Chapter 4

Polynomial Approximations

Introduction

Polynomial computation is an important phase of numerical work. In calculus, a student learns that a polynomial is the simplest type of mathematical function. Thus it may be a surprise to him that polynomial computation presents difficulties and that a careful error analysis is necessary when the polynomial is of moderate or high degree.

In this chapter we derive the Newton and Lagrange polynomials as well as the respective truncation errors. We then address the topic of best approximation to a continuous function by a polynomial of degree n. The norm we choose for error analysis dominates construction of such a polynomial. Two norm options we explore are the L_∞ or Chebyshev norm and the ℓ_2 norm. With respect to the Chebyshev norm we first consider a special case where the approximation leads to the Chebyshev polynomials. We then explicity construct a polynomial of degree n that best approximates a function at $n+2$ select points. Given sufficient continuity conditions the result extends to the entire interval $[a,b]$. With respect to the ℓ_2 norm we seek the best least square approximation on a set of $m > n$ points. This final topic falls into the realm of optimization.

4.1 Properties of Polynomials

We denote an algebraic polynomial of degree n in x as

$$p_n(x) = a_0 + a_1 x + \cdots + a_n x^n \tag{4.1.1}$$

For the present discussion we assume that the a_i are real numbers. If $a_n \neq 0$ then the fundamental theorem of algebra guarantees the existence of n roots x_i where $i = 1, 2, \ldots, n$ such that

$$p_n(x) = a_n(x - x_1)(x - x_2) \cdots (x - x_n) . \tag{4.1.2}$$

We do not assume the roots are real nor do we require unicity since the equation above allows for multiplicities.

A powerful consequence of the fundamental theorem of algebra is that if two polynomials of degree n share the same ordinates at $n + 1$ distinct points then the polynomials are identical. To see this, let $p(x)$ and $q(x)$ be polynomials of degree n such that $p(x_i) = q(x_i)$ at x_i where $i = 0, 1, \ldots, n$. The polynomial

$$r(x) = p(x) - q(x)$$

is clearly of degree n and vanishes at $n+1$ distinct points . Using the roots x_i where $i = 1, \ldots, n$ and the fundamental theorem of algebra

$$r(x) = c(x - x_1) \cdots (x - x_n) .$$

But $r(x_0)$ also vanishes, hence, $c \equiv 0$.

Problems

4.1.1 Prove that if the coefficients of a polynomial of degree n are real then any complex roots that occur must do so in conjugate pairs. That is, for α and β real if $z = \alpha + i\beta$ is a root of the polynomial then so is $\alpha - i\beta$.

4.1.2 Show that if the roots of a polynomial of degree n are real and the coefficent of x^n is real then necessarily the coefficients of x^j where $j = n - 1, \ldots, 0$ are real.

4.2 Newton–Lagrange Interpolation

In this section we introduce interpolatory polynomials. A polynomial $p(x)$ interpolates a function $f(x)$ at $x_0, x_1, \ldots, x_n$ if $p(x_i) = f(x_i)$ for $i = 0, 1, \ldots, n$. Many recipes exist for building $p(x)$, two we examine in detail and others very briefly. We first look closely at the Lagrange polynomial and then develop the Newton polynomial from the Lagrange. Theoretically these represent the same polynomial but, just as with linear systems, alternative constructions lead to different strenths and weaknesses. The discussion below hopefully implies that the Newton polynomial is more practical for computing whereas the Lagrange polynomial is a powerful algebraic tool.

Lagrange recognized the connection between the polynomial that interpolates a function at $n+1$ given points x_i and the $n+1$ polynomials that vanish on the distinct subsets of n points from $\{x_i : \ i = 0, 1, \ldots, n\}$. For instance, polynomial of degree n that vanishes at $x_1, x_2, \ldots, x_n$ must have the form

$$\ell_0(x) = c(x - x_1)(x - x_2) \cdots (x - x_n)\,.$$

Requiring that $\ell_0(x_0) = 1$ determines c and gives the result

$$\ell_0(x) = \frac{(x - x_1)(x - x_2) \cdots (x - x_n)}{(x_0 - x_1)(x_0 - x_2) \cdots (x_0 - x_n)}\,.$$

The same argument produces the polynomial $\ell_i(x)$ that vanishes at $x = x_0, x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n$ and takes the value 1 at $x = x_i$; i.e.,

$$\ell_i(x) = \frac{(x - x_0) \cdots (x - x_{i-1})(x - x_{i+1}) \cdots (x - x_n)}{(x_i - x_0) \cdots (x_i - x_{i-1})(x_i - x_{i+1}) \cdots (x_i - x_n)}\,. \tag{4.2.1}$$

In shorthand notation, $\ell_i(x)$ satisfies

$$\ell_i(x_j) = \delta_{ij} \equiv \begin{cases} 1 & \text{if} \quad i = j \\ 0 & \text{if} \quad i \neq j\,. \end{cases} \tag{4.2.2}$$

where i and j run from 0 to n . We refer to the polynomials $\ell_i(x)$ as Lagrange basis functions. If we now multiply $\ell_i(x)$ by $f(x_i)$ and sum over i, we produce the Lagrange interpolating representation

$$L_n(x) = \sum_{i=0}^{n} f(x_i)\ell_i(x)\,. \tag{4.2.3}$$

It follows immediately from the construction argument that $L_n(x) = f(x)$ for $x = x_0, x_1, \ldots, x_n$. Furthermore, the fundamental theorem of algebra implies that $L_n(x)$ is unique. (Why?)

An alternative notation for Lagrange basis functions is useful when we later derive the error associated with Lagrange interpolation. We remark that this notation makes sense if we interpret the quotient x^2/x as x and say that we can extend the domain of a function to include those points where the Taylor series converges. Let $q_{n+1}(x)$ be

$$q_{n+1}(x) = (x - x_0)(x - x_1) \cdots (x - x_n) . \tag{4.2.4}$$

Then applying the product rule for derivatives and evaluating the result at x_i yields the denominator of $\ell_i(x)$

$$q'_{n+1}(x_i) = (x_i - x_0) \cdots (x_i - x_{i-1})(x_i - x_{i+1}) \cdots (x_i - x_n) . \tag{4.2.5}$$

This identity gives a new representation for the Lagrange basis functions, namely,

$$\ell_i(x) = \frac{q_{n+1}(x)}{(x - x_i) q'_{n+1}(x_i)} . \tag{4.2.6}$$

The Lagrange polynomial $L_n(x)$ now becomes

$$L_n(x) = \sum_{i=0}^{n} \frac{f(x_i) q_{n+1}(x)}{(x - x_i) q'_{n+1}(x_i)} . \tag{4.2.7}$$

Note that the Lagrange basis functions provide a convenient algebraic tool for constructing an interpolatory polynomial but they are impractical for calculation. For instance, evaluating (4.2.7) at a point other than an interpolating point is costly. It is possible to simplify $L_n(x)$ with a computer routine but to justify a particular simplification it must lead to fewer calculations that are also stable. Further, a viable simplification must allow for the inclusion of additional interpolatory points without starting from scratch. This directs us toward the Newton polynomial.

We have chosen to derive the Newton polynomial based on the Lagrange. This is historically unjustified but our method nicely shows the connection between the two representations. Observe that $q_{n+1}(x)/(x - x_i)$ is a polynomial of degree n for which the coefficient of x^n is 1. Thus the separate contributions from the Lagrange basis functions that determine the coefficient of x^n in (4.2.7) are easy to identify. If the symbol $f[x_0, x_1, \ldots, x_n]$ represents the coefficient of x^n in $L_n(x)$ then

$$f[x_0, x_1, \ldots, x_n] \equiv \sum_{i=0}^{n} \frac{f(x_i)}{q'_{n+1}(x_i)} . \tag{4.2.8}$$

The notation on the left emphasizes that the coefficient of x^n is a function of $x_0, x_1, \ldots, x_n$ but is not meant to convey any particular ordering of points.

In deriving the Lagrange polynomial we made no assumption concerning the order of the x_i. If we relabel the x_i in any order whatsoever, we derive the same polynomial $L_n(x)$; therefore, the value $f[x_0, x_1, \ldots, x_n]$ is independent of ordering. Our subsequent argument relies on this identity.

Suppose that $L_m(x)$ is the Lagrange polynomial of degree m in x, that interpolates $f(x)$ at the points $x_0, x_1, \ldots, x_m$ and that $m = 0, 1, \ldots, n$. By (4.2.8) the coefficient of x^m appearing in $L_m(x)$ is $f[x_0, x_1, \ldots, x_m]$. Now consider $L_m - L_{m-1}$. We make two observations regarding this difference. The first is that the coefficient of x^m must be the same as that appearing in L_m since $L_{m-1}(x)$ is a polynomial of degree at most $m-1$. The second is that $L_m - L_{m-1}$ vanishes at $x_0, \ldots, x_{m-1}$, hence, up to an arbitrary constant c

$$L_m(x) - L_{m-1}(x) = c(x - x_0)(x - x_1)\cdots(x - x_{m-1})\,. \tag{4.2.9}$$

Combined with our first remark we have

$$c = f[x_0, x_1, \ldots, x_m]\,.$$

If we systematically replace the differences in the identity

$$L_n(x) = L_0(x) + [L_1(x) - L_0(x)] + \cdots + [L_n(x) - L_{n-1}(x)]\,.$$

using (4.2.9) then we have the Newton polynomial

$$\begin{aligned} p_n(x) \quad = \quad & f(x_0) + f[x_0, x_1](x - x_0) + f[x_0, x_1, x_2](x - x_0)(x - x_1) \\ & + \cdots + f[x_0, x_1, \ldots, x_n](x - x_0)(x - x_1)\cdots(x - x_{n-1})\,. \end{aligned} \tag{4.2.10}$$

We denote the Newton polynomial by p_n to distinguish the represenation from L_n.

The coefficients $f[\cdots]$ admit convenient calculation. To see why, we first evaluate (4.2.10) at $x = x_1$. This simplifies to

$$p_n(x_1) = f(x_1) = f(x_0) + f[x_0, x_1](x_1 - x_0)$$

so that

$$f[x_0, x_1] = \frac{f(x_1) - f(x_0)}{x_1 - x_0}\,. \tag{4.2.11}$$

The expression in (4.2.11) is usually called the first divided difference of $f(x)$ at x_0 and x_1. Continuing, if $x = x_2$ in (4.2.10) then

$$\begin{aligned} p_n(x_2) = f(x_2) \quad = \quad & f(x_0) + f[x_0, x_1](x_2 - x_0) \\ & + f[x_0, x_1, x_2](x_2 - x_0)(x_2 - x_1) \end{aligned}$$

or

$$f[x_0, x_1, x_2](x_2 - x_0)(x_2 - x_1) = f(x_2) - f(x_0) - f[x_0, x_1](x_2 - x_0)\,.$$

This gives the second divided difference

$$f[x_0, x_1, x_2] = \frac{f[x_2, x_0] - f[x_0, x_1]}{x_2 - x_1}$$

but the divided difference is symmetric is x_0 and x_1 hence

$$f[x_0, x_1, x_2] = f[x_1, x_0, x_2] = \frac{f[x_2, x_1] - f[x_1, x_0]}{x_2 - x_0}\,.$$

By induction, we deduce the general relation for the m-th divided difference

$$f[x_0, x_1, \ldots, x_m] = \frac{f[x_1, x_2, \ldots, x_m] - f[x_0, x_1, \ldots, x_{m-1}]}{x_m - x_0} \tag{4.2.12}$$

where $f[x_i] \equiv f(x_i)$. In general the first divided difference for any two points $(x_i, f(x_i))$ and $(x_j, f(x_j))$ is given by

$$f[x_i, x_j] \equiv \frac{f(x_i) - f(x_j)}{x_i - x_j} \tag{4.2.13}$$

with higher order divided differences following as above. We can think of $f[x_0, x_1, \ldots, x_m]$ as a linear combination of $f(x_0), f(x_1), \ldots, f(x_m)$ where $x_0, x_1, \ldots, x_m$ determine the exact form of the combination.

Example 4.2.1 Let us find the Lagrange polynomial that passes through the points (0,0), (1,1), and (2,2). Here $q_3(x) = x(x-1)(x-2)$, and so

$$q_3'(0) = (-1)(-2) = 2;\quad q_3'(1) = 1(-1) = -1;\quad q_3'(2) = 2(1) = 2\,.$$

Hence

$$L_2(x) = \frac{0 \cdot q_3(x)}{x(2)} + \frac{1 \cdot q_3(x)}{(x-1)(-1)} + \frac{2 \cdot q_3(x)}{(x-2)(2)} = x\,.$$

Note that, while each polynomial $\ell_i(x)$ of the Lagrange polynomial $L_n(x)$ is of degree $n = 2$ the polynomial $L_n(x)$ may be of lower degree.

Example 4.2.2 The computation of the Newton polynomial is usually carried out according to the following array:

$$\begin{array}{cc|c|c}
x_i & y_i & & \\
0 & 1 & & \\
 & & f[x_0,x_1]=\frac{3-1}{1-0}=2 & \\
1 & 3 & & f[x_0,x_1,x_2]=\frac{7-2}{2-0}=\frac{5}{2} \\
 & & f[x_1,x_2]=\frac{10-3}{2-1}=7 & \\
2 & 10 & &
\end{array}$$

From this array and (4.2.10) we have $p_2(x) = 1 + 2x + \frac{5}{2}\,x(x-1)$.

The ordering of the points in the array is immaterial. For example,

$$\begin{array}{cc|c|c}
x_i & y_i & & \\
2 & 10 & & \\
 & & \frac{1-10}{0-2}=\frac{9}{2} & \\
0 & 1 & & \frac{2-9/2}{1-2}=\frac{5}{2} \\
 & & \frac{3-1}{1-0}=2 & \\
1 & 3 & &
\end{array}$$

Here we have $p_2(x) = 10 + \frac{9}{2}\,(x-2) + \frac{5}{2}\,(x-2)x$ but this is identical to the first expression.

Note that if the assumption $x_i \neq x_j$ whenever $i \neq j$, is not satisfied, then both the Lagrange method and the Newton method of obtaining an interpolation polynomial fail. We now turn our attention to a selection of topics that arise from general interpolation.

4.2.1 Divided Difference Tables

In the past, the practical use of interpolation polynomials was mainly confined to table–making , an activity of diminishing importance. For convenience tables are usually given at equispaced arguments of the interpolated function.

We begin with Newton polynomials where we assume equispaced points x_i. If h is the tabular interval then the grid points x_i expressed in terms of x_0 and h are

$$x_m = x_0 + mh\,.$$

We reformulate the Newton divided differences by introducing the forward differencing operator Δ. If $f_m \equiv f(x_m)$ then

$$\begin{aligned}
\Delta f_m &\equiv f_{m+1} - f_m \\
\Delta^2 f_m &\equiv \Delta\left(f_{m+1} - f_m\right) = f_{m+2} - 2f_{m+1} + f_m
\end{aligned}$$

and in general

$$\Delta^n f_m = \Delta(\Delta^{n-1} f_m \,.$$

With the definition above, the first and second Newton divided differences are

$$\begin{aligned} f[x_0, x_1] &= \frac{f(x_0) - f(x_1)}{x_0 - x_1} = \frac{f_1 - f_0}{h} = \frac{\Delta f_0}{h} \\ f[x_0, x_1, x_2] &= \frac{f[x_1, x_2] - f[x_0, x_1]}{x_2 - x_0} = \frac{f_2 - 2f_1 + f_0}{2h^2} = \frac{\Delta^2 f_0}{2!\, h^2} \end{aligned}$$

while induction implies

$$f[x_0, x_1, \ldots, x_m] = \frac{\Delta^m f_0}{m!\, h^m} \,.$$

When these results are substituted into (4.2.10) the Newton polynomial becomes

$$\begin{aligned} p_n(x) &= f_0 + \frac{(x - x_0)}{h} \Delta f_0 + \frac{(x - x_0)(x - x_1)}{2!\, h^2} \Delta^2 f_0 + \cdots \\ &\quad + \frac{(x - x_0)(x - x_1) \cdots (x - x_{n-1})}{n!\, h^n} \Delta^n f_0 \end{aligned}$$

or, in terms of the variable ξ where

$$x = x_0 + \xi h \quad x - x_m = (\xi - m)h$$

$$p_n(x) = \sum_{k=0}^{n} \binom{\xi}{k} \Delta^k f_0 \,. \tag{4.2.14}$$

This last identity points to a disadvantage of equispaced arguments, that is, some coefficients of the resulting polynomial are very large when n is large. For instance, the coefficient of $\Delta^m f_0$ is

$$\binom{\xi}{m} \equiv \frac{\xi(\xi - 1) \cdots (\xi - m + 1)}{m!} \,.$$

The greatest absolute value of this coefficient for ξ in the range $0 \le \xi \le n$ is not easy to obtain. For the case $m = 2n + 1$ we find a rough estimate on the magnitude of this coefficient by taking $\xi = n - \frac{1}{2} = 2m + \frac{1}{2}$ and then using Stirling's formula $\Gamma(n+a) \sim (2\pi)^{1/2} n^{n+a-1/2} e^{-n}$ as $n \to \infty$ (see e.g. Chapter 6) to derive

$$\begin{aligned}
\frac{\xi(\xi-1)\cdots(\xi-m+1)}{m!} &= \frac{\left(2m+\frac{1}{2}\right)\left(2m-\frac{1}{2}\right)\cdots\left(m+\frac{1}{2}\right)}{m!} \\
&= \frac{\Gamma\left(2m+\frac{3}{2}\right)}{\Gamma\left(m+\frac{3}{2}\right)\Gamma(m+1)} \\
&= \frac{2^{2m+1}}{\sqrt{\pi}}\,\frac{\Gamma\left(2m+\frac{3}{2}\right)}{\Gamma(2m+2)} \sim \frac{2^{2m+1}}{(2\pi m)^{1/2}}
\end{aligned}$$

as $m \to \infty$.

The Lagrange polynomial for equispaced abscissae is sometimes useful for mathematical tables. As with the Newton polynomials we expect some simplification in the representation of the Lagrange polynomial. If the interpolating points are

$$x_m = \pm mh\ , \quad m = 0, 1, \ldots, n$$

then the Lagrange basis functions are of degree $2n+1$ with the form

$$\ell_m(x) = \frac{c_m x \prod_{r=1}^{n} (x^2 - r^2h^2)}{x - mh}\ ,$$

where $c_m = 1/q_{2n+1}(x_m)$, or setting $x = \xi h$

$$\ell_m(x) = c_m h^n \, \frac{\xi \prod_{r=1}^{n} (\xi^2 - r^2)}{\xi - m}\ .$$

Representing $L_{2n+1}(x)$ consists of minor details.

Considering the close link between Lagrange and Newton polynomials we expect large terms in the Lagrange expansion for certain values of $x = \xi h$. To exhibit this we develop a relation between gamma functions and the Langrange basis functions. From Appendix 3, we see that the polynomial

$$\Gamma(\xi+n+1)/\Gamma(\xi-n)$$

vanishes at the $2n+1$ points $\xi = \pm m$, $m = 0, 1, 2, \ldots, n$; therefore , $(\xi - m)\ell_m(x)$ differs from $\Gamma(\xi+n+1)/\Gamma(\xi-n)$ by at most a multiplicative constant. If we denote this constant by α then we may write

$$\ell_m(x) = \alpha\,\frac{\Gamma(\xi+n+1)}{(\xi-m)\Gamma(\xi-n)}\ .$$

Two further identities that we need are

$$\Gamma(x)\Gamma(1-x) = \pi \sin(\pi x)$$

and

$$\sin \pi(\xi - n) = (-1)^{n-m} \sin \pi(\xi - m)\,.$$

Evaluating the first at $x = \xi - n$ and then using the second we now have

$$\ell_m(x) = \alpha(-1)^{n-m}\Gamma(\xi + n + 1)\Gamma(n + 1 - \xi)\,\frac{\sin \pi(\xi - m)}{\pi(\xi - m)}\,.$$

Finally we determine α by requiring that $\ell_m(mh) = 1$ so that

$$\ell_m(x) = \frac{\Gamma(n+1+\xi)\Gamma(n+1-\xi)}{\Gamma(n+1+m)\Gamma(n+1-m)}\,\frac{\sin \pi(\xi - m)}{\pi(\xi - m)}\,.$$

We examine the magnitude of $\ell_m(x)$ for the particular value $\xi = (2n - 1)/2$. In this case

$$|\ell_m((2n-1)h/2)| = \frac{\Gamma\left(2n + \frac{1}{2}\right)\Gamma\left(\frac{3}{2}\right)}{\Gamma(n+1+m)\Gamma(n+1-m)}\,\frac{1}{\pi\left(n - m - \frac{1}{2}\right)}\,.$$

Thus for example,

$$|\ell_0((2n-1)h/2)| = \frac{\Gamma\left(2n + \frac{1}{2}\right)\Gamma\left(\frac{3}{2}\right)}{\Gamma^2(n+1)} \cdot \frac{1}{\pi\left(n - \frac{1}{2}\right)} \sim \frac{2^{2n}}{\pi n}$$

as $n \to \infty$. Other convenient approximation methods exist that do not use equispaced arguments and are free from the disadvantage of large coefficients. Methods based on Chebyshev polynomials are one such instance.

4.2.2 Van der Monde Determinants

The Vandermonde determinant of the form

$$r_2(x) = \begin{vmatrix} 1 & x & x^2 \\ 1 & x_0 & x_0^2 \\ 1 & x_1 & x_1^2 \end{vmatrix}$$

arises frequently in polynomial formulations. Clearly, $r_2(x)$ is a polynomial of degree 2 in x that vanishes at $x = x_0$ and at $x = x_1$ since in either of these cases two rows of the determinant are identical. Hence

$$r_2(x) = c_2(x - x_0)(x - x_1)$$

To evaluate c_2 note that the coefficient of x^2 is

$$\begin{vmatrix} 1 & x_0 \\ 1 & x_1 \end{vmatrix}$$

(expansion in minors about row 1). So by uniqueness

$$r_2(x) = (x_1 - x_0)(x - x_0)(x - x_1)\,.$$

Similarly, it may be shown that

$$r_n(x) = \begin{vmatrix} 1 & x & x^2 & \dots & x^n \\ 1 & x_0 & x_0^2 & \dots & x_0^n \\ \vdots & & & & \\ 1 & x_{n-1} & x_{n-1}^2 & \dots & x_{n-1}^n \end{vmatrix} = c_n \prod_{i=0}^{n-1} (x - x_i) \tag{4.2.15}$$

with c_n given by

$$c_n = \begin{vmatrix} 1 & x_0 & x_0^2 & \dots & x_0^{n-1} \\ 1 & x_1 & x_1^2 & \dots & x_1^{n-1} \\ \vdots & & & & \\ 1 & x_{n-1} & x_{n-1}^2 & \dots & x_{n-1}^{n-1} \end{vmatrix} = \prod_{i>j} (x_i - x_j)\,. \tag{4.2.16}$$

The latter equality is obtained by the recursive process in Problem 4.2.6.

If $y(x)$ is the polynomial of degree 2 that passes through the points (x_i, y_i), $i = 0, 1, 2$, then $y(x) = y$ expressed as a functions of the determinant is

$$\begin{vmatrix} y & 1 & x & x^2 \\ y_0 & 1 & x_0 & x_0^2 \\ y_1 & 1 & x_1 & x_1^2 \\ y_2 & 1 & x_2 & x_2^2 \end{vmatrix}\,.$$

The Lagrange polynomial $y(x)$ can be recovered from this equation by expanding the determinant along the first column of the matrix. For example, the coefficient of y_0 in the expansion of the determinant, i.e., $\ell_0(x)$ is given by

$$\ell_0(x) = \begin{vmatrix} 1 & x & x^2 \\ 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \end{vmatrix} \div \begin{vmatrix} 1 & x_0 & x_0^2 \\ 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \end{vmatrix} = \frac{(x-x_1)(x-x_2)}{(x_0-x_1)(x_0-x_2)} .$$

The condition that $y(x)$ be of degree 2 (and no less) requires that the coefficient of x^2 in the determinant be non–zero, i.e.

$$\begin{vmatrix} y_0 & 1 & x_0 \\ y_1 & 1 & x_1 \\ y_2 & 1 & x_2 \end{vmatrix} \neq 0 .$$

The geometric interpretation of this result is that the triangle formed by the points (x_0, y_0), (x_1, y_1), (x_2, y_2) does not collapse into a straight line. These results and their geometric interpretations extend easily to the case of a polynomial of degree n.

4.2.3 Hermite Interpolation

A certain type of confluence leads to a Lagrange–type polynomial of degree $2n+1$ that interpolates both f and f' at $x = x_0, x_1, \ldots, x_n$. If we set

$$H_{2n+1}(x) = \sum_{i=0}^{n} [\alpha_i + \beta_i(x - x_i)]\ell_i^2(x)$$

where $\ell_i(x)$ is defined as in (4.2.11), and demand that $H_{2n+1}(x_i) = f(x_i)$, $H'_{2n+1}(x_i) = f'(x_i)$ for $i = 0, 1, \ldots, n$, then we find that

$$\begin{aligned} H_{2n+1}(x) &= \sum_{i=0}^{n} \{ f(x_i)[1 - 2(x - x_i)\ell_i'(x_i)]\ell_i^2(x) \\ &\quad + f'(x_i)(x - x_i)\ell_i^2(x) \} \end{aligned} \qquad (4.2.17)$$

This is called Hermite's interpolation polynomial and should not be confused with the orthogonal Hermite polynomial that is important in the theory of Gaussian quadrature. Clearly, the above process can be extended. For example we can find the constants a_1, b_1, and c_1 such that the polynomial

$$p_{3n+2}(x) = \sum_{i=0}^{n} [a_i + b_i(x - x_i) + c_i(x - x_i)^2]\ell_i^3(x)$$

satisfies

$$p_{3n+2}^{(k)}(x_i) = f^{(k)}(x_i), \; i = 0, 1, \ldots, n; \quad k = 0, 1, 2\,.$$

Polynomials that interpolate function values plus higher derivatives are generically called osculating polynomials.

4.2.4 Aitken's Variant of the Newton Polynomial

Let us suppose that $n+1$ points $(x_i, f(x_i))$ are given, $i = 0, 1, \ldots, n$, and that we want to evaluate the polynomial $p(\xi)$ of degree n in ξ, at $\xi = x$. We examine

$$P_{0,j}(x) \equiv \frac{\begin{vmatrix} f_0 & x_0 - x \\ f_j & x_j - x \end{vmatrix}}{x_j - x_0}\,, \quad j \neq 0 \tag{4.2.18}$$

and observe that $P_{0,j}(x)$ is the linear polynomial that interpolates f_0 and f_j at x_0 and x_j, respectively. Consider now the quadratic polynomials

$$P_{0,1,j}(x) \equiv \frac{\begin{vmatrix} P_{0,1}(x) & x_1 - x \\ P_{0,j}(x) & x_j - x \end{vmatrix}}{x_j - x_1}\,, \quad j \neq 0, 1$$

that have the property that $P_{0,1,j}(x)$ interpolates f_0, f_1 and f_j at x_0, x_1 and x_j. Now assume that $P_{0,1,2,\ldots,k-1,j}(x)$ has been defined (inductively) to be the polynomial of degree k that interpolates the data $\{f_0, f_1, \ldots, f_{k-1}, f_j\}$ at $\{x_0, x_1, \ldots, x_{k-1}, x_j\}$, $j = k, \ldots, n$, and define

$$P_{0,1,2,\ldots,k-1,k,j}(x) = \frac{\begin{vmatrix} P_{0,1,\ldots,k-1,k}(x) & x_k - x \\ P_{0,1,\ldots,k-1,j}(x) & x_j - x \end{vmatrix}}{x_j - x_k}\,, \quad j \geq k+1\,. \tag{4.2.19}$$

This is the polynomial of degree $k+1$ that interpolates $\{f_0, \ldots, f_k, f_j\}$ at $\{x_0, x_1, \ldots, x_k, x_j\}$, $j = k+1, \ldots, n$. This development is conveniently displayed in tabular form as follows $(k+1 = n = 4)$:

If the columns are enumerated as in the table then the column heading is the degree of the interpolant. The index of $P\ldots$ are the indices of the nodes at which $P\ldots$ interpolates the data. For each fixed x note that an entry in the table is a linear interpolant of two entries in the column to the immediate left, e.g. $P_{0,1,4}$ is the linear interpolant of $P_{0,1}$ and $P_{0,4}$ at x_1 and

Table 4.2.1: Aitken's Difference Table

Column		1	2	3	4
x_0	f_0				
x_1	f_1	$P_{0,1}$			
x_2	f_2	$P_{0,2}$	$P_{0,1,2}$		
x_3	f_3	$P_{0,3}$	$P_{0,1,3}$	$P_{0,1,2,3}$	
x_4	f_4	$P_{0,4}$	$P_{0,1,4}$	$P_{0,1,2,4}$	$P_{0,1,2,3,4}$

x_4 and $P_{0,1,2,3}$ is the linear interpolant of $P_{0,1,2}$ and $P_{0,1,3}$ at x_2 and x_3 (see (4.2.19)).

Problems

4.2.1 If $\ell_m(x)$ is a Lagrange basis function where we are interpolating at the points $x_0, x_1, \ldots, x_n$ show that

$$\sum_{m=0}^{n} \ell_m(x) = 1$$

and

$$\sum_{m=0}^{n} (x_m - x)^j \ell_m(x) = 0 \ , \quad j = 1, 2, \ldots, n \, .$$

4.2.2 Show that

$$f(n) = \sum_{m=0}^{n} \binom{n}{m} \Delta^m f(0) \, .$$

4.2.3 Let $f \in C[a,b]$ and let $x_0, \ldots, x_n$ be points in $[a,b]$. If we define L_n as the operator such that $L_n f$ is a polynomial of degree $\leq n$ which agrees with f at $x_0, \ldots, x_n$ then show that L_n is a linear operator, i.e., $L_n(\alpha f) = \alpha L_n f$ and $L_n(f+g) = L_n f + L_n g$.

4.2.4 Show that $L_n f = f$ if and only if f is a polynomial of degree $\leq n$.

4.2.5 Show that $||L_n f|| \leq \alpha ||f||$ where $\alpha = \max_x \sum |\ell_i(x)|$ and that no smaller α has this property for all $f \in C[a,b]$.

4.2.6 By induction show that

$$c_n = \begin{vmatrix} 1 & x_0 & x_0^2 & \dots & x_0^{n-1} \\ 1 & x_1 & x_1^2 & \dots & x_1^{n-1} \\ \vdots & & & & \\ 1 & x_{n-1} & x_{n-1}^2 & \dots & x_{n-1}^{n-1} \end{vmatrix} = \prod_{i>j} (x_i - x_j) \, .$$

Reduce c_n to a determinant of order $n-1$ as follows: For columns $i = 1, 2, \dots, n-1$ multiply column i by x_0 and subtract the result from column $i+1$. Then factor and use the induction hypothesis.

4.2.7 Prove that if the $n+1$ numbers x_i, $i = 0, 1, \dots, n$ are distinct, then the system of equations

$$\begin{bmatrix} 1 & x_0 & x_0^2 & \dots & x_0^n \\ 1 & x_1 & x_1^2 & \dots & x_1^n \\ \dots & & & & \\ 1 & x_n & x_n^2 & \dots & x^n \end{bmatrix} \begin{bmatrix} b_0 \\ b_1 \\ \vdots \\ b \end{bmatrix} = \begin{bmatrix} y_0 \\ y_1 \\ \vdots \\ y \end{bmatrix}$$

always has a solution for any given set of numbers $y_0, y_1, \dots, y_n$. How is this solution
related to the polynomial of degree n that interpolates the data
$(x_0, y_0), (x_1, y_1), \dots, (x_n, y_n)$? (We do not recommend this as a practical method of constructing an interpolation polynomial.)

4.2.8 Find a_i, b_i, and c_i such that the polynomial $p_{3n+2}(x)$ in (2.20) satisfies $p_{3n+2}(x_i) = f(x_i)$, $p'_{3n+2}(x_i) = f'(x_i)$, $p''_{3n+2}(x_i) = f''(x_i)$, for $i = 0, 1, \dots, n$.

4.2.9 Let $p(x)$ be a polynomial of degree 4 in x, such that $p(0) = 5$, $p(3.5) = 2$, $p(1.7) = 8$, $p(9.6) = 4$, $p(11.2) = -6$. Use the Aitken algorithm to evaluate $p(2.8)$.

4.2.10 Neville's algorithm for evaluating the polynomial $p(x)$ that interpolates $f(x)$ at $x = x_0, x_2, \dots, x_n$ is described as follows:

$$\begin{aligned} p_k^{(0)}(x) &= f(x_k) \; , \quad k = 0, 1, \dots, n \\ p_k^{(s)}(x) &= \frac{1}{x_k - x_{k-s}} \begin{vmatrix} p_k^{(s-1)}(x) & p_{k-1}^{(s-1)}(x) \\ x_{k-s} - x & x_k - x \end{vmatrix} \end{aligned}$$

$k = s, s+1, \dots, n$; $s = 1, 2, \dots, n$. Show that $p_n^{(n)}(x) = p(x)$.

4.2.11 Thiele's algorithm for constructing a continued fraction representation of a rational function $r(x)$ that interpolates $f(x)$ at $x_0, x_1, \ldots, x_n$ takes the form given in the figure below, where the circled term is dropped for purposes of interpolation. The reciprocal divided differences $p(\ldots)$ are defined as follows:

$$\begin{aligned}
p(x_0x_1) &= \frac{x_0 - x_1}{f_0 - f_1} \\
p(x_0x_1x_2) &= \frac{x_0 - x_2}{p(x_0x_1) - p(x_1x_2)} + f_1 \\
p(x_0x_1x_2x_3) &= \frac{x_0 - x_3}{p(x_0x_1x_2) - p(x_1x_2x_3)} + p(x_1x_2) \\
&\cdots \\
p(x_0x_1\ldots x_n) &= \frac{x_0 - x_n}{p(x_0x_1\ldots x_{n-1}) - p(x_1x_2\ldots x_n)} + p(x_1x_2\ldots x_{n-1})
\end{aligned}$$

where $f_i = f(x_i)$.

(a) Show that if the circled term in the figure below is retained, then $r(x_i) = f(x_i)$ where $i = 0, 1, \ldots, n$.

(b) Show that if the circled term in the figure is dropped, then $r(x) = \frac{p(x)}{q(x)}$, where p and q are polynomials. Furthermore, if $x, x_0, \ldots, x_n \in [a, b]$,

$$f(x) - r(x) = \frac{(x - x_0)(x - x_1)\cdots(x - x_n)}{(n+1)!\,\{q(x)\}^2} \left(\frac{d}{dy}\right)^{n+1} \left[f(y)\{q(y)\}^2\right]_{y=\xi}$$

for some $\xi \in (a, b)$. Notice that

$$f(y) = q^2(y)\left[f(x) - \frac{p(y)}{q(y)}\right] - q^2(x)\left[f(x) - \frac{p(x)}{q(x)}\right]\frac{q_{n+1}(y)}{q_{n+1}(x)}$$

where

$$q_{n+1}(y) = \prod_{i=0}^{n}(y - x_i)$$

has $(n+2)$ zeros and

$$\frac{d^{n+1}}{dy^{n+1}}(qp) = 0\,.$$

where the last two terms of the continued fraction take the form

4.3 The Truncation Error of Interpolating Polynomials

In keeping with the purpose of this chapter, the previous discussion was mainly algebraic. The discussion and the illustrative examples were limited to an understanding of interpolating polynomials, and some extensions of these. The following discussion of the truncation error of polynomial approximation is similarly limited.

Theorem 4.3.1 (Rolle's Theorem): *If*

(i) $f(x)$ *is continuous in* $[a, b]$,

(ii) $f(a) = f(b)$, *and*

(iii) $f'(x)$ *exists in* (a, b),

then there is a point ζ *in* (a, b) *such that* $f'(\zeta) = 0$.

We omit the proof of Rolle's theorem, since it is included in most calculus texts and instead exhibit applications. Our first is the Mean Value Theorem.

Theorem 4.3.2 (Mean Value Theorem): *If* $f(x)$ *is continuous in* $[a, b]$ *and differentiable in* (a, b) *then there exists a point* $\zeta \in (a, b)$ *such that*

$$f'(\zeta) = \frac{f(b) - f(a)}{b - a}.$$

Proof: The linear polynomial that interpolates f at $x = a$ and $x = b$ is

$$p_1(x) = f(a) + f[a, b](x - a).$$

Let

$$\epsilon_1(x) \equiv f(x) - p_1(x),$$

then $\epsilon_1(x)$ satisfies the hypotheses of Rolle's Theorem; hence, there exists $\zeta \in (a, b)$ such that

$$0 = \epsilon_1'(\zeta) = f'(\zeta) - f[a, b].$$

This immediately gives

$$f'(\zeta) = \frac{f(b) - f(a)}{b - a}.$$

One way to interpret the Mean Value Theorem is that it guarantees the existence of at least one point $\zeta \in (a, b)$ such that the tangent line to $f(\zeta)$ is parallel to the secant line between $(a, f(a))$ and $(b, f(b))$. A second interpretation is that among the points $\zeta \in (a, b)$ that satisfy $f'(\zeta) = f[a, b]$ at least one corresponds to $\max\{|\epsilon_1(x)| : a \le x \le b\}$. With respect to interpolation our interest is an error estimate that depends more directly on $x \in [a, b]$ and so gives more information for a particular value of x. For linear interpolation this means determining the exact form of $\epsilon_1(x)$. We know that $f(x)$ and $p_1(x)$ match at $x = a$ and $x = b$. To extend the results of the Mean Value Theorem and at the same time use Rolle's Theorem we need another function that produces a zero at any fixed $x \in (a, b)$. One such function is

$$\begin{aligned} g(t) &= [f(t) - p_1(t)](x-a)(x-b) - [f(x) - p_1(x)](t-a)(t-b) \\ &= [f(t) - p_1(t)](x-a)(x-b) - (t-a)(t-b)\epsilon_1(x)\,. \end{aligned}$$

A simple check verifies that $g(t) = 0$ when $t = a, b, x$. Applying Rolle's Theorem once implies that $g'(t)$ vanishes at least twice in (a, b) and a second application gives the existence of $\zeta = \zeta(x)$ such that

$$0 = g''(\zeta) = [f''(\zeta) - p_1''(\zeta)](x-a)(x-b) - 2\epsilon_1(x)\,.$$

But $p_1(x)$ is linear so $p_1''(\zeta(x)) = 0$ for all x and solving for $\epsilon_1(x)$ we have

$$\epsilon_1(x) = \frac{f''(\zeta(x))}{2!}\,(x-a)(x-b)\,.$$

The more general statement we seek follows.

Theorem 4.3.3 *Suppose $f^{(n)}(x)$ is continuous is $[a, b]$ and $f^{(n+1)}(x)$ exists (a, b). Let $p_n(x)$ be a polynomial of degree n in x such that $p_n(x_i) = f(x_i)$ where $i = 0, 1, \ldots, n$ and we assume the x_i are distinct and located in the interval $[a, b]$. Given these conditions, for all $x \in [a, b]$ there exists $\zeta = \zeta(x) \in (a, b)$ such that*

$$f(x) - p_n(x) = \frac{f^{(n+1)}(\zeta(x))}{(n+1)!}\; q_{n+1}(x) \tag{4.3.1}$$

where, as before,

$$q_{n+1}(x) = (x - x_0)(x - x_1)\cdots(x - x_n)$$

Proof: Let us define a function $g(t)$ by

$$g(t) = [f(t) - p_n(t)]q_{n+1}(x) - [f(x) - p_n(x)]q_{n+1}(t)\,, \tag{4.3.2}$$

where x is fixed in $[a, b]$. By induction we can show that $g^{(n+1)}(t)$ vanishes at least once in (a, b), say at $t = \zeta$. We also know that $p_n^{(n+1)}(t) = 0$ for all t and $q_{n+1}^{(n+1)}(t) = (n+1)!$. Differentiating (4.3.1) we conclude that

$$0 = g^{(n+1)}(\zeta(x)) = f^{(n+1)}(\zeta(x))q_{n+1}(x) - [f(x) - p_n(x)](n+1)!\,,$$

or

$$\epsilon_n(x) \equiv f(x) - p_n(x) = \frac{f^{(n+1)}(\zeta(x))}{(n+1)!}\, q_{n+1}(x)\,.$$

4.3.1 The Equispaced Newton Polynomial

We can use the interpolation error to interpret the coefficients of the Newton polynomial. This is easier to do assuming equispaced grid points that also satisfy $x_0 < x_1 < \cdots < x_n$. We begin by noting that the polynomial $q_{n+1}(x)$ simplifies to

$$\begin{aligned} q_{n+1}(x) &= h^{n+1}\xi(\xi-1)\cdots(\xi-n) \\ &= h^{n+1}(n+1)!\binom{\xi}{n+1} \end{aligned}$$

where $x = x_0 + \xi h$ and $\xi \in [0, n]$. If f is an arbitrary function such that $f^{(n+1)}(x)$ exists for $x \in (x_0, x_n)$ then we may write f in terms of the Newton polynomial and error as

$$f(x_0 + \xi h) = \sum_{k=0}^{n} \binom{\xi}{k} \Delta^k f(x_0) + h^{n+1}\binom{\xi}{n+1} f^{(n+1)}(\zeta(x)) \tag{4.3.3}$$

where ζ lies between x_0 and $x_0 + nh$. Similarly, the represenation for f using the Newton polynomial of degree $n-1$ is

$$f(x_0 + \xi h) = \sum_{k=0}^{n-1} \binom{\xi}{k} \Delta^k f(x_0) + h^{n}\binom{\xi}{n} f^{(n)}(\nu(x)) \tag{4.3.4}$$

where ν lies between x_0 and x_{n-1}. If we now subtract (4.3.4) from (4.3.3), we have

$$\frac{\Delta^n f(x_0)}{h^n} = f^n(\nu(x)) - h\,\frac{\xi - n}{n+1} f^{(n+1)}(\zeta(x))\,.$$

Finally, taking $\xi = n$ and letting $h \to 0$, we find

$$f^{(n)}(x_0) = \lim_{h\to 0} \frac{\Delta^n f(x_0)}{h^n}\,. \tag{4.3.5}$$

If we drop the assumption of equispaced points then we can still arrive at a result similar to the one above. Again we assume that $x_0 < x_1 < \cdots < x_n$. Consider the two expressions

$$\begin{aligned} f(x) &= f(x_0) + f[x_0, x_1](x - x_0) + \cdots \\ &\quad + f[x_0, x_1, \ldots, x_n](x - x_0)(x - x_1)\cdots(x - x_{n-1}) \\ &\quad + \frac{q_{n+1}}{(n+1)!}\, f^{(n+1)}(\xi(x)) \\ &= f(x_0) + f[x_0, x_1]q_0(x) + \cdots + \\ &\quad f[x_0, x_1, \ldots, x_n]q_n(x) + \frac{q_{n+1}(x)}{(n+1)!}\, f^{(n+1)}(\xi(x)) \end{aligned} \tag{4.3.6}$$

and

$$\begin{aligned} f(x) &= f(x_0) + f[x_0, x_1]q_0(x) + \cdots + \\ &\quad f[x_0, x_1, \ldots, x_{n-1}]q_{n-1}(x) + \frac{q_n(x)}{n!}\, f^{(n)}(\nu(x))\,. \end{aligned} \tag{4.3.7}$$

Subtracting the second from the first we have

$$\frac{f^{(n)}(\nu(x))}{n!} = f[x_0, x_1, \ldots, x_n] + \frac{x - x_n}{(n+1)!}\, f^{(n+1)}(\xi(x))\,.$$

Taking $x = x_n$ and letting $x_n \to x_0$ we find that

$$\frac{f^{(n)}(x_0)}{n!} = \lim_{x_n \to x_0} f[x_0, x_1, \ldots, x_n]\,. \tag{4.3.8}$$

Also, by substituting this result into (4.3.6), it follows that

$$f(x) = \lim_{x_n\to x_0} \{f(x_0) + f[x_0, x_1](x - x_0) + \cdots +$$

$$f[x_0,\ldots,x_n](x-x_0)\cdots(x-x_n)+\frac{q_{n+1}(x)}{(n+1)!}\,f^{(n+1)}(\zeta(x))\}$$

$$(4.3.9)$$

$$\begin{aligned}&= f(x_0)+f'(x_0)(x-x_0)+\cdots+\frac{f^{(n)}(x_0)}{n!}\,(x-x_0)^n\\&\quad+\frac{f^{(n+1)}(\zeta(x))}{(n+1)!}\,(x-x_0)^{n+1},\end{aligned}$$

where $\zeta(x)$ lies between x_0 and x. We therefore see that the Taylor polynomial can be regarded as a special case of the Newton polynomial.

Example 4.3.4 If in (4.3.3) $f(x)=a^x$ and $x_0=0$ by induction we can show that $\Delta^k a^{x_0}=(a^h-1)^k$. Since $f^{(n+1)}(x)=a^x(\log a)^{n+1}$ we now have

$$a^{\xi h}=\sum_{k=0}^{n}\binom{\xi}{k}(a^h-1)^k+(h\log a)^{n+1}\binom{\xi}{n+1}a^{\nu}$$

where ν is between the largest and smallest of 0, ξh and nh. But

$$\begin{aligned}\binom{\xi}{n+1} &= \frac{\xi(\xi-1)\cdots(\xi-n)}{(n+1)!}\\ &= \frac{(-1)^{n+1}(n-\xi)(n-\xi-1)\cdots(1-\xi)(-\xi)}{(n+1)\Gamma(n+1)}\\ &= \frac{(-1)^{n+1}\Gamma(n+1-\xi)}{\Gamma(-\xi)(n+1)\Gamma(n+1)}.\end{aligned}$$

Stirling's formula gives

$$\frac{\Gamma(n+1-\xi)}{\Gamma(n+1)}=\frac{n^{n+1-\xi-1/2}}{n^{n+1-1/2}}=n^{-\xi}$$

so finally

$$\left|\binom{\xi}{n+1}\right|=\frac{1}{(n+1)n^{\xi}\Gamma(-\xi)}\quad\text{as } n\to\infty\,.$$

However, this last term decays to zero as $n\to\infty$ if $\xi>-1$. If we also assume that $|a^h h\log a|=|h\log a\exp(h\log a)|<1$ then

$$\sum_{k=0}^{n}\binom{\xi}{k}(a^h-1)^k\to a^{\xi h} \tag{4.3.10}$$

as $n\to\infty$.

The need for care in the practical use of interpolating polynomials has already been suggested in the previous section. Although electronic computers easily evaluate polynomials of high order, it is nevertheless necessary to examine closely the error. One criterion for estimating the error of interpolation is given by (4.3.2). The error is a product of $f^{(n+1)}(\xi(x))/(n+1)!$ and $q_{n+1}(x)$. In the next section we thoroughly examine the dependence of the error on $q_{n+1}(x)$. For example, if $|f^{(n+1)}(x)| \leq M < \infty$ for all x in $[a, b]$, then $|f^{(n+1)}(\xi(x))| \leq M$ for all x in $[a, b]$, and therefore

$$|f(x) - p_n(x)| \leq \frac{M}{(n+1)!} \, |q_{n+1}(x)| \, . \tag{4.3.11}$$

The maximum value of the bound on the right is minimized if we choose $q_{n+1}(x)$ so that its maximum value on the interval $[a, b]$ is a minimum. Specifying $q_{n+1}(x)$ refers to selecting the points $x_0, x_1, \ldots, x_n$. We discuss the minimizations of the maximum of $|q_{n+1}(x)|$ in the following section.

For now, we comment on the effect of the error term $f^{(n+1)}$. To a great extent this term indicates how closely the polynomial preserves the character of the interpolated function. We see that at times polynomials are inadequate tools of approximation. For instance, suppose that $f(x) = 0$ for $x = x_1$, x_2 and x_3, and that $f(x_0) = y_0$ where $x_0 = 0$ and $x_0 < x_1 < x_2 < x_3$. The value y_0 may or may not be finite. If $f(x)$ is unbounded at $x = x_0$ then y_0 simply distinguishes between $\pm\infty$. The polynomial that interpolates $f(x)$ at these four points is

$$p_3(x) = \frac{-y_0(x - x_2)(x - x_3)(x - x_4)}{x_2 x_3 x_4} \, .$$

This polynomial is shown as the dotted curve in Figure 4.3.1.
The polynomial $p_3(x)$ is a good approximation to $f(x)$ if $f(x)$ admits close approximation by a polynomial throughout the interval $[0, x_4]$. This assumption is false when $f(x)$ is infinite at $x = x_1$. In this case no polynomial provides a satisfactory approximation. The rational function

$$g(x) = \frac{-y_0(x - x_2)(x - x_3)(x - x_4)}{x x_2 x_3 x_4} = \frac{p_3(x)}{x}$$

may now be a better approximation. The function $g(x)$ is represented by the continuous curve in Figure 4.3.1. Of course, if we know the rate at which $f(x)$ blows up at $x = x_0$ we can build this information into the rational approximation. For example, if

$$f_1(x) = \frac{\sin x}{x^2} \quad \text{and} \quad f_2(x) = \frac{\sin x}{x^3}$$

Figure 4.3.1: Example of Interpolation

then the rational interpolant for $f_1(x)$ should behave like $1/x$ near $x = 0$ while the interpolant behavior for $f_2(x)$ should mimic $1/x^2$ near $x = 0$.

This example is not artificial; rather, it emphasizes an essential limitation of polynomial approximation. Several excellent references are given at the end of this chapter. We note that the theory of the Gamma function is the most completely worked out theory of an interpolating function.

Problems

4.3.1 Verify (4.3.10).

4.3.2 *(This problem should be performed with the aid of a computer package such as MATLAB.)* Consider the function $f(x) = 1/(1+25x^2)$ on the interval $[-1, 1]$. Interpolate $f(x)$ at the points $x_i = -1 + ih$ where $h = .2$ and $i = 0, 1, \ldots, 11$. Graph $f(x)$ with a solid line and the interpolant with a dashed line. Comment on the quality of interpolant. Extend the interval to $[-2, 2]$ and perform a similar analysis.

4.3.3 Prove that the polynomial $H_{2n+1}(x)$ in (4.2.17) satisfies

$$f(x) - H_{2n+1}(x) = \frac{f^{(2n+2)}(\xi(x))}{(2n+2)!}\,(x - x_0)^2(x - x_1)^2 \cdots (x - x_n)^2$$

where $\xi(x)$ is between the largest and smallest of $x_0, x_1, \ldots, x_n$ and x.

4.3.4 (a) Show that $f[x,x] = \dfrac{d}{dx} f[x]$. Hint: Let $x_1 = x + \delta$ and consider $\lim\limits_{\delta \to 0} f[x_1, x]$.

(b) Show that $f[x_0, x_1, \ldots, x_m, x, x] = \dfrac{d}{dx} f[x_0, x_1, \ldots, x_m, x]$.

4.3.5 Theorem 4.3.3 gives us the exact from of $\epsilon_1(x)$ where

$$f(x) = f[x_0] + (x - x_0) f[x_0, x_1] + \epsilon_1(x)\,.$$

Using the relation above and the identity

$$f[x_0, x_1, x] = \frac{f[x_0, x] - f[x_0, x_1]}{x - x_1}$$

show that

$$f[x_0, x_1, x] - \frac{f''(\zeta(x))}{2!}\,, \quad \zeta \in (x_0, x_1)\,.$$

In general show that if $x_0, x_1, \ldots, x_n$ lie in $[a,b]$ then

$$f[x_0, x_1, \ldots, x_n, x] = \frac{f^{(n+1)}(\zeta(x))}{(n+1)!}\,, \quad \zeta \in (a,b)\,.$$

4.3.6 Show that if $f(x) = \mathrm{erf} x$ where

$$\mathrm{erf} x \equiv \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2}\, dt$$

then $|\epsilon_1(x)|$ cannot exceed $(x_1 - x_0)^2 / (2\sqrt{2\pi e})$ between x_0 and x_1.

4.3.7 Suppose that $f(x) = \int_0^x \sin t^2\, dt$ is tabulated for $0 \le x \le 1$ with uniform spacing h. What spacing should be used to guarantee an accuracy within 0.005?

4.3.8 Let $x_0, x_1, \ldots, x_n$ lie in the interval $[a,b]$. If $f(z)$ is analytic in a simply connected region D fo the complex plane which includes the line segment $[a,b]$ and if C is a simple closed curve in D that surrounds $[a,b]$ show that

$$f[x_0, x_1, \ldots, x_n, x] = \frac{1}{2\pi i} \oint_C \frac{f(z)\, dz}{q_{n+1}(z)(z - x)}$$

where, as usual, $q_{n+1}(x) = (x - x_0)(x - x_1) \cdots (x - x_n)$. Then show that

$$|\epsilon_n(x)| \le \frac{ML}{2\pi d} \max_{z \in C} \left| \frac{q_{n+1}(x)}{q_{n+1}(z)} \right|$$

where $|f(z)| \le M$ on C, L is the length of C and d is the shortest distance from the real point x to a point on C.

4.4 The Polynomial of Least Deviation from Zero in $[-1, 1]$

In the previous section we derived the error statement

$$f(x) - p_n(x) = \epsilon_n(x) \equiv \frac{f^{(n+1)}(\zeta(x))q_{n+1}(x)}{(n+1)!} \tag{4.4.1}$$

assuming that $x, x_i \in [a, b]$, and $f^{(n+1)}(x)$ exists for all $x \in (a, b)$. In this case $\zeta(x) \in (a, b)$ if $x \in [a, b]$; therefore we can bound $\epsilon_n(x)$ if $f^{(n+1)}(x)$ is bounded on $[a, b]$ as follows:

$$|\epsilon_n(x)| \leq \left(\max_{x \in [a,b]} \left| \frac{f^{(n+1)}(x)}{(n+1)!} \right| \right) \left(\max_{x \in [a,b]} |q_{n+1}(x)| \right) . \tag{4.4.2}$$

We make two comments regarding this approach. The first is that we have implicity chosen to discuss error in terms of the infinity norm. The second is that separating the maximization of $|f^{(n+1)}(x)|$ from that of $|q_{n+1}(x)|$ is not the only way to proceed. The next two sections offer alternatives to these options. The proper choice of norm depends on a given application and more than one choice is often valid. We do not say more in this regard. However, with respect to segregating the minimization of $\max |q_{n+1}(x)|$ we see the advantage of a more universally applicable result since a reasonable small value for this term is satisfactory for the interpolation of many functions.

We begin by translating the abstract problem of minimizing the $\max |q_{n+1}(x)|$ to a concrete problem with a list of criteria. Recall that

$$q_{n+1}(x) = (x - x_0)(x - x_1) \cdots (x - x_n) .$$

One feature this polynomial has independent of x_i is that the coefficient of x^{n+1} is $a_{n+1} = 1$. We now define the minimum deviation problem as finding the polynomial of degree $n + 1$ whose leading coefficient is 1, whose zeros $x_0, x_1, \ldots, x_n$ lie in $[a, b]$, and such that its uniform norm on $[a, b]$ is as small as possible. We must select the proper set of zeros to complete the minimax problem. Rather than perform the analysis on $[a, b]$ we choose to work on $[-1, 1]$ and then explain how to transfer the result to any finite interval $[a, b]$. Hence we are now assuming $x \in [-1, 1]$.

We show that the solution of

$$\min_{x_i \in [-1,1]} \max |q_{n+1}(x)|$$

is satisfied by the choice

$$q_{n+1}(x) = t_{n+1}(x)$$

where

$$t_n(x) = \frac{\cos n\theta}{2^{n-1}} \quad , \quad x = \cos\theta \quad , \quad \theta \in [0, \pi] \tag{4.4.3}$$

is known as the n-th *Chebyshev* polynomial of the second kind. Several comments are in order. First the mapping $x = \cos\theta$ is one–to–one from $[0, \pi]$ onto $[-1, 1]$. Next, $t_n(x)$ is indeed a polynomial. To prove this note that

$$\begin{aligned} t_0(x) &= 2 \\ t_1(x) &= x\,. \end{aligned}$$

Assuming that $t_j(x)$ where $j = 2, \ldots, n$ are polynomials and using the trigonometric identity

$$\cos(a + b) + \cos(a - b) = 2 \cos a \cos b$$

gives

$$\cos(n + 1)\theta = 2 \cos n\theta \cos\theta - \cos(n - 1)\theta$$

or

$$t_{n+1}(x) = x t_n(x) - \frac{1}{4}\, t_{n-1}(x)\,.$$

Hence, by induction we have shown that $t_n(x)$ is a polynomial of degree n whose leading coefficient is $a_n = 1$. We also have a recursion relation so that, for instance,

$$\begin{aligned} t_2(x) &= x t_1(x) - \tfrac{1}{4} t_0(x) &&= x^2 - \frac{1}{2} \\ t_3(x) &= x\left(x^2 - \tfrac{1}{2}\right) - \tfrac{1}{4}\, x &&= x^3 - \frac{3}{4}\, x\,. \end{aligned}$$

Next, we must examine the zeros and critical points of $t_{n+1}(x)$. The former correspond to

$$\theta_k = \frac{(2k + 1)\pi}{2(n + 1)} \;, \quad k = 0, 1, \ldots, n \tag{4.4.4}$$

and the latter critical points occur when

$$t'_{n+1}(x) = \frac{n + 1}{2^n (1 - x^2)^{1/2}} \sin(n + 1)\theta = 0\,.$$

To distinguish between the zeros of $t_{n+1}(x)$ and the critical points we denote the critical points as $\hat{\theta}_k$ where, from the equation above,

$$\hat{\theta}_k = \frac{k\pi}{n+1}\,, \quad k = 0, 1, \ldots, n\,. \tag{4.4.5}$$

Since $|\cos(n+1)\hat{\theta}_k| = 1$ for all k we have

$$\max_{-1 \le x \le 1} |t_{n+1}(x)| = 2^{-n}\,. \tag{4.4.6}$$

Suppose that we interpolate a function f on the interval $[-1, 1]$ at the points $x_i = \cos\left[\frac{(2i+1)\pi}{2n+2}\right]$, $i = 0, 1, \ldots, n$ using the Lagrange representation. In this case

$$\ell_k(x) = \frac{t_{n+1}(x)}{(x - x_k)[t'_{n+1}(x)]\,|_{x=x_k}}\,.$$

Now

$$\frac{d}{dx}\cos(n+1)\theta = \frac{(n+1)\sin(n+1)\theta}{(1-x^2)^{1/2}}$$

and so

$$\frac{d}{dx}\cos(n+1)\theta\,|_{\theta=\theta_k} \;= \frac{(n+1)(-1)^k}{(1-x_k^2)^{1/2}}\,.$$

Substituting into the expression for $\ell_k(x)$ gives

$$\ell_k(x) = \frac{(-1)^k(1-x_k^2)^{1/2}2^n t_{n+1}(x)}{(n+1)(x-x_k)}\,. \tag{4.4.7}$$

¿From the results above we now have

$$\left| f(x) - \sum_{k=0}^{n} \frac{f(x_k)(-1)^k(1-x_k^2)^{1/2}2^n t_{n+1}(x)}{(n+1)(x-x_k)} \right| \le 2^{-n} \left| \frac{f^{(n+1)}(\zeta(x))}{(n+1)!} \right|\,. \tag{4.4.8}$$

It may also be shown that (Problem 4.4.3).

$$|\ell_i(x)| \le 2 \;\text{ if }\; |x| \le 1\,. \tag{4.4.9}$$

Let us now state and prove a theorem that characterizes the polynomial $t_{n+1}(x)$.

Theorem 4.4.1 *Let*

$$p_{n+1}(x) = x^{n+1} + a_n x^n + \cdots + a_0 \tag{4.4.10}$$

and let us set

$$G(p_{n+1}) = \max_{-1 \le x \le 1} |p_{n+1}(x)| . \tag{4.4.11}$$

Then $G(p_{n+1}) \ge 2^{-n}$. *Equality occurs if and only if*

$$p_{n+1}(x) = t_{n+1}(x) = 2^{-n} \cos((n+1) \arccos x) .$$

Proof: (by contradiction) Suppose that there is a polynomial $p_{n+1}(x)$, say

$$p_{n+1}(x) = x^{n+1} + b_n x^n + \cdots + b_0$$

such that

$$G(p_{n+1}) < 2^{-n} .$$

Then

$$r_n(x) = p_{n+1}(x) - t_{n+1}(x)$$

is a polynomial of degree n since both coefficients of x^{n+1} in $p_{n+1}(x)$ and $t_{n+1}(x)$ are 1. Furthermore $r_n(x)$ is not identically zero. At the points $x_k = \cos(\hat{\theta}_k)$ where $k = 0, 1, \ldots, n$

$$t_{n+1}(x_k) = (-1)^k 2^{-n}$$

but by hypothesis

$$|p_{n+1}(x_k)| < 2^{-n} .$$

Therefore $r_n(x_0) < 0$, $r_n(x_1) > 0, \ldots, r_n(x_k)(-1)^k < 0, \ldots$ or, more simply, $r_n(x)$ changes sign $n+1$ times on the interval $(-1, 1)$. Now the polynomial $r_n(x)$ is only of degree n in x, hence $r_n(x) \equiv 0$, a contradiction. This gives $G(p_{n+1}) \ge 2^{-n}$ but equality follows immediately from the construction of $t_{n+1}(x)$.

We close this section by transferring our results to the interval $[a, b]$. The interval $-1 \le \xi \le 1$ maps one–to–one onto the interval $a \le x \le b$ via

$$x = \frac{a+b}{2} + \frac{b-a}{2} \xi .$$

If $x_i = (a+b)/2 + (b-a)\xi_i/2$ then

$$x - x_i = \frac{b-a}{2} (\xi - \xi_i)$$

and

$$q_{n+1}(x) = \left(\frac{b-a}{2}\right)^{n+1} \prod_{i=0}^{n} (\xi - \xi_i)\,.$$

¿From the analysis above we know that $\max\{|q_{n+1}(x)| \,:\, a \le x \le b\}$ is minimized by selecting

$$x_k = \frac{a+b}{2} + \frac{b-a}{2}\cos\theta_k$$

where $\theta_k = \cos((2k+1)\pi/(2n+2))$. With this set of zeros we denote $q_{n+1}(x)$ by $\hat{t}_{n+1}(x)$ and note that

$$\begin{aligned}
\max_{a\le x\le b} |\hat{t}_{n+1}(x)| &= \left(\frac{b-a}{2}\right)^{n+1} max_{-1\le\xi\le1}|t_{n+1}(x)| \\
&= \left(\frac{b-a}{2}\right)^{n+1} 2^{-n} \\
&= \frac{(b-a)^{n+1}}{2^{2n+1}}\,.
\end{aligned}$$

Problems

4.4.1 Show that $t_n(x)$ is even or odd respectively when n is even or odd.

4.4.2 The polynomial $T_n(x) = \cos n\theta$, where $x = \cos\theta$, is called the Chebyshev polynomial of degree n in x. Clearly

$$T_n(x) = 2^{n-1} t_n(x).$$

Show that the Chebyshev polynomials have the following properties.

(a) $(1-x^2)T_n''(x) - xT_n'(x) + n^2 T_n(x) = 0$

(b) $T_n(x) = 2xT_{n-1}(x) - T_{n-2}(x)$ where $n = 2, 3, \ldots$

(c) Orthogonality

$$\int_{-1}^{1} T_n(x)T_m(x)\,\frac{dx}{(1-x^2)^{1/2}} = \begin{cases} 0 & n \ne m \\ \pi & n = m = 0 \\ \pi/2 & n = m \ne 0 \end{cases}$$

(d) Semi-group property

$$T_m(T_n(x)) = T_{mn}(x) \ ; \quad m, n > 0$$

4.4.3 Show that if

$$x_i = \cos\left[\frac{(2i+i)\pi}{2(n+1)}\right] \quad , \quad i = 0, 1, \ldots, n \, ,$$

and

$$\ell_i(x) = \frac{(-1)^i(1-x_i^2)\cos[(n+1)\arccos x]}{(n+1)(x-x_i)}$$

then $|\ell_i(x)| \leq 2$ for $-1 \leq x \leq 1$.

4.4.4 Let $p_n(x)$ be the polynomial of degree n in x that interpolates e^x at the zeros of $T_n(x) = \cos(n \arccos x)$. Prove that for $-1 \leq x \leq 1$,

$$|e^x - p_n(x)| \leq \frac{e}{2^n n!} \, .$$

4.5 Best Approximation

The theory and background of best approximation is a deep subject that evolved over many years. We can not duplicate this work in one section but we can and do highlight the major results.

Let f be a given continuous function on the interval $[-1,1]$. We seek a polynomial $p_{n+1} \in \Pi_{n+1}$ that approximates f well. If by "approximating well" we mean exact interpolation of f at the $(n+2)$ points $\{x_i\}_{i=0}^{n+1}$ where $-1 \leq x_0 < x_1 < \cdots < x_{n+1} \leq 1$ then a solution of the problem is provided by the Lagrange interpolate L_{n+1} of f at $\{x_i\}$. However, the error between f and L_{n+1} away from the nodes can be huge. If L_{n+1} is used in place of f in a numerical process such as integration then L_{n+1} may not approximate f well. We must always specify the approximation considered as well as the norm measuring error. Throughout this chapter we examine only interpolation via polynomials. In the next section we measure error in terms of the L_2 norm while in this section we turn our attention to the L_∞ norm.

Before stating the precise definition of best approximation in the L_∞ norm we motivate the discussion with the following simple situation. Find $b_1(x) = a_1 x + a_0$ so that for $f \in C[-1,1]$

$$E_1(f) = \max_{-1 \leq x \leq 1} |f(x) - b_1(x)| \leq \max_{-1 \leq x \leq 1} |f(x) - p(x)| \qquad (4.5.1)$$

for all $p \in \Pi_1$. In minimax language define

$$E_1(f) = \max_{-1\leq x\leq 1} |f(x) - b_1(x)| = \min_{p\in\Pi_1} \left\{ \max_{-1\leq x\leq 1} |f(x) - p(x)| \right\}$$

where for now we assume the problem is well defined and $b_1(x)$ exists. The solution for a given f is justifiable geometrically. Referring to Figure 4.5.1 note that the dotted line corresponding to $d_1(x) = \alpha x + \beta$ cannot be the solution since by lowering the line

$$|f(x) - \tilde{d}_1(x)| < |f(x) - d_1(x)| \, .$$

On the other hand, the solid curve $b_1(x)$ looks to be a better approximate

Figure 4.5.1:

to f than either d_1 or $\tilde{d}_1$. We eventually show that $b_1(x)$ is the unique best approximate for the problem defined by (4.5.1). The distinguishing feature $b_1(x)$ exhibits is that the error $E_1(f)$ occurs at three points and with alternate signs. Hopefully, this provides some motivation for the general setting outlined in the following two theorems.

Theorem 4.5.1 *Let $f \in C[a,b]$. Then there exists a unique $b_n \in \Pi_n$ satisfying*

$$E_n(f) \equiv \max_{x\in[a,b]} |f(x) - b_n(x)| \leq \max_{x\in[a,b]} |f(x) - p_n(x)| \qquad (4.5.2)$$

for all $p_n \in \Pi_n$. *Moreover there are at least* $n+2$ *points* $x_0 < x_1 < \cdots < x_{n+1}$ *in* $[a,b]$ *so that*

$$f(x_i) - b_n(x_i) = (-1)^i E_n(f)\,,\ i = 0, \ldots, n+1\,. \tag{4.5.3}$$

The set $\{x_i\}_{i=0}^{n+1}$ in Theorem 4.5.1 is called an alternating set for the function f and the polynomial b_n is called the best approximate to f from Π_n. The "equi–oscillation" of the error in (4.5.3) characterizes best approximates. We emphasize that while b_n is unique this is not necessarily true of the alternate set.

Theorem 4.5.2 *Assume that* $f \in C[a,b]$ *and there is a* $p_n \in \Pi_n$ *satisfying*

$$\max_{a \le x \le b} |f(x) - p_n(x)| = \mu\,.$$

If there is a set of $n+2$ *points* $a \le x_0 < x_1 < \cdots < x_{n+1} \le b$ *so that*

$$f(x_i) - p_n(x_i) = (-1)^i \mu\,, \quad i = 0, 1, \ldots, n+1$$

then $\mu = E_n(f)$ *and* $p_n \equiv b_n$.

To gain insight into the previous theorems and what is involved in constructing a best approximate, consider the following example.

Example 4.5.3 The function $f(x) = |x|$ is in $C[-1,1]$. Suppose we seek the best quadratic approximate b_2 to f. Convince yourself that $b_2(x) = \alpha x^2 + \beta$ since f is even; indeed, for the same reason $b_3(x) = b_2(x)$. An alternation set for f contains at least $n+2 =$ 4 points but due to symmetry actually contains 5 points. Label these $\{-1, -\xi, 0, \xi, 1\}$ (Problem 4.5.1 asks you to show that 0 and ± 1 belong to the alternation set). Using (4.5.3) we have

$$\begin{aligned}
f(-1) - b_2(-1) = 1 - (\alpha + \beta) &\quad = -E_1(f) && \text{(a)}\\
f(-\xi) - b_2(-\xi) = \xi - (\alpha\xi^2 + \beta) &\quad = +E_1(f) && \text{(b)}\\
f(0) - b_2(0) = -\beta &\quad = -E_1(f) && \text{(c)}
\end{aligned}$$

Equation (c) gives $\beta = E_1(f)$; then using this result along with equation (a) we have $\alpha = 1$. The alternating set determines the value of β. We know that ξ corresponds to a critical point of (b). Differentiating equation (b) with respect to ξ shows that $\xi = 1/2$. Finally, solving (b) given α and ξ we have $\beta = 1/8$. Hence $b_2(x) = x^2 + 1/8$.

The previous example is typical. That is, the $n+2$ equations in (4.5.3) defining $E_n(f)$ and the $n+1$ coefficients of b_n are nonlinear in the sense that we must determine an alternating set. Theorem 4.5.2 is a basis for approximation schemes that compute $b_n(x)$ but if a high degree approximation is

Figure 4.5.2: Best Approximation to $f(x) = |x|\ , \quad -1 < x < 1\,.$

sought then determining an alternating set valid on the interval $[a, b]$ is often prohibitively expensive. Hence we defer further discussion of this topic to the various approximation theory books of the appendix and instead consider other options. One option is that we split $[a, b]$ into subintervals and calculate low degree approximations on each. This is the basis of splines. The second option is that we seek a best approximation on a set of points in $[a, b]$ rather than on the entire interval. This is the subject of the next two theorems.

Theorem 4.5.2 remains valid if f is defined at $n+2$ points $x_0 < x_1 < \cdots < x_{n+1}$ and the maximum is taken over this finite set. On the finite set we can explicity calculate the best approximation and the error. The disadvantage is that the error may not extend to the entire interval. The theorem that covers this setting is easier to state if we first construct the best approximate on the set of points. This construction combines the ideas of Lagrange interpolation and alternating sets. Note at the conclusion that the algorithm is linear. Let S_{n+2} denote a finite set of $n+2$ points (x_i, y_i),

$i = 0, 1, \ldots, n+1$. We seek the unique polynomial $b_n^*(x) \in \Pi_n$ such that

$$E_n^*(f) = \max_{(x_i, y_i) \in S_{n+2}} |y_i - b_n^*(x_i)| \leq \min_{p_n \in \Pi_n} \left\{ \max_{(x_i, y_i) \in S_{n+2}} |y_i - p_n(x_i)| \right\} . \tag{4.5.4}$$

Recall that the Lagrange polynomial that interpolates the function values $y_i = f(x_i)$ for $i = 0, 1, \ldots, n = 1$ is

$$L_{n+1}(x) = \sum_{i=0}^{n+1} \frac{y_i q_{n+2}}{(x - x_i) q'_{n+2}(x_i)}$$

where

$$q_{n+1}(x) = \prod_{i=0}^{n+1} (x - x_i) .$$

Clearly $L_{n+1}(x)$ is not the polynomial we seek since $L_{n+1}(x) \in \Pi_{n+1}$ and we have not yet introduced an alternating set. However, a sketch of $q_{n+2}(x)$ and $q'_{n+2}(x)$ suggests a way to introduce such a set. This polynomial has the properties that $q'_{n+2}(x_i) \neq 0$ and $(-1)^i q'_{n+2}(x_i)$ has the same sign for $i = 0, 1, \ldots, n+1$. The first property gives a nonzero scaling factor while the second guarantees points of alternation so that the polynomial

$$u_{n+1}(x) = \sum_{i=0}^{n+1} \frac{(-1)^i q_{n+2}(x)}{(x - x_i) q'_{n+2}(x_i)}$$

satisfies the identity $u_{n+1}(x_i) = (-1)^i$. Further, the sign property implies that the coefficient of x^{n+1} of $u_{n+1}(x)$ does not vanish. Hence, in general both $L_{n+1}(x)$ and $u_{n+1}(x)$ are polynomials of degree $n+1$. The coefficient of x^{n+1} in $L_{n+1}(x)$ is

$$\sum_{i=0}^{n+1} \frac{y_i}{q'_{n+2}(x_i)}$$

while the corresponding coefficient for $u_{n+1}(x)$ is

$$\sum_{i=0}^{n+1} \frac{(-1)^i}{q'_{n+2}(x_i)} .$$

Given that we know the coefficients of x^{n+1} an explicit method exists for combining $L_{n+1}(x)$ and $u_{n+1}(x)$ so that

$$b_n^*(x) = L_{n+1}(x) - \mu u_{n+1}(x) \tag{4.5.5}$$

where

$$\mu = \sum_{i=0}^{n+1} \frac{y_i}{q_{n+1}(x_i)} \Big/ \sum_{i=0}^{n+1} \frac{(-1)^i}{u'(x_i)} . \tag{4.5.6}$$

We have not shown that this combination actually gives $b_n^*(x)$ but the statement is the next theorem.

Theorem 4.5.4 *The polynomial $b_n^*(x)$ defined in (4.5.5) and (4.5.6) is the unique polynomial that satisfies (4.5.4) with $m = n + 2$.*

In the case of a set S_m of m points (x_i, y_i), where $m \geq n+2$, and where we assume that the x_i are all distinct. From S_m we can select $\binom{m}{n+2}$ different sets $S_{m,k}$, $k = 1, 2, \ldots, \binom{m}{n+2}$, where

$$S_{m,k} = \left\{ \left(x_0^{(k)}, y_0^{(k)}\right), \left(x_1^{(k)}, y_1^{(k)}\right), \ldots, \left(x_{n+1}^{(k)}, y_{n+1}^{(k)}\right) \right\}$$

and where $x_0^{(k)} < x_1^{(k)} < \cdots < x_{n+1}^{(k)}$. On each of the sets $S_{m,k}$ we approximate the ordinates by a polynomial $p_{n,k}^*(x)$ chosen from Π_n and constructed along the lines following (4.5.4) such that

$$y_j^{(k)} - p_{n,k}^* \left(x_j^{(k)}\right) = (-1)^{j+1} \mu_k , \quad j = 0, 1, \ldots, n+1 . \tag{4.5.7}$$

Let ℓ be the integer chosen from $\left\{1, 2, \ldots, \binom{m}{n+2}\right\}$ such that $|\mu_k| \leq |\mu_\ell|$ for $k = 1, 2, \ldots, \binom{m}{n+2}$, and let $p_{n,\ell}^*(x)$ be the corresponding polynomial be the polynomial corresponding to the set $S_{m,\ell}$.

Theorem 4.5.5 *Let $p_{n,\ell}^*$ be the polynomial defined as in the above paragraph. Then $p_{n,\ell}^*(x)$ is the unique polynomial in Π_n that satisfies (4.5.4). Moreover, on the set of point $\{x_i\}_{i=0}^m$*

$$|y_i - p_{n\ell}^*(x_i)| \leq |\mu_\ell| . \tag{4.5.8}$$

Proof: We only prove the uniqueness of $p_{n,\ell}^*(x)$; we illustrate the construction of $p_{n,\ell}^*(x)$ via the exchange algorithm, by means of examples.

Let us assume that there are two distinct polynomials, $p_{n,\ell}^*(x)$ and $\hat{p}_n(x)$, both of which are in Π_n, and satisfy

$$|y_i - p_{n,\ell}^*(x_i)| \leq |\mu_\ell| , \quad |y_i - \hat{p}_n(x_i)| \leq |\mu_\ell| , \quad i = 0, 1, \ldots, m . \tag{4.5.9}$$

We set $r_n(x) = \frac{1}{2}\left[p_{n,\ell}^*(x) + \hat{p}_n(x)\right]$. Then there are again $n+2$ points (x_{s_i}, y_{s_i}) in S_m, $x_{s_0} < x_{s_1} < \cdots < x_{s_{n+1}}$ (why?) such that

$$y_{s_i} - r_n(x_i) = (-1)^{i+1} \mu , \quad i = 0, 1, \ldots, n+1 ,$$

where $\mu = \pm|\mu_\ell|$ (why?). That is,

$$[y_{s_i} - p^*_{n,\ell}(x_{s_i})] + [y_{s_i} - \hat{p}_n(x_{s_i})] = 2(-1)^{i+1}\mu\,.$$

However, because of (4.5.9), it follows that

$$y_{s_i} - p^*_{n,\ell}(x_{s_i}) = y_{s_i} - \hat{p}_n(x_{s_i}) = (-1)^{i+1}\mu\,,$$

that is, $p^*_{n,\ell}(x_{s_i}) = \hat{p}_n(x_{s_i})$. Since there are $n+2$ such points, $p^*_{n,\ell}(x) = \hat{p}_n(x)$ for all x.

Example 4.5.6 For given $S_3 = \{(x_i, y_i) : i = 0, 1, 2\}$ let us construct the straight line, $b_1(x) = ax + b$ that satisfies $y_i - p_i(x_i) = (-1)^{i+1}\mu$, $i = 0, 1, 2$. However, instead of using the formulas (4.5.5) to (4.5.6), we carry out the computation directly by solving the linear system

$$\begin{aligned} ax_0 + b &= y_0 - \mu \\ ax_1 + b &= y_1 + \mu \\ ax_2 + b &= y_2 - \mu \end{aligned}$$

which yields

$$\begin{aligned} a &= \frac{y_2 - y_0}{x_2 - x_0} \\ b &= \frac{1}{2}\,\frac{y_0(x_1 + x_2) + y_1(x_2 - x_0) - y_2(x_0 + x_1)}{x_2 - x_0} \\ \mu &= y_0 - ax_0 - b \\ &= \frac{1}{2}\,\frac{y_0(x_2 - x_1) - y_1(x_2 - x_0) + y_2(x_1 - x_0)}{x_2 - x_0} \end{aligned}$$

If this procedure is used to find the best linear approximation to the data $(-1, 1)$, $(0, 0)$, and $(1, 1)$ then $a = 0$, $b = 1/2$ and $b_1(x) = 1/2$. Compare this with the best linear approximation to $f(x) = |x|$, $x \in [-1, 1]$ in Example 4.5.3. What happens in the present case if the data is changed to the set $(-1, 1)$, $(0, 0)$, $(1/2, 1/2)$? (Answer: $b_1(x) = (1-x)/3$.)

Example 4.5.7 Construct the best linear approximation to the data

$$S_4 = \{(-1, 1/2), (-1/2, 4/5), (1/2, 4/5), (1, 1/2)\}\,.$$

This reflects the situation in the construction preceeding Theorem 4.5.5; e.g. $m = 4$ and $n = 1$ (linear). There are $\binom{4}{3} = 4$ sets $S_{4,k}$, namely

$$\begin{aligned} S_{4,1} &= \{(-1, 1/2), (-1/2, 4/5), (1/2, 4/5)\} \\ S_{4,2} &= \{(-1, 1/2), (-1/2, 4/5), (1, 1/2)\} \\ S_{4,3} &= \{(-1, 1/2), (1/2, 4/5), (1, 1/2)\} \\ S_{4,4} &= \{(1/2, 4/5), (1/2, 4/5), (1, 1/2)\} \end{aligned} \tag{4.5.10}$$

As in the lines preceeding Theorem 4.5.5 label the points of $S_{4,1}$ $(x_0^1, y_0^1) = (-1, 1/2)$; $(x_1^1, y_1^1) = (-1/2, 4/5)$ and $(x_2^1, y_2^1) = (1/2, 4/5)$. Using (4.5.6) and (4.5.8) with $n = 1$ one finds

$$\hat{p}_2(x) = 4x^2 + 2x - 1 \quad \text{and} \quad p_2(x) = -2/5\, x^2 + 9/10\,.$$

Instead of using (4.5.6) to find $\hat{p}_2$ it is perhaps simpler to put $\hat{p}_2(x) = ax^2 + bx + c$ and require $\hat{p}_2(x_i^1) = (-1)^i$, $i = 0, 1, 2$. Similarly for the construction of $p_2(x)$. Hence, using (4.5.5) we have

$$\begin{aligned} p_{1,1}^*(x) &= p_2(x) - \mu_1 \hat{p}_2(x) \\ &= (-2/5 - 4\mu_1)x^2 - 2\mu_1 x + (9/10 + \mu_1) \end{aligned}$$

so that (requiring $p_{1,1}^* \in \Pi_1$) gives

$$\mu_1 = \frac{-2/5}{4} = -1/10$$

and

$$p_{1,1}^*(x) = 1/5\, x + 4/5\,.$$

In an exactly similar fashion one finds for the set $S_{4,2}$ in (4.5.10)

$$\hat{p}_2(x) = 8/3\, x^2 - 5/3 \quad \text{and} \quad p_2(x) = -2/5\, x^2 + 9/10$$

so that

$$\mu_2 = \frac{-2/5}{8/3} = -3/20$$

and

$$p_{1,2}^*(x) = 13/20\,.$$

You are asked in Problem **??** to conclude that $p_{1,3}^*$ and $p_{1,4}^*$ (best approximation on the sets $S_{1,3}$ and $S_{1,4}$, respectively) are given by $p_{1,2}^*$ and $-1/5\, x + 4/5$. Theorem 4.5.5 guarantees that $p_{1,2}^*(x)$ is the unique best approximation to the data S_4 ($|\mu_1| < |\mu_2|$).

As indicated by the previous examples computing the best approximate to a continuous function or the best approximate on a discrete point set can be an arduous process. To highlight the next theorem let us compute the Lagrange polynomial of degree two that interpolates $f(x) = |x|$ at the nodes $\{-\sqrt{3}/2, 0, \sqrt{3}/2\}$. The result is $L_2(x) = \frac{2\sqrt{3}}{3}\, x^2$. Let us compute

$$\begin{aligned} \delta_1(|x|) &= \max_{-1 \le x \le 1} \big| |x| - L_2(x) \big| \\ &= \max_{-1 \le x \le 1} \Big| |x| - \frac{2\sqrt{3}}{3}\, x^2 \Big|\,. \end{aligned}$$

By considering the function $g(x) = x - \frac{2\sqrt{3}}{3}\, x$ on $[0, 1]$ and some elementary calculus one finds that

$$\max_{-1 \le x \le 1} |g(x)| = g\left(\frac{3}{4\sqrt{3}}\right) = \frac{\sqrt{3}}{8}$$

Figure 4.5.3: Best Linear Approximation to S_4.

so that (by symmetry)

$$\delta_1(|x|) = \frac{\sqrt{3}}{8} \cong .2165\,.$$

Recall, from Example 4.5.3, that $E_1(|x|) = 1/8 = .125$ (error in the best linear approximate to $|x|$). Hence the error in the Lagrange interpolant (on the "special" points of this example) is less than twice the error of the best approximate. The important point here is that the Lagrange interpolant is much easier to compute than is the best approximate. That this situation persists for larger degrees is the content of the following theorem.

Theorem 4.5.8 *Let $f \in C[-1,1]$ and put*

$$E_n(f) = \max_{-1\le x\le 1} |f(x) - b_n(x)| \le \max_{-1\le x\le 1} |f(x) - p_n(x)|$$

for all $p_n \in \Pi_n$. Denote by

$$x_j = \cos\left(\frac{(2j-1)\pi}{2n+2}\right) \quad , \quad j = 0, 1, \ldots, n+1$$

the zeros of $T_{n+1}(x) = \cos((n+1)\cos^{-1}(x))$ (see ??) and $L_n(x)$ the Lagrange polynomial satisfying

$$L_n(x_j) = f(x_j) \quad , \quad j = 1, \ldots, n+1\,.$$

Finally, put

$$\delta_n(f) = \max_{-1\leq x\leq 1} |f(x) - L_n(x)| .$$

Then

$$\delta_n(f) \leq 4\, E_n(f) \ , \quad 1 \leq n \leq 20$$

and

$$\delta_n(f) \leq 5\, E_n(f) \ , \quad 21 \leq n \leq 100 .$$

Example 4.5.9 If $f(x)$ is continuous and twice differentiable in $[a,b]$, and if $f''(x) > 0$ on $[a,b]$ then we can find the polynomial in b_n that satisfies equation (4.5.3) with $n = 1$, for all $b_n(x)$ in Π_n. Setting $b_1(x) = a_0 x + a_1$ where

$$\begin{aligned} a_0 &= \frac{f(b) - f(a)}{b - a} \\ a_1 &= \frac{f(c) - f(a)}{2} - \frac{c + a}{2}\, a_0 \end{aligned}$$

where c is determined such that $f'(a) = a_0$. Writing

$$r(x) = f(x) - a_0 x - a_1$$

we have

$$r'(x) = f'(x) - a_0$$

and $r'(x)$ is strictly increasing, since $f''(x) > 0$. There is a unique point c in $[a,b]$ such that $r'(c) = 0$. Thus $r(a) = -r(c) = r(b)$, i.e. b_1 is the best linear approximate of f.

Example 4.5.10 Symmetry can sometimes be used as an aid in constructing polynomials of best approximation. Consider for example, the approximation of $\cosh x$ in $[-1,1]$ by a polynomial of degree ≤ 1.

¿From Figure 4.5.4 the best approximant of degree ≤ 1 is a polynomial of degree zero, i.e., the constant

$$c = \frac{1 + \cosh 1}{2} \cong 1.22$$

For, the deviations at $x = -1, 0, 1$ must be equal and alternating, i.e.

$$\cosh 1 - c = -[\cosh 0 - c] \quad (\cong .38)$$

which yields the value c given above. Notice that this result is a special case of Example 4.5.9.

The polynomial of best approximation $b_2(x)$ in Π_2 can be obtained in a similar way. By symmetry, the polynomial is even, and may be denoted by $ax^2 + b$. The difference, $r(x) = \cosh x - ax^2 - b$ in the half interval $[0,1]$ has

Figure 4.5.4: The Approximation of $\cosh x$ by a Constant

maxima or minima where $r'(x) = \sinh x - 2ax = 0$, i.e. at $x = 0$, and at x_1, where $x_1 = \sinh x_1/2a$. The greatest positive deviations in $[0, 1]$ are at $x = 0, 1$, the deviation at x_1 being negative. The equations

$$r(0) = -r(x_1) = r(1)$$

yield

$$1 - b = -[\cosh x_1 - ax_1^2 - b] = \cosh 1 - a - b$$

or

$$a \cong 54 \ , \ b \cong .995 \ , \ x_1 \cong .70$$

the greatest deviation is $1 - b \cong .005$.

It can be seen that the polynomial of degree ≤ 2 yields five extrema; the theory requires at least four. The reason for this is symmetry. The polynomial of best approximation of degree ≤ 3 is the same as the polynomial of best approximation of degree 2.

The approximation $b_0(x) = 1.22$ and $b_2(x) = .54x^2 + .995$ may be compared with the Taylor polynomial of degrees zero and 2, $q_0(x) = 1$, $q_2(x) = 1 + 1/2\, x^2$. For the latter polynomials we find that

$$\begin{aligned} \cosh x - 1 &\cong .60 \quad \text{at} \quad x = 1 \\ \cosh x - 1 - \frac{1}{2}x^2 &\cong .04 \quad \text{at} \quad x = 1\,. \end{aligned}$$

Ex. 5: The best straight line approximation of $\sinh x$ in $[-1, 1]$ is given by a line $b_1(x) = mx$ (the approximation must be odd in x). The difference,

$r(x) = \sinh x - mx$ has a minimum (Figure ???5.7) at x_1, where $r'(x_1) = \cosh x_1 - m = 0$. Consider the half–interval $[0, 1]$. Here the deviations at x_1 and 1 must be equal and opposite

$$\sinh 1 - m = -[\sinh x_1 - mx_1]$$

or

$$(1 + x_1)\cosh x_1 = \sinh 1 - \sinh x_1 \,..$$

Solving this equation, we have $x_1 = 0.49$ and $m = 1.12$. Here there are 4 extrema, at $\pm x_1$ and ± 1, and in analogy with Ex. 4, the best degree 2 polynomial approximation is the same as the best straight line approximation.

Figure 4.5.5: Approximation of $\sinh x$ by a straight line

Ex. 6: If $f(x)$ is continuous in $[a, b]$ and

$$m = \min_{x\in[a,b]} f(x)\ , \quad M = \max_{x\in[a,b]} f(x)$$

then the polynomial $b_0(x)$ in Π_0 that satisfies Theorem 4.5.2, is given by $b(x) = (m + M)/2$ and so

$$\max_{x\in[a,b]} |f(x) - b_0(x)| = 1/2(M - m)\,.$$

Example 4.5.11 The functions $\cos(\pi x/2)$, $1 - x^2$, and $1 - x^4$ all have $1/2$ for the best–approximant of degree zero (cf. Ex. 6) in $[-1, 1]$. The function $(1 - x^2)(1 + 2x^2)$ has $9/16$ for approximant of degree zero in $[-1, 1]$. In general, $b_n(x)$ ($b_n(x)$ in Π_n) is the best polynomial approximation of degree $\leq n$ to the set of functions $f_i(x)$ continuous in $[a, b]$ if the set of remainders $r_i(x) = f_i(x) - b_n(x)$ have common extrema at $n + 2$ points of $[a, b]$.

Example 4.5.12 If $T_n(x) = \cos n\theta$, where $x = \cos\theta$ then x is the best straight line approximation to $T_3(x) + x$, $T_5(x) + x$, $T_7(x) + x$, $\ldots$.

Problems

4.5.1 $*$ Write a computer algorithm to find the polynomial $p_n \in \Pi_n$ that best approximates $m > n + 1$ points in a minimax sense.

4.5.2 Show that the minimax approximation of $q_4(x) \equiv (x^2 - 1)(x^2 - 4)$ in $[-1, 1]$ by a quadratic $q_2(x)$ has the property that $r_4(x) \equiv q_4(x) - q_2(x)$ has 5 extrema of equal magnitude and alternating sign (the preceding theory requires at least 4). Sketch $r_4(x)$.

4.5.3 Approximate $q_5(x) = x^5$ in $[-1, 1]$ by a linear function, using the procedure $q_5(x)t_5(x) - (5/4)t_3(x)$ to eliminate the terms in x^5 and x^3. Show that the error $E(x) = q_5(x) - \{q_5(x) - t_5(x) - 5/4\, t_3(x)\}$ is bounded by $1/16 + (5/4)(1/4) = 3/8$ on $[-1, 1]$, and that this error bound is sharp. This procedure is called telescoping. Verify that the maximum and minimum points of $E(x)$ are at $-2^{1/4}/4$, $2^{1/4}/2$ and $-2^{1/4}/2$, $2^{1/4}/4$ respectively.

4.5.4 Approximate x^5 in $[-1, 1]$ by the linear function $5a^4x$ in the minimax sense. Show that the minimax approximation is achieved for some solution a of the equation $(a + 1)^2(4a^3 - 3a^2 + 2a - 1) = 0$. Show that the cubic on the left–hand side of this equation has one real zero $a \cong 0.6055$, and that the modulus of the maximum error of the approximation is approximately 0.326. Show also that the error curve has four extrema.