $\bar{x}$?

Note that we could add $\bar{x}$ as another interpolation point and create an interpolating polynomial $\bar{p}(x)$ of degree $n+1$ that interpolates also at this point,

$$\bar{p}(x) = p(x) + F[x_0, \ldots, x_n, \bar{x}](x - x_0)\cdots(x - x_n).$$

Then $\bar{p}(\bar{x}) = F(\bar{x})$ and so

$$F(\bar{x}) - p(\bar{x}) = F[x_0, \ldots, x_n, \bar{x}](\bar{x} - x_0)\cdots(\bar{x} - x_n).$$

Using (B.11), we obtain an error formula similar to the remainder formula for Taylor series. If $p(x)$ is given by (B.6), then at any point x,

$$F(x) - p(x) = \frac{1}{n!}F^{(n)}(\xi)(x - x_0)\cdots(x - x_n), \tag{B.13}$$

where ξ is some point lying in $\mathrm{Int}(x, x_0, \ldots, x_n)$. How large this is depends on

- how close the point x is to the interpolation points $x_0, \ldots, x_n$, and
- how small the derivative $F^{(n)}(\xi)$ is over this interval, i.e., how smooth the function is.

For a given x we don't know exactly what ξ is in general, but we can often use this expression to obtain an *error bound* of the form

$$|p(x) - F(x)| \le K|(x - x_0)\cdots(x - x_n)|, \tag{B.14}$$

where

$$K = \frac{1}{n!}\max_{\xi \in \mathrm{Int}(x_0, \ldots, x_n)} |F^{(n)}(\xi)|.$$

Note that the bound (B.14) involves values of the polynomial $Q(x) \equiv \prod_{i=1}^{n}(x - x_i)$, the polynomial with roots at the interpolation points x_i and with leading coefficient 1 (i.e., a monic polynomial). If we want to minimize the error over some interval, then we might want to choose the interpolation points to minimize the maximum value that $Q(x)$ takes over that interval. We will return to this in Section B.3.2, where we will see that Chebyshev polynomials satisfy the required optimality condition. These are a particular class of orthogonal polynomials, as described in the next section.

B.3 Orthogonal polynomials

If $w(x)$ is a function on an interval $[a, b]$ that is positive everywhere on the interval, then we can define the *inner product* of two functions $f(x)$ and $g(x)$ on this interval by

$$\langle f, g\rangle = \int_a^b w(x)f(x)g(x)\,dx. \tag{B.15}$$

We say that two functions $f(x)$ and $g(x)$ are *orthogonal* in a given interval $[a,b]$ with respect to the given weight function $w(x)$ if the inner product of f and g is equal to zero.

For a given $[a,b]$ and $w(x)$, one can define a sequence of orthogonal polynomials $P_0(x)$, $P_1(x)$, ... of increasing degree which have the property that

$$\begin{aligned} &P_m(x) \in \mathcal{P}_m \quad \text{(the set of polynomials of degree } m\text{)}, \\ &\langle P_m, P_n \rangle = 0 \quad \text{for } m \neq n. \end{aligned} \tag{B.16}$$

The sequence of polynomials is said to be *orthonormal* if, in addition,

$$\langle P_m, P_m \rangle = 1 \quad \text{for all } m. \tag{B.17}$$

Orthogonal polynomials have many interesting and useful properties and arise in numerous branches of numerical analysis. In particular, the Chebyshev polynomials described in Section B.3.2 are used in several contexts in this book.

The set of polynomials $P_0(x),\ P_1(x),\ \ldots,\ P_k(x)$ forms an *orthogonal basis* for $\mathcal{P}_k$. Any polynomial $p \in \mathcal{P}_k$ can be uniquely expressed as a linear combination of $P_0,\ \ldots,\ P_k$. Note that each P_m must have an exact degree of m, meaning the coefficient of x^m is nonzero. Otherwise it would be a linear combination of previous polynomials in the sequence and could not be orthogonal to them all.

Note that if P_m is orthogonal to $P_0,\ P_1,\ \ldots,\ P_{m-1}$, then in fact P_m is orthogonal to all polynomials $p \in \mathcal{P}_{m-1}$ of degree less than m, since p can be written as $p(x) = c_0P_0(x) + \cdots + c_{m-1}P_{m-1}(x)$ and so

$$\langle p, P_m \rangle = c_0\langle P_0, P_m \rangle + \cdots c_{m-1}\langle P_{m-1}, P_m \rangle = 0.$$

We say that P_m is orthogonal to the space $\mathcal{P}_{m-1}$.

Sequences of orthogonal polynomials can be built up by a Gram–Schmidt process, analogous to the manner in which a sequence of linearly independent vectors is transformed into a sequence of orthogonal vectors. Suppose $P_0,\ P_1,\ \ldots,\ P_m$ are already mutually orthogonal with P_n having exact degree n. We wish to construct $P_{m+1}(x)$, a polynomial of exact degree $m+1$ that is orthogonal to all of these. Start with the polynomial

$$Q(x) = \alpha_m x P_m(x)$$

for some $\alpha_m \neq 0$. This polynomial has exact degree $m+1$ and hence is linearly independent from $P_0,\ P_1,\ \ldots,\ P_m$. Moreover, it is already orthogonal to $P_0,\ P_1,\ \ldots,\ P_{m-2}$, since

$$\langle Q, P_n \rangle = \langle xP_m, P_n \rangle = \int_a^b w(x)xP_m(x)P_n(x)\,dx = \langle P_m, xP_n \rangle = 0$$

for $n \le m-2$, since $xP_n \in \mathcal{P}_{m-1}$ and P_m is orthogonal to this space. We wish to make Q orthogonal to P_{m-1} and P_m and, as in the Gram–Schmidt process for vectors, we do this by subtracting multiples of P_{m-1} and P_m from Q:

$$P_{m+1}(x) = \alpha_m x P_m(x) - \beta_m P_m(x) - \gamma_m P_{m-1}(x). \tag{B.18}$$

Requiring $\langle P_{m+1}, P_m\rangle = 0$ determines

$$\beta_m = \frac{\langle P_m, \alpha_m x P_m\rangle}{\langle P_m, P_m\rangle} \tag{B.19}$$

and then $\langle P_{m+1}, P_{m-1}\rangle = 0$ gives

$$\gamma_m = \frac{\langle P_{m-1}, \alpha_m x P_m\rangle - \beta_m\langle P_{m-1}, P_m\rangle}{\langle P_{m-1}, P_{m-1}\rangle}. \tag{B.20}$$

The relation (B.18) is a *three-term recurrence relation* for the sequence of orthogonal polynomials and can be used to generate the entire sequence once P_0 and P_1 are specified. For many useful sets of orthogonal polynomials the coefficients α_m, β_m, and γ_m take particularly simple forms.

B.3.1 Legendre polynomials

The sequence of polynomials that are orthogonal on $[-1, 1]$ with weight function $w(x) = 1$ are called the *Legendre polynomials*. We must also choose some normalization to uniquely define this sequence (since multiplying two orthogonal polynomials by arbitrary constants leaves them orthogonal). This amounts to choosing the nonzero constants α_m in (B.18). One might choose the polynomials to be orthonormal (i.e., normalize so that (B.17) is satisfied), but this leads to messy coefficients. The traditional choice is to require that $P_m(1) = 1$ for all m. The first few Legendre polynomials are then

$$\begin{aligned}
P_0(x) &= 1,\\
P_1(x) &= x,\\
P_2(x) &= \frac{3}{2}x^2 - \frac{1}{2},\\
P_3(x) &= \frac{5}{2}x^3 - \frac{3}{2}x.
\end{aligned} \tag{B.21}$$

These polynomials satisfy a three-term recurrence relation (B.18) with

$$\alpha_m = \frac{2m+1}{m+1}, \quad \beta_m = 0, \quad \gamma_m = \frac{m}{m+1}.$$

The roots of the Legendre polynomials are of importance in various applications. In particular, they are the nodes for *Gaussian quadrature* formulas for approximating the integral of a function; see, e.g., [16], [90]. There is no simple expression for the location of the roots, but they can be found as the eigenvalues of a tridiagonal matrix in MATLAB by the following code (adapted from the program `gauss.m` in [90], which also computes the weights for the associated Gauss quadrature rules):

```
Toff = .5./sqrt(1-(2*(1:m-1)).^(-2));
T = diag(Toff,1) + diag(Toff,-1);
xi = sort(eig(T));
```

These points are also sometimes used as grid points in spectral methods.

B.3.2 Chebyshev polynomials

Several topics discussed in this book involve the Chebyshev polynomials $T_m(x)$. These are a sequence of polynomials that are orthogonal on the interval $[-1, 1]$ with the weight function

$$w(x) = (1 - x^2)^{-1/2}. \tag{B.22}$$

The first few Chebyshev polynomials are

$$\begin{aligned} T_0(x) &= 1, \\ T_1(x) &= x, \\ T_2(x) &= 2x^2 - 1, \\ T_3(x) &= 4x^3 - 3x. \end{aligned} \tag{B.23}$$

Again these are normalized so that $T_m(1) = 1$ for $m = 0, 1, \ldots$. Chebyshev polynomials satisfy a particularly simple three-term recurrence,

$$T_{m+1}(x) = 2xT_m(x) - T_{m-1}(x). \tag{B.24}$$

While the weight function (B.22) may seem to be a less natural choice than $w(x) = 1$, the Chebyshev polynomials have a number of valuable properties, a few of which are listed below and are used elsewhere in this text.

Property 1. The Chebyshev polynomial $T_m(x)$ *equioscillates* $m + 1$ times in the interval $[-1, 1]$, i.e., $|T_m(x)|$ is maximized at $m + 1$ points $x_0, x_1, \ldots, x_m$ in the interval, points where $T_m(x)$ takes the values

$$T_m(x_j) = (-1)^j.$$

These *Chebyshev extreme points* are given by

$$x_j = \cos(j\pi/m), \quad j = 0, 1, \ldots, m. \tag{B.25}$$

(Note that these are labeled in decreasing order from $x_0 = 1$ to $x_m = -1$.) Figure B.1 shows a plot of $T_7(x)$, for example. This set of extreme points will be useful for spectral methods as discussed in Section 2.21.

Property 2. For x in the interval $[-1, 1]$, the value of $T_m(x)$ is given by

$$T_m(x) = \cos(m \arccos x). \tag{B.26}$$

This does not look much like a polynomial, but it is since $\cos(m\theta)$ can be written as a polynomial in $\cos(\theta)$ using trigonometric identities, and then set as $x = \cos(\theta)$. Note that from this formulation it is easy to check that (B.25) gives the desired extreme points.

Outside this interval there is an analogous formula in terms of the hyperbolic cosine,

$$T_m(x) = \cosh(m \cosh^{-1} x) \qquad \text{for } |x| \geq 1, \tag{B.27}$$

an expression that is used in the analysis of the convergence of conjugate gradients. Note that outside the unit interval the Chebyshev polynomials grow very rapidly.

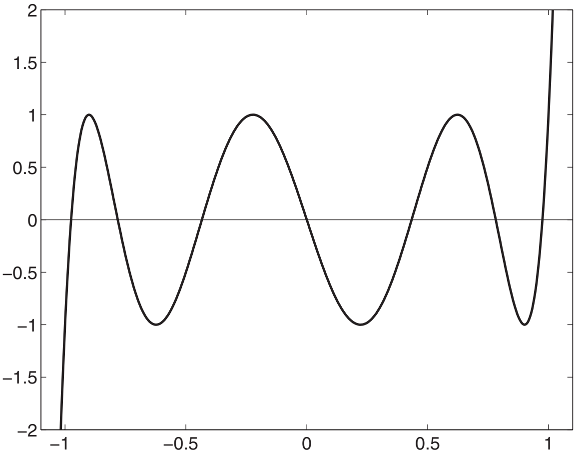

Figure B.1. *The Chebyshev polynomial $T_7(x)$ of degree 7.*

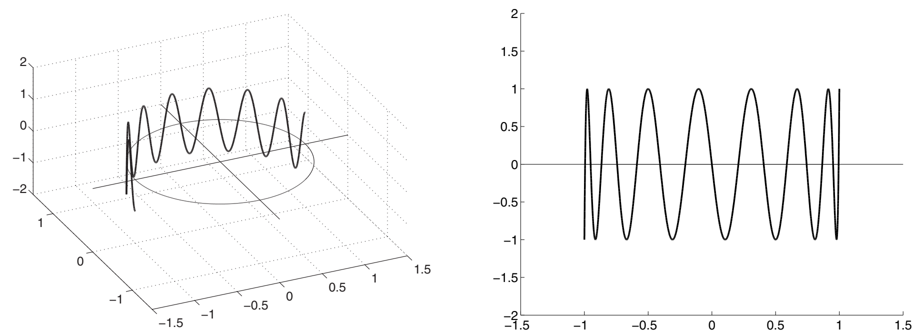

Figure B.2. *The Chebyshev polynomial viewed as a function $C_m(\theta)$ on the unit disk $e^{i\theta}$ and when projected on the x-axis, i.e., as a function of $x = \cos(\theta)$. Shown for $m = 15$.*

Property 3. Consider the function

$$C_m(\theta) = \mathrm{Re}(e^{im\theta}) = \cos(m\theta) \tag{B.28}$$

for $0 \le \theta \le \pi$. We can view this as a function defined on the upper half of the unit circle in the complex plane. If we identify $x = \cos(\theta)$ or $\theta = \arccos x$, then this reduces to (B.26), so we can view the Chebyshev polynomial on the interval $[-1, 1]$ as being the projection of the function (B.28) onto the real axis, as illustrated in Figure B.2. This property is useful in relating polynomial interpolation at Chebyshev points to trigonometric interpolation at equally spaced points on the unit circle and allows the use of the Fast Fourier Transform (FFT) algorithm to efficiently implement Chebyshev spectral methods. Orthogonality of the Chebyshev polynomials with respect to the weight function (B.22) also can be easily interpreted in terms of orthogonality of the trigonometric functions $\cos(m\theta)$ and $\cos(n\theta)$.

Property 4. The m roots of $T_m(x)$ all lie in $[-1, 1]$, at the points

$$\xi_j = \cos\left(\frac{(j-1/2)\pi}{m}\right) \qquad \text{for } j = 1,\ 2,\ \ldots,\ m. \tag{B.29}$$

This follows directly from the representation (B.26).

Property 5. The Chebyshev polynomial $T_m(x)$ solves the mini-max optimization problem

$$\text{find } p \in \mathcal{P}_m^1 \text{ to minimize } \max_{-1\le x\le 1} |p(x)|, \tag{B.30}$$

where $\mathcal{P}_m^1$ is the set of mth degree polynomials satisfying $p(1) = 1$. Recall that $T_m(x)$ equioscillates with extreme values ± 1 so $\max_{-1\le x\le 1} |T_m(x)| = 1$.

A slightly different formulation of this property is sometimes useful: the scaled Chebyshev polynomial $2^{1-m}T_m(x)$ is the monic polynomial of degree m that minimizes $\max_{-1\le x\le 1} |p(x)|$. A monic polynomial has leading coefficient 1 on the x^m term. The scaled Chebyshev polynomial $2^{1-m}T_m(x) = \prod_{j=1}^m (x - \xi_j)$ has leading coefficient 1 and equioscillates between the values $\pm 2^{1-m}$. If we try to reduce the level of any of these peaks by perturbing the polynomial slightly, at least one of the other peaks will increase in magnitude.

Note that 2^{1-m} decays to zero exponentially fast as we increase the degree. This is responsible for the spectral accuracy of Chebyshev spectral methods and this optimality is also used in proving the rapid convergence of the conjugate gradient algorithm (see Section 4.3.4).

Returning to the formula (B.14) for the error in polynomial interpolation, we see that if we are interested in approximating the function $F(x)$ uniformly well on the interval $[-1, 1]$, then we should use the Chebyshev roots (B.29) as interpolation points. Then (B.14) gives the bound

$$|p(x) - F(x)| \le K\, 2^{1-n}.$$

On a different interval $[a, b]$, we can use the shifted Chebyshev polynomial

$$T_n\left(\frac{2x - (a+b)}{(b-a)}\right). \tag{B.31}$$

The corresponding Chebyshev extreme points and Chebyshev roots are then

$$x_j = \frac{a+b}{2} + \frac{(b-a)}{2}\cos\left(\frac{j\pi}{m}\right) \qquad \text{for } j = 0,\ 1,\ \ldots,\ m \tag{B.32}$$

and

$$\xi_j = \frac{a+b}{2} + \frac{(b-a)}{2}\cos\left(\frac{(j-1/2)\pi}{m}\right) \qquad \text{for } j = 1,\ 2,\ \ldots,\ m, \tag{B.33}$$

respectively.