5

Interpolation and Approximation

5.1 INTRODUCTION

An important problem often encountered in scientific work is that of approximating some very "complicated" function, $f(x)$, by a "simpler" function, $p(x)$. The reader has already seen one form of this problem in calculus: if $f(a)$, $f'(a), f''(a), \ldots, f^{(k)}(a)$ are known at some point a, then the truncated Taylor's expansion,

$$p(x) = f(a) + f'(a)(x - a) + \frac{f''(a)}{2!}(x - a)^2 + \cdots + \frac{f^{(k)}(a)}{k!}(x - a)^k,$$

is a kth degree polynomial that approximates $f(x)$ for x near a. If $f^{(k+1)}(x)$ is continuous, then the error of the approximation at x is given by

$$\frac{f^{(k+1)}(\xi)}{(k + 1)!}(x - a)^{k+1}$$

where ξ lies between x and a. There are, however, several obvious drawbacks to this type of an approximation: the derivatives may not exist, calculating them may be very difficult [or even impossible if, for instance, $f(x)$ is given in tabular form instead of in terms of a formula], $f^{(k+1)}(\xi)$ may become quite large and thus make the error large, etc. Even for a "nice" function, such as $f(x) = \cos(x)$ with $a = 0$ where the Taylor (Maclaurin) expansion converges to $f(x)$ for any x and $f^{(k)}(0)$ is known for any k, this type of approximation is not usually practical since many terms may be required to maintain accuracy for values of x somewhat removed from 0 (see Problem 2).

Thus we see that the problem of approximation of functions must be analyzed quite carefully to obtain practical computational procedures. In this chapter we shall be mainly concerned with polynomial interpolation, spline interpolation, and Fourier approximations. In the following chapter, we will study their use in such problems as numerical integration and numerical differentiation. For example if $f(x)$ is approximated by $p(x)$ and we wish to find a numerical approximation for $\int_a^b f(x)\,dx$ or for $f'(\alpha)$ for some value α, then we could use $\int_a^b p(x)\,dx$ or $p'(\alpha)$, respectively, as these approximations. However, it may be that while $\int_a^b p(x)\,dx$ is a good approximation for $\int_a^b f(x)\,dx$, $p'(\alpha)$ is a poor approximation for $f'(\alpha)$ (and vice versa). Therefore an integral part of designing an approximation for a function is the knowledge of how the approximation is to be used subsequently. Thus Chapters 5 and 6 are intrinsically related in that Chapter 5 develops some approximation techniques and the main part of Chapter 6 discusses their utilization in integration and differentiation. Then with this background, Chapter 6 concludes by discussing some further approximation techniques.

We shall be principally interested in approximating continuous real-valued functions, $f(x)$, where x belongs to the closed finite interval $[a, b]$. We use the standard notation $C[a, b]$ to denote the set of real-valued functions that are continuous on $[a, b]$. If $p(x)$ is an approximation to $f(x)$ where $f \in C[a, b]$, then we need some way of measuring how good this approximation is. That is, we must be able to answer this question: How "close" is $p(x)$ to $f(x)$? We will naturally say that $p(x)$ is a good approximation to $f(x)$ if the function $f(x) - p(x)$ is small in some sense. So to measure closeness we need some sort of a measure of size for functions in $C[a, b]$. In (5.1), we have listed three commonly used measures for the size of the function $(f - p)$. These measures are called *norms*, and the magnitude of the number $\|f - p\|$ provides us with a quantitative way of gauging how good an approximation $p(x)$ is to $f(x)$. [The reader who has covered the material on vector norms in Chapter 2 will recognize the norms defined in (5.1) as natural extensions of the ℓ_p vector norms where $(f - p)$ is used in place of the vector $(x - y)$ and integration is used in place of summation.]

$$\|f - p\|_1 \equiv \int_a^b |f(x) - p(x)|\,w(x)\,dx \tag{5.1a}$$

$$\|f - p\|_2 \equiv \left(\int_a^b (f(x) - p(x))^2 w(x)\,dx\right)^{1/2} \tag{5.1b}$$

$$\|f - p\|_\infty \equiv \max_{a \le x \le b} |f(x) - p(x)|. \tag{5.1c}$$

In (5.1a) and (5.1b) the function $w(x)$ is a fixed weighting function that provides us with some flexibility in measuring closeness. In all the cases we consider, $w(x)$ is continuous and nonnegative on (a, b), $\int_a^b w(x)\,dx$ exists, and $\int_a^b w(x)\,dx > 0$. Finally, by way of notation, we shall denote the "zero function" as $\theta(x)$ where $\theta(x) \equiv 0$ for all x in $[a, b]$; and we will let $\mathcal{P}_n$ denote the set of all polynomials of degree n or less. The following example illustrates that two functions can be "close" in one norm but *not* in another.

EXAMPLE 5.1. Let $f(x) = \theta(x)$ and let $w(x) \equiv 1$ for x in $[a, b]$ with $a = 0$ and $b = 3$. For any positive integer k, let $f_k(x)$ be given by (see Fig. 5.1)

$$f_k(x) = \begin{cases} k(k^2x - 1), & \text{for} \quad 1/k^2 \le x \le 2/k^2 \\ -k(k^2x - 3), & \text{for} \quad 2/k^2 \le x \le 3/k^2 \\ 0, & \text{otherwise.} \end{cases}$$

Using these three formulas, we obtain

$$\|f - f_k\|_1 = 1/k, \qquad \|f - f_k\|_2 = \sqrt{2}/\sqrt{3}, \qquad \|f - f_k\|_\infty = k.$$

Thus, in the sense of (5.1a), the distance between $f(x)$ and $f_k(x)$ becomes small for large k; for (5.1b) the distance is constant for any k; and for (5.1c) the distance is large for large k. Now we consider $\{\|f - f_k\|\}_{k=1}^{\infty}$ as a sequence of real numbers and see that

$$\lim_{k \to \infty} \|f - f_k\|_1 = 0,$$

$$\lim_{k \to \infty} \|f - f_k\|_2 = \sqrt{2}/\sqrt{3},$$

and

$$\lim_{k \to \infty} \|f - f_k\|_\infty = \infty.$$

In this context we therefore say that the sequence of functions, $\{f_k(x)\}_{k=1}^{\infty}$, *converges to* $f(x)$ only in terms of the $\|\cdot\|_1$ distance measurement. [A sequence of functions, $\{q_k(x)\}_{k=1}^{\infty}$, is said to converge to a function $g(x)$ with respect to a given norm, $\|\cdot\|$, if and only if

$$\lim_{k \to \infty} \|q_k - g\| = 0.]$$

The preceding example illustrates that the quality of an approximation is completely dependent on how we choose to measure distance between func-

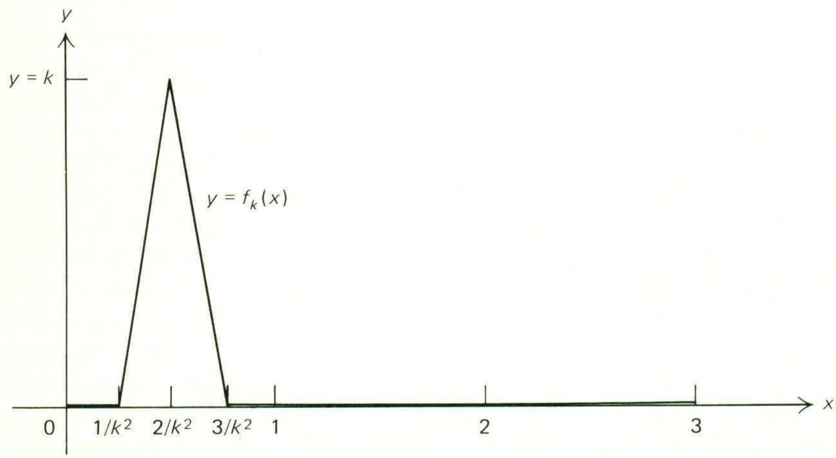

Figure 5.1 Graph of $y = f_k(x)$; $k \ge 1$.

tions. In most practical problems we would probably accept $f_k(x)$ for large k as a "good" approximation for $f(x) = \theta(x)$ since it is "bad" only in a small neighborhood of a single point. However, there are practical problems in which even moderate errors in a small neighborhood are unacceptable. For example, the cosine routine in a computer must provide *uniformly* good approximations for all x in $[0, \frac{\pi}{2}]$. The choice of a distance measurement (norm) is dependent on the underlying physical or mathematical problem. We note here that

$$\|f - p\|_1 = \int_a^b |f(x) - p(x)|\, w(x)\, dx \le \max_{a \le x \le b} |f(x) - p(x)| \int_a^b w(x)\, dx$$

$$= \|f - p\|_\infty \left(\int_a^b w(x)\, dx\right)$$

and similarly

$$\|f - p\|_2 \le \|f - p\|_\infty \left(\int_a^b w(x)\, dx\right)^{\frac{1}{2}}.$$

The number $\int_a^b w(x)\, dx$ is a constant; so if $\|f - p\|_\infty$ is "small," then both $\|f - p\|_1$ and $\|f - p\|_2$ are also "small." Thus the $\|\cdot\|_\infty$ norm is "stronger" than the other two norms, and we strive for goodness of approximation with respect to the $\|\cdot\|_\infty$ norm whenever possible. (The reader with some advanced calculus background will easily recognize that convergence of a sequence of functions in the $\|\cdot\|_\infty$ norm is equivalent to *uniform convergence*.)

As we shall see shortly, we will normally be approximating functions with polynomials. Two important reasons for this practice are that polynomials are easy to use (for example, in integration and differentiation as in Problem 1) and that polynomials can be used to provide very good approximations for functions in $C[a, b]$. The first reason is obvious to the reader and evidence for the second is provided by the following classical result (given here without proof) of Weierstrass.

Theorem 5.1 Weierstrass

Let $f \in C[a, b]$. For each $\varepsilon > 0$ there exists a polynomial $p(x)$ of degree N_ε (N_ε depends on ε) such that $\|f - p\|_\infty < \varepsilon$.

This theorem says that any continuous function on the finite interval $[a, b]$ may be uniformly approximated by some polynomial. We conclude this section with another classical result (again without proof).

Theorem 5.2

Let $f(x)$ be given in $C[a, b]$ and let n be a fixed positive integer. If $\|\cdot\|$ is any one of the three norms given above, then there exists a unique polynomial $p^*(x)$ of degree n or less such that $\|f - p^*\| \le \|f - p\|$ for all $p(x) \in \mathcal{P}_n$.

This theorem tells us that there is a unique "best" nth degree polynomial approximation to $f(x)$ with respect to any of the above norms. The reader should be warned that the polynomial that is best with respect to one norm is usually *not* the same polynomial that is best with respect to another. We shall see in Section 5.3.1 that the polynomial that minimizes $\|f - p\|_2$ for $p(x) \in \mathcal{P}_n$ can be explicitly constructed whereas this is not usually true with respect to the other two norms. However, Theorem 5.2 still has important theoretical implications with respect to these norms. Finally, let us define $E_n(f) \equiv \|f - p_n^*\|_\infty$ where $p_n^*(x) \in \mathcal{P}_n$ satisfies $\|f - p_n^*\|_\infty \leq \|f - p\|_\infty$ for all $p(x) \in \mathcal{P}_n$. [Usually, $E_n(f)$ is called the *degree of approximation for* $f(x)$ and $p_n^*(x)$ is called the *best nth degree uniform approximation to* $f(x)$.] It is easily seen that if $m > n$, then $E_m(f) \leq E_n(f)$. So, by Theorem 5.1, we obtain the result that $\lim_{n\to\infty} E_n(f) = 0$ for any $f(x)$ in $C[a, b]$.

PROBLEMS, SECTION 5.1

1. One argument for using polynomials $p(x)$ to approximate complicated functions $f(x)$ is that polynomials are easy to integrate, evaluate, and differentiate on the computer. Write a program that accepts the coefficients of any 20th-degree or less polynomial $p(x)$ as input, together with either an interval $[a, b]$ or a number c. Develop the program to have the capability of calculating $\int_a^b p(x)\,dx$ and the ith derivative of $p(x)$ evaluated at $x = c$, for any i, $0 \leq i \leq 20$.
2. Find the truncated kth degree Taylor's series expansion $p_k(x)$ for $f(x) = \cos(x)$ with $a = 0$ and k arbitrary. Show that a bound for the error $|f(x) - p_k(x)|$ is given by $|x^{k+1}|/(k + 1)!$. Given this bound, how large must k be in order that $|p_k(x) - f(x)| \leq 10^{-6}$ for $x = 1$, for $x = 2$, and for $x = 3$? For $k = 6$, 8, and 10, write a short program that calculates and lists $p_k(x)$, $\cos(x)$, and $\cos(x) - p_k(x)$ for x varying between 0 and 3 in steps of 0.1.
3. Verify that the function $\|\cdot\|_\infty$ defined on $C[a, b]$ by (5.1c) satisfies the following three conditions (where f and g are in $C[a, b]$):

   a) $\|g\|_\infty \geq 0$ and $\|g\|_\infty = 0$ if and only if $g(x) \equiv \theta(x)$

   b) $\|\alpha g\|_\infty = |\alpha|\,\|g\|_\infty$ for any scalar α

   c) $\|f + g\|_\infty \leq \|f\|_\infty + \|g\|_\infty$.

   To do this problem, recall that a continuous function defined on $[a, b]$ attains its maximum at some point x_0 in $[a, b]$.
4. The *sine-integral*, $\mathrm{Si}(x)$, occurs frequently in certain applied problems where $x > 0$ and $\mathrm{Si}(x)$ is defined by

$$\mathrm{Si}(x) = \int_0^x \frac{\sin(t)}{t}\,dt.$$

One way to estimate Si(x) is to take the truncated kth-degree Taylor's series expansion $p_k(t)$, for $f(t) = \sin(t)$ with $a = 0$, and use

$$\int_0^x \frac{p_k(t)}{t}\,dt$$

as an approximation to Si(x). How large must k be in order that

$$\left|\int_0^4 \frac{p_k(t)}{t}\,dt - \mathrm{Si}(4)\right| \le 10^{-6}?$$

To bound the error, use the fact that if $|h(t)| \le |q(t)|$ for $0 \le t \le x$, then $\int_0^x |h(t)|\,dt \le \int_0^x |q(t)|\,dt$.

5.2 POLYNOMIAL INTERPOLATION

Perhaps the simplest and best known way to construct an nth-degree polynomial approximation $p(x)$ to a function $f(x)$ in $C[a, b]$ is by interpolation. Let x_0, $x_1, \ldots, x_n$ be $(n + 1)$ distinct points in the interval $[a, b]$. Then $p(x) \in \mathcal{P}_n$ is said to *interpolate* $f(x)$ at these points if $p(x_j) = f(x_j)$ for $0 \le j \le n$. For example, the second-degree polynomial $p(x) = -(4/\pi^2)x^2 + (4/\pi)x$ interpolates $f(x) = \sin(x)$ at the points $x_0 = 0$, $x_1 = \pi/2$ and $x_2 = \pi$. We must first show that such an interpolating polynomial always exists and is unique. In doing so we shall present two separate proofs since they both illustrate ways of constructing the interpolating polynomial.

Theorem 5.3

Let $\{x_j\}_{j=0}^n$ be $(n + 1)$ distinct points in the interval $[a, b]$, and let $\{y_j\}_{j=0}^n$ be any set of $(n + 1)$ real numbers. Then there exists a unique polynomial $p(x)$ in $\mathcal{P}_n$ such that $p(x_j) = y_j$ for $0 \le j \le n$. [Often the y_j's are determined by some function, $f(x)$; so we have the property that $p(x_j) = f(x_j)$ for $0 \le j \le n$.]

Proof 1. For each j, $0 \le j \le n$, let $\ell_j(x)$ be the nth-degee polynomial defined by

$$\ell_j(x) \equiv \frac{(x - x_0)(x - x_1)\cdots(x - x_{j-1})(x - x_{j+1})\cdots(x - x_n)}{(x_j - x_0)(x_j - x_1)\cdots(x_j - x_{j-1})(x_j - x_{j+1})\cdots(x_j - x_n)}$$
$$\equiv \prod_{\substack{i=0 \\ i \ne j}}^{n} \frac{(x - x_i)}{(x_j - x_i)}. \tag{5.2}$$

Then it follows for each i and j, that $\ell_i(x_j) = \delta_{ij}$. (The symbol δ_{ij} is defined to be 1 when $i = j$ and 0 otherwise. The symbol δ_{ij} is called the Kronecker delta.)

Since the sum of nth-degree polynomials is again a polynomial of at most nth degree, the polynomial $p(x)$ defined by (5.3) is in $\mathcal{P}_n$:

$$p(x) \equiv \sum_{j=0}^{n} y_j \ell_j(x). \tag{5.3}$$

Moreover, for $0 \le i \le n$, we have

$$p(x_i) = y_0 \ell_0(x_i) + \cdots + y_n \ell_n(x_i) = y_i \ell_i(x_i) = y_i.$$

If $f(x)$ is a function such that $f(x_i) = y_i$, $0 \le i \le n$, then it has in $\mathcal{P}_n$ an interpolating polynomial of the form

$$p(x) = \sum_{j=0}^{n} f(x_j) \ell_j(x). \tag{5.4}$$

To establish uniqueness, let us assume that there are two different polynomials, $p(x)$ and $q(x)$ in $\mathcal{P}_n$, such that $p(x_j) = q(x_j) = y_j$ for $0 \le j \le n$. If $r(x) \equiv p(x) - q(x)$, then $r(x) \in \mathcal{P}_n$ and furthermore $r(x_j) = p(x_j) - q(x_j) = 0$ for $0 \le j \le n$. By the Fundamental Theorem of Algebra, $r(x) \equiv 0$; so $p(x) \equiv q(x)$, which contradicts our assumption.

Proof 2. Let $p(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n$ where the coefficients, a_j, are to be determined. Consider the $(n + 1)$ equations

$$p(x_j) = a_0 + a_1 x_j + a_2 x_j^2 + \cdots + a_n x_j^n = y_j, \qquad 0 \le j \le n. \tag{5.5}$$

In matrix form, Eqs. (5.5) become

$$\begin{bmatrix} 1 & x_0 & x_0^2 & \cdots & x_0^n \\ 1 & x_1 & x_1^2 & \cdots & x_1^n \\ \vdots & & & & \\ 1 & x_n & x_n^2 & \cdots & x_n^n \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_n \end{bmatrix} = \begin{bmatrix} y_0 \\ y_1 \\ \vdots \\ y_n \end{bmatrix} \tag{5.6}$$

or $V\mathbf{a} = \mathbf{y}$ where V is the coefficient matrix in (5.6), $\mathbf{a} = [a_0, a_1, \ldots, a_n]^T$, and $\mathbf{y} = [y_0, y_1, \ldots, y_n]^T$. The matrix V is called a *Vandermonde matrix*, and it is easy to see that V is nonsingular when $x_0, x_1, \ldots, x_n$ are distinct. To see this, recall from Section 2.1 that V is nonsingular if and only if θ is the only solution of $V\mathbf{a} = \theta$. So if the vector $\mathbf{a}$ is any solution of $V\mathbf{a} = \theta$, then $p(x) = a_0 + a_1 x + \cdots + a_n x^n$ is an nth-degree polynomial such that $p(x_j) = 0$ for $j = 0, 1, \ldots, n$. Since the only nth-degree polynomial with $(n + 1)$ zeros is the zero polynomial, it must be $\mathbf{a} = \theta$ and hence V is nonsingular. Thus Eqs. (5.5) have a unique solution for the a_i's, and the polynomial $p(x)$ found by solving (5.6) is the unique interpolating polynomial in $\mathcal{P}_n$. ∎

From the uniqueness of the interpolating polynomial in $\mathcal{P}_n$, both (5.4) and (5.6) must yield the same polynomial even though it is written in different forms. The form given by (5.4) is called the *Lagrange form* whereas the construction of $p(x)$ by (5.6) is called the *method of undetermined coefficients*. We should note three obvious facts here. First, if $f(x)$ is itself in $\mathcal{P}_n$, then $f(x) \equiv p(x)$ for all x by the uniqueness property. Second, it is possible that the unique solution of (5.6) may yield $a_n = 0$ (and possibly other coefficients may be zero also); so $p(x)$ may be a polynomial of degree strictly less than n. Third, if $m > n$, there is an infinite number of polynomials, $q(x)$, in $\mathcal{P}_m$ which satisfy $q(x_j) = y_j$, $0 \le j \le n$. [Note that we can arbitrarily specify another point, x_{n+1}, in $[a, b]$ and any other value y_{n+1}, and then construct $q(x) \in \mathcal{P}_{n+1}$ such that $q(x_j) = y_j$, $0 \le j \le n + 1$. Thus the uniqueness holds only for $\mathcal{P}_n$.]

EXAMPLE 5.2 As a simple example of a problem involving data fitting, suppose we want a second-degree polynomial, $p(x)$, such that $p(0) = -1$, $p(1) = 2$, and $p(2) = 7$. Using the Lagrange form, we have

$$\ell_0(x) = \frac{(x-1)(x-2)}{2}, \qquad \ell_1(x) = -x(x-2), \qquad \text{and} \qquad \ell_2(x) = \frac{x(x-1)}{2}.$$

Thus $p(x)$ is given by the formula $p(x) = -\ell_0(x) + 2\ell_1(x) + 7\ell_2(x)$, or upon simplification $p(x) = x^2 + 2x - 1$. Alternatively, using the method of undetermined coefficients with $x_0 = 0$, $x_1 = 1$ and $x_2 = 2$ in (5.5), we obtain

$$\begin{aligned} a_0 \phantom{{}+2a_1+4a_2} &= -1 \\ a_0 + a_1 + a_2 &= 2 \\ a_0 + 2a_1 + 4a_2 &= 7. \end{aligned}$$

Solving this system gives $p(x) = x^2 + 2x - 1$.

The method of undetermined coefficients is a procedure that has wide application to other types of interpolation problems. For example, let $x_0 = 0$, $x_1 = \pi/6$, $x_2 = \pi/4$, and $x_3 = \pi/3$ with $y_0 = 0$, $y_1 = 2 - \sqrt{3}/2$, $y_2 = 1 + \sqrt{2}/2$, and $y_3 = 3/2$. Suppose we wish to find a *trigonometric* polynomial, $p(x)$, of the form $p(x) = C_0 + C_1 \cos(x) + C_2 \cos(2x) + C_3 \cos(3x)$ such that $p(x_j) = y_j$ for $0 \le j \le 3$. These four equations yield the matrix equation [as in the derivation of (5.6)]

$$\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & \sqrt{3}/2 & 1/2 & 0 \\ 1 & \sqrt{2}/2 & 0 & -\sqrt{2}/2 \\ 1 & 1/2 & -1/2 & -1 \end{bmatrix} \begin{bmatrix} C_0 \\ C_1 \\ C_2 \\ C_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 2 - \sqrt{3}/2 \\ 1 + \sqrt{2}/2 \\ 3/2 \end{bmatrix}. \tag{5.7}$$

The solution of this system by Gauss elimination yields $C_0 = 1$, $C_1 = -1$, $C_2 = 2$, $C_3 = -2$. Therefore $p(x) = 1 - \cos(x) + 2\cos(2x) - 2\cos(3x)$ is a trigonometric polynomial satisfying $p(x_i) = y_i$, $i = 0, 1, 2, 3$.

5.2.1. Divided Differences and the Newton Form of the Interpolating Polynomial

If $f(x)$ is a function defined on $[a, b]$ and if $x_0, x_1, \ldots, x_n$ are distinct points in $[a, b]$, then there is a unique polynomial $p(x)$ in $\mathcal{P}_n$ that satisfies $p(x_i) = f(x_i)$, $0 \le i \le n$. There are many ways to represent this interpolating polynomial $p(x)$, but some representations are more useful for computation than others (in mathematical terms, we would like to use a basis for $\mathcal{P}_n$ that is convenient with respect to computation). For example, although we can immediately express $p(x)$ in the Lagrange form

$$p(x) = c_0\ell_0(x) + c_1\ell_1(x) + \cdots + c_n\ell_n(x)$$

simply by choosing $c_i = f(x_i)$, $0 \le i \le n$, this form is cumbersome to use in common operations such as differentiation and integration. The most convenient form for these operations is usually $p(x) = b_0 + b_1x + \cdots + b_nx^n$; but here, to find the values of the b_i's we must go through the effort of solving the system in (5.6) or collecting coefficients of like powers of x from another form of $p(x)$ such as the Lagrange form. Hence, the manner in which $p(x)$ is to be used is an important factor in the choice of the form of $p(x)$. In this section we consider yet another representation for $p(x)$, the *Newton form*, which facilitates computations with $p(x)$ when the interpolation points are equally spaced or when we wish to add further interpolation points to get a higher-degree approximation.

The Newton form for $p(x)$ arises when we express the interpolating polynomial $p(x)$ in the form

$$p(x) = a_0 + a_1(x - x_0) + a_2(x - x_0)(x - x_1) + a_3(x - x_0)(x - x_1)(x - x_2) + \cdots + a_n(x - x_0)(x - x_1)(x - x_2)\cdots(x - x_{n-1}). \tag{5.8}$$

If we impose the interpolatory constraints $p(x_0) = f(x_0)$, $p(x_1) = f(x_1)$, . . . , $p(x_n) = f(x_n)$, then it is clear that we can solve for the coefficients a_0, a_1, . . . , a_n in (5.8). For example, setting $x = x_0$ in (5.8), we have $f(x_0) = a_0$. Next setting $x = x_1$, we have $f(x_1) = a_0 + a_1(x_1 - x_0)$ or $a_1 = [f(x_1) - f(x_0)]/(x_1 - x_0)$. For $x = x_2$, we have $f(x_2) = a_0 + a_1(x_2 - x_0) + a_2(x_2 - x_0)(x_2 - x_1)$; and it is not hard to see that this leads to

$$a_2 = \frac{\dfrac{f(x_2) - f(x_1)}{x_2 - x_1} - \dfrac{f(x_1) - f(x_0)}{x_1 - x_0}}{x_2 - x_0}.$$

Clearly, determining the coefficients a_i in (5.8) is recursive in nature and we can mechanize the procedure (in a way that is suited for computation) by introducing the related ideas of divided differences and the divided difference table.

Given x_0, x_1, . . . , x_n in $[a, b]$, we define the *first divided difference*, $f[x_i, x_{i+1}]$, by

$$f[x_i, x_{i+1}] = \frac{f(x_{i+1}) - f(x_i)}{x_{i+1} - x_i}, \qquad 0 \le i \le n - 1.$$

We define the kth divided difference inductively by

$$f[x_i, x_{i+1}, \ldots, x_{i+k}] = \frac{f[x_{i+1}, x_{i+2}, \ldots, x_{i+k}] - f[x_i, x_{i+1}, \ldots, x_{i+k-1}]}{x_{i+k} - x_i} \tag{5.9}$$

where $0 \le i \le n - k$. For example,

$$f[x_2, x_3] = \frac{f(x_3) - f(x_2)}{x_3 - x_2}, \quad f[x_3, x_4, x_5] = \frac{f[x_4, x_5] - f[x_3, x_4]}{x_5 - x_3},$$

$$f[x_1, x_2, x_3, x_4] = \frac{f[x_2, x_3, x_4] - f[x_1, x_2, x_3]}{x_4 - x_1}.$$

We will show shortly that the coefficients a_i in (5.8) are given precisely by $a_i = f[x_0, x_1, \ldots, x_i]$. (Note that as found above, $a_1 = f[x_0, x_1]$ and $a_2 = f[x_0, x_1, x_2]$.)

The calculation of divided differences can be organized by constructing a *divided difference table*, (see Table 5.1). Table 5.1 presents the definition (5.9) in a rather schematic form; entries in the kth column are obtained by "differencing" successive entries in the $(k-1)$st column and then dividing by the appropriate difference $x_{i+k} - x_i$

TABLE 5.1 A divided difference table.

x_0	$f(x_0)$				
		$f[x_0, x_1]$			
x_1	$f(x_1)$		$f[x_0, x_1, x_2]$		
		$f[x_1, x_2]$		$f[x_0, x_1, x_2, x_3]$	
x_2	$f(x_2)$		$f[x_1, x_2, x_3]$		$f[x_0, x_1, x_2, x_3, x_4]$
		$f[x_2, x_3]$		$f[x_1, x_2, x_3, x_4]$	
x_3	$f(x_3)$		$f[x_2, x_3, x_4]$		.
		$f[x_3, x_4]$		.	.
x_4	$f(x_4)$		.	.	.
.	.	.	.	.	
.	.	.	.		
.	.	.			
			$f[x_{n-2}, x_{n-1}, x_n]$		
		$f[x_{n-1}, x_n]$			
x_n	$f(x_n)$				

EXAMPLE 5.3. For $f(x) = x^5 - x^4 + 2x^2 + 1$, we construct the divided difference table for the data

x	$f(x)$
−2	−39
−1	1
0	1
1	3
2	25
3	181
4	801

The divided difference table is

−2	−39						
		40					
−1	1		−20				
		0		7			
0	1		1		−1		
		2		3		1	
1	3		10		4		0
		22		19		1	
2	25		67		9		
		156		55			
3	181		232				
		620					
4	801						

For instance, $x_3 = 1$, $x_4 = 2$, $x_5 = 3$; so $f[x_3, x_4] = 22$, $f[x_4, x_5] = 156$, and $f[x_3, x_4, x_5] = 67$. To display the recursive nature of (5.9), we consider the entries along the upper diagonal, beginning with $f(x_0) = -39$. The next term along the diagonal is $f[x_0, x_1] = (1 - (-39))/1 = 40$, followed by $f[x_0, x_1, x_2] = (0 - 40)/2 = -20$; then $f[x_0, x_1, x_2, x_3] = (1 - (-20))/3 = 7$; then $f[x_0, x_1, x_2, x_3, x_4] = (3 - 7)/4 = -1$, etc.

A divided difference table contains a good deal of information that is relevant to interpolation. The downward slanting diagonals contain the coefficients of various interpolating polynomials, and we will see that the columns contain information that can be used to estimate interpolation errors. Our first objective is to find the coefficients a_i in (5.8), and to that end we state Theorem 5.4 below.

Theorem 5.4

Suppose that $f(x)$ is defined on $[a, b]$ and suppose that $x_0, x_1, \ldots, x_n$ are distinct points in $[a, b]$. The kth degree polynomial $p(x)$ interpolating $f(x)$ at $x_i, x_{i+1}, \ldots, x_{i+k}$ is given by

$$p(x) = f(x_i) + f[x_i, x_{i+1}](x - x_i) + f[x_i, x_{i+1}, x_{i+2}](x - x_i)(x - x_{i+1}) + \cdots + f[x_i, x_{i+1}, \ldots, x_{i+k}](x - x_i)(x - x_{i+1}) \ldots (x - x_{i+k-1}).$$

Since the notation is somewhat involved, we give several illustrations of the theorem before sketching the proof. For example, according to Theorem 5.4, the quadratic polynomial interpolating $f(x)$ at x_2, x_3, x_4 is given by $p(x) = f(x_2) + f[x_2, x_3](x - x_2) + f[x_2, x_3, x_4](x - x_2)(x - x_3)$; and we observe that the coefficients $f(x_2)$, $f[x_2, x_3]$, $f[x_2, x_3, x_4]$ are found in Table 5.1 on the downward diagonal that begins at $f(x_2)$. As another special case, the cubic polynomial interpolating $f(x)$ at x_0, x_1, x_2, x_3 is (according to Theorem 5.4) given by

$$p(x) = f(x_0) + f[x_0, x_1](x - x_0) + f[x_0, x_1, x_2](x - x_0)(x - x_1) + f[x_0, x_1, x_2, x_3](x - x_0)(x - x_1)(x - x_2).$$

Again, we note that the coefficients of $p(x)$ are found on the downward diagonal that starts at $f(x_0)$ (see Table 5.1).

Proof. The proof of Theorem 5.4 is by induction on k. The proof is simple and we sketch only a part of it here; the rest is left to the exercises. For $k = 1$, Theorem 5.4 asserts that $p(x) = f(x_i) + f[x_i, x_{i+1}](x - x_i)$ is the first-degree polynomial interpolating $f(x)$ at x_i and x_{i+1}. Clearly $p(x)$ is a linear polynomial, and it is also obvious that $p(x_i) = f(x_i)$. At $x = x_{i+1}$, we have $p(x_{i+1}) = f(x_i) + f[x_i, x_{i+1}](x_{i+1} - x_i) = f(x_i) + (f(x_{i+1}) - f(x_i))$; so $p(x_{i+1}) = f(x_{i+1})$. This argument shows that Theorem 5.4 is valid for $k = 1$.

Rather than doing the general induction step, we show that Theorem 5.4 is valid for $k = 2$, given that the theorem holds for $k = 1$ (virtually the same proof will work for general k and is left to the exercises). Let $p(x)$ be the polynomial of degree two interpolating $f(x)$ at x_i, x_{i+1}, x_{i+2}. We know we can represent $p(x)$ in the form

$$p(x) = a_0 + a_1(x - x_i) + a_2(x - x_i)(x - x_{i+1}) = g(x) + a_2(x - x_i)(x - x_{i+1}).$$

Now since $p(x_i) = f(x_i)$ and $p(x_{i+1}) = f(x_{i+1})$, it follows that $g(x_i) = f(x_i)$ and $g(x_{i+1}) = f(x_{i+1})$. Since $g(x)$ is linear and since Theorem 5.4 is valid for $k = 1$, we know that $g(x) = f(x_i) + f[x_i, x_{i+1}](x - x_i)$. From this, $p(x)$ has the form

$$p(x) = f(x_i) + f[x_i, x_{i+1}](x - x_i) + a_2(x - x_i)(x - x_{i+1}),$$

and all that remains is to show that $a_2 = f[x_i, x_{i+1}, x_{i+2}]$. To show this, let $h(x)$ be the linear polynomial that interpolates $f(x)$ at x_{i+1} and x_{i+2} so that $h(x) = f(x_{i+1}) + f[x_{i+1}, x_{i+2}](x - x_{i+1})$. Define $Q(x)$ by

$$Q(x) = \frac{(x - x_i)h(x) - (x - x_{i+2})g(x)}{x_{i+2} - x_i}.$$

Note that $Q(x)$ has degree two, and by direct substitution, $Q(x)$ interpolates $f(x)$ at x_i, x_{i+1}, x_{i+2}. Since polynomial interpolation is unique, $Q(x) \equiv p(x)$ and hence the coefficient of x^2 must be the same in both $Q(x)$ and $p(x)$. However, the coefficient of x^2 in $Q(x)$ is

$$\frac{f[x_{i+1}, x_{i+2}] - f[x_i, x_{i+1}]}{x_{i+2} - x_i};$$

so we have shown that $a_2 = f[x_i, x_{i+1}, x_{i+2}]$. ■

The Newton form of the interpolating polynomial is given explicitly in a corollary to Theorem 5.4.

Corollary

The kth degree polynomial interpolating $f(x)$ at $x_0, x_1, \ldots, x_k$ is given by

$$p(x) = f(x_0) + f[x_0, x_1](x - x_0) + f[x_0, x_1, x_2](x - x_0)(x - x_1) + \cdots$$
$$+ f[x_0, x_1, \ldots, x_k](x - x_0)(x - x_1) \ldots (x - x_{k-1}).$$

As an example, consider the function $f(x)$ from Example 5.3. According to Theorem 5.4, the cubic polynomial interpolating $f(x)$ at -2, -1, 0, 1 is

$$q(x) = -39 + 40(x + 2) - 20(x + 2)(x + 1) + 7(x + 2)(x + 1)x \quad (5.10)$$

and this fact can be verified directly. An observation that may not be obvious is that when we use the Newton form, we can add another interpolation constraint without sacrificing our previous effort. To be explicit, suppose $p_k(x)$ interpolates $f(x)$ at $(k + 1)$ points $x_0, x_1, \ldots, x_k$. To construct a polynomial $p(x)$ that interpolates $f(x)$ at $(k + 2)$ points $x_0, x_1, \ldots, x_k, x_{k+1}$, we merely add another term to $p_k(x)$:

$$p(x) = p_k(x) + f[x_0, x_1, \ldots, x_k, x_{k+1}](x - x_0)(x - x_1) \ldots (x - x_k). \quad (5.11)$$

The observation above is one computational feature that recommends the Newton form of the interpolating polynomial.

EXAMPLE 5.4. As an illustration of (5.11), we can build the polynomial $p(x)$ that interpolates $f(x)$ at -2, -1, 0, 1, 2 from $q(x)$ in (5.10):

$$p(x) = q(x) - (x + 2)(x + 1)x(x - 1).$$

A variety of different interpolating polynomials can be found from a divided difference table such as the one in Example 5.3. For instance, the cubic polynomial interpolating $f(x)$ at 0, 1, 2, 3 is given by

$$p(x) = 1 + 2x + 10x(x - 1) + 19x(x - 1)(x - 2). \quad (5.12)$$

The coefficients of $p(x)$ in (5.12) are found (using Theorem 5.4) from the downward diagonal that starts at $f(0) = 1$. Finally, it should be clear that additional data can easily be added to either end of a divided difference table. To illustrate, if we add the point $x = -3$ and $f(x) = -305$ to the head of the table in Example 5.3, we can quickly construct a new diagonal that incorporates this additional information into the existing table.

PROBLEMS, SECTION 5.2.1

1. For each set of tabulated data below, let $p(x)$ be the cubic interpolating polynomial. Using the Lagrange form of $p(x)$, evaluate $p(x)$ at $x = -2$. [Note that you need not find the coefficients of x^j in $p(x)$; you need only evaluate $\ell_j(-2)$ for $0 \le j \le 3$.]

a)

x	y
-1	-1
0	3
2	11
3	27

b)

x	y
-1	-3
0	1
1	1
1	3

2. Repeat Problem 1, but use the method of undetermined coefficients to find $p(x)$.

3. For each set of data in Problem 1, set up the divided difference table and use it to construct $p(x)$ as in the corollary to Theorem 5.4. For each interpolating polynomial $p(x)$, evaluate $p(-2)$. [Again, note that you need not write $p(x)$ in the form $a_3x^3 + a_2x^2 + a_1x + a_0$ to calculate $p(-2)$.]

4. Add the data point $x = 4$, $y = 10$ to each of the tables in Problem 1. Form the divided difference table for these data by adding an upward slanting diagonal to the table constructed in Problem 3. Form the quartic polynomial $q(x)$ that interpolates these five data points, and evaluate $q(x)$ at $x = -2$.

5. Using the divided difference table in Example 5.3, construct the polynomial interpolating $f(x)$ at

a) $x = -1, 0, 1$ b) $x = -1, 0, 1, 2$ c) $x = -1, 0, 1, 2, 3$

d) $x = 0, 1$ e) $x = 0, 1, 2, 3, 4$

6. Use the method of undetermined coefficients to find a quadratic polynomial, $p(x)$ such that $p(1) = 0$, $p'(1) = 7$, and $p(2) = 10$. [Note the derivative evaluation. You must modify (5.6) slightly.]

7. Suppose we know that $f(t)$ has the form $f(t) = ae^{2t} + be^{-t}$ where a and b are unknown. Use the method of undetermined coefficients to find a and b given that $f(t_0) = 1$ and $f(t_1) = 7.5$ where $t_0 = 0$ and $t_1 = \ln(2)$. Suppose next that $f(t)$ is to be determined by some experimental observations in which $f(t_0) = 1$, $f(t_1) = 7$, and $f(t_2) = 9$ with $t_2 = \ln(3)$. Since we have three data points and only two parameters (a and b), we do not expect to be able to find a and b by the method of undetermined coefficients. Use the procedure of Section 2.5 to find the best least-squares solution for a and b.

8. Write a subroutine for polynomial interpolation. Your subprogram should accept an integer N (N $\le$ 20), N-dimensional arrays of data X(I) and Y(I), and an evaluation point ALPHA. The subroutine should calculate the divided difference table, construct the interpolating polynomial (of degree N − 1), and return the value of the polynomial at x = ALPHA. Test your routine with the input data (x_i, y_i) where $x_i = i$ and $y_i = p(x_i)$, $0 \le i \le 4$, $p(x) = x^4 + 1$. Evaluate at $\alpha = -1 + k/2$, $k = 0, 1, \ldots, 12$; and note that $p(\alpha)$ should equal $\alpha^4 + 1$ for any α.

9. To see the sorts of results that can be expected, use the program in Problem 8 to interpolate $f(x) = \sin(x)$ at the knots $x_k = .1 + k/10$, $0 \le k \le 9$. If $p(x)$ denotes the ninth degree interpolating polynomial for $f(x)$, print $f(\alpha)$, $p(\alpha)$, and the relative error $(f(\alpha) - p(\alpha))/f(\alpha)$ for $\alpha = .05 + k/10$, $0 \le k \le 11$.

10. Repeat Problem 9 for $f(x) = 1/(1 + x^2)$ and the knots $x_k = -5 + k$, $0 \le k \le 10$ and $\alpha = -5.5 + k$, $0 \le k \le 12$. Also call a plotting routine and plot $f(x)$ and $p(x)$ over