3 Polynomial Interpolation

**Definition 3.1.** Interpolation constructs new data points within the range of a discrete set of known data points, usually by generating an interpolating function whose graph goes through all known data points.

**Example 3.1.** The interpolating function may be piecewise constant, piecewise linear, polynomial, spline, or other non-polynomial functions.

### 3.1 The Vandermonde determinant

**Definition 3.2.** For \( n + 1 \) given points \( x_0, x_1, \ldots, x_n \in \mathbb{R} \), the associated Vandermonde matrix \( V \in \mathbb{R}^{(n+1) \times (n+1)} \) is

\[
V(x_0, x_1, \ldots, x_n) = \begin{pmatrix}
1 & x_0 & x_0^2 & \cdots & x_0^n \\
1 & x_1 & x_1^2 & \cdots & x_1^n \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
1 & x_n & x_n^2 & \cdots & x_n^n 
\end{pmatrix}.
\]

**Lemma 3.3.** The determinant of a Vandermonde matrix can be expressed as

\[
\det V(x_0, x_1, \ldots, x_n) = \prod_{i>j}(x_i - x_j). \quad (3.2)
\]

**Proof.** Consider the function

\[
U(x) = \det V(x_0, x_1, \ldots, x_n, x) = \begin{vmatrix}
1 & x_0 & x_0^2 & \cdots & x_0^n \\
1 & x_1 & x_1^2 & \cdots & x_1^n \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
1 & x_n & x_n^2 & \cdots & x_n^n \\
1 & x & x^2 & \cdots & x^n 
\end{vmatrix}. \quad (3.3)
\]

Clearly, \( U(x) \in \mathbb{P}_n \) and it vanishes at \( x_0, x_1, \ldots, x_n \) since inserting these values in place of \( x \) yields two identical rows in the determinant. It follows that

\[
U(x_0, x_1, \ldots, x_n, x) = A \prod_{i=0}^{n-1} (x - x_i),
\]

where \( A \) depends only on \( x_0, x_1, \ldots, x_n \). Meanwhile, the expansion of \( U(x) \) in \( (3.3) \) by minors of its last row implies that the coefficient of \( x^n \) is \( U(x_0, x_1, \ldots, x_n) \). Hence we have

\[
U(x_0, x_1, \ldots, x_n, x) = U(x_0, x_1, \ldots, x_n) \prod_{i=0}^{n-1} (x - x_i),
\]

and consequently the recursion

\[
U(x_0, x_1, \ldots, x_n, x) = U(x_0, x_1, \ldots, x_n) \prod_{i=0}^{n-1} (x - x_i).
\]

An induction based on \( U(x_0, x_1) = x_1 - x_0 \) yields \( (3.2) \). \qed

**Remark 3.2** (Alternative proof of Lemma 3.3). For any \( i = 1, 2, \ldots, n \) and \( j \leq i \), replace \( x_i \) with \( x_j \) in the Vandermonde matrix \( (3.1) \) and we have \( \det V = 0 \). The total number of these replacements is \( 1 + 2 + \cdots + n = \frac{n(n+1)}{2} \). On the other hand, \( \det V \) can be considered as a polynomial in the variables \( x_0, x_1, \ldots, x_n \). Hence \( \prod_{i>j}(x_i - x_j) \) is a factor of \( \det V \). By definition of determinants, the total degree of each polynomial term in \( \det V \) is also \( 1 + 2 + \cdots + n = \frac{n(n+1)}{2} \). Then the proof is completed by noticing that the term \( x_1x_2^2 \cdots x^n \) has coefficient 1.

**Theorem 3.4.** Given distinct points \( x_0, x_1, \ldots, x_n \in \mathbb{C} \) and corresponding values \( f_0, f_1, \ldots, f_n \in \mathbb{C} \). Denote by \( \mathbb{P}_n \) the class of polynomials of degree at most \( n \). There exists a unique polynomial \( p_n(x) \in \mathbb{P}_n \) such that

\[
\forall i = 0, 1, \ldots, n, \quad p_n(x_i) = f_i. \quad (3.4)
\]

**Proof.** Set up a polynomial \( \sum_{i=0}^n a_ix^i \) with \( n + 1 \) undetermined coefficients \( a_i \). The condition \( (3.4) \) leads to the system of \( n + 1 \) equations:

\[
a_0 + a_1x_1 + a_2x_1^2 + \cdots + a_nx_1^n = f_1,
\]

where \( i = 0, 1, \ldots, n \). By Lemma 3.3, the determinant of the system is \( \prod_{i>j}(x_i - x_j) \). The proof is completed by the distinctness of the points and Cramer’s rule. \( \square \)

### 3.2 The Cauchy Remainder

Suppose the values \( f_i \)’s come from a given function \( f \) as \( f_i = f(x_i) \). Apart from the fact that \( p_n(x) \) agrees with \( f(x) \) on the given points \( x_i \), how much would \( p_n(x) \) differ from \( f(x) \) for \( x \neq x_i \)?

**Theorem 3.5** (Generalized Rolle). Let \( n \geq 2 \). Suppose that \( f \in \mathcal{C}^n[a, b] \) and \( f^{(n)}(x) \) exists at each point of \( (a, b) \). Suppose that \( f(x_0) = f(x_1) = \cdots = f(x_n) = 0 \) for \( a \leq x_0 < x_1 < \cdots < x_n \leq b \). Then there is a point \( \xi \in (x_0, x_n) \) such that \( f^{(n)}(\xi) = 0 \).

**Proof.** Applying Rolle’s theorem T0.32 on the \( n-1 \) intervals \( (x_i, x_{i+1}) \) yields \( n-1 \) points \( \zeta_i \) where \( f^{(i)}(\zeta_i) = 0 \). Consider \( f^{(i)}, f^{(i+1)}, \ldots, f^{(n-2)} \) as new functions. Repeatedly applying the above arguments completes the proof. \( \square \)

**Theorem 3.6** (Cauchy remainder of polynomial interpolation). Let \( f \in \mathcal{C}^n[a, b] \) and suppose that \( f^{(n+1)}(x) \) exists at each point of \( (a, b) \). Let \( p_n(f; x) \) denote the unique polynomial in \( \mathbb{P}_n \) that coincides with \( f \) at \( x_0, x_1, \ldots, x_n \). Define

\[
R_n(f; x) := f(x) - p_n(f; x) \quad (3.5)
\]

as the Cauchy remainder of the polynomial interpolation. If \( a \leq x_0 < x_1 < \cdots < x_n \leq b \), then there exists some \( \xi \in (a, b) \) such that

\[
R_n(f; x) = \frac{f^{(n+1)}(\xi)}{(n+1)!} \prod_{i=0}^{n}(x - x_i) \quad (3.6)
\]

where the value of \( \xi \) depends on \( x, x_0, x_1, \ldots, x_n \), and \( f \).
Proof. Since \( f(x_k) = p_n(f; x_k) \), the remainder \( R_n(f; x) \) vanished at \( x_k \)'s. Fix \( x \neq x_0, x_1, \ldots, x_n \) and define
\[
K(x) = \frac{f(x) - p_n(f; x)}{\prod_{i=0}^{n}(x-x_i)},
\]
and a function of \( t \)
\[
W(t) = f(t) - p_n(f; t) - K(x) \prod_{i=0}^{n}(t-x_i).
\]
The function \( W(t) \) vanishes at \( t = x_0, x_1, \ldots, x_n \). In addition \( W(x) = 0 \). By Theorem 3.5, \( W^{(n+1)}(x) = 0 \) for some \( \xi \in (a, b) \), i.e.
\[
0 = W^{(n+1)}(x) = f^{(n+1)}(x) - (n + 1)K(x).
\]
Hence \( K(x) = f^{(n+1)}(x)/(n + 1)! \) and \((3.6)\) holds. \( \square \)

**Proposition 3.9.** Define a symmetric polynomial
\[
\pi_n(x) = \begin{cases} 1, & n = 0; \\ \prod_{i=0}^{n}(x-x_i), & n > 0. \end{cases} \quad (3.10)
\]
Then the fundamental polynomial for pointwise interpolation can be expressed as
\[
\forall x \neq x_k, \quad \ell_k(x) = \frac{\pi_n(x)}{(x-x_k)^{n}(x-x_k)\pi_n(x)}.
\]

Proof. By the chain rule, \( \pi_n(x) \) is the summation of \( n + 1 \) terms, each of which is a product of \( n \) terms. When \( x \) is replaced with \( x_k \), all of the \( n + 1 \) terms vanish except one. \( \square \)

**Remark 3.6.** Theorems 3.4 and 3.6 dictate that the polynomial interpolation of degree \( k \) of a polynomial \( q \) of degree no greater than \( k \) is \( q \) itself.

**Proposition 3.10.** The fundamental polynomials \( \ell_k(x) \) satisfy the Cauchy relations as follows.
\[
\sum_{k=0}^{n} \ell_k(x) \equiv 1 \quad (3.12)
\]
\[
\forall j = 1, \ldots, n, \quad \sum_{k=0}^{n} (x_k - x)\ell_k(x) \equiv 0 \quad (3.13)
\]

Proof. By Theorems 3.4 and 3.6, for each \( q(x) \in \mathbb{P}_n \) we have \( p_n(q; x) \equiv q(x) \). Interpolating the constant function \( f(x) \equiv 1 \) with the Lagrange formula yields \((3.12)\). Similarly, \((3.13)\) can be proved by interpolating the polynomial \( q(u) = (u - x)^j \) for each \( j = 1, \ldots, n \) with the Lagrange formula. \( \square \)

**Remark 3.7** (An alternative proof of the Cauchy relation). Suppose that \((3.13)\) does not hold for some \( x = x_c \). Then interpolating the polynomial \( q(x) = (x-u)^j \) for each \( j = 1, \ldots, n \) with the Lagrange formula yields
\[
(x-u)^j = p_n(q; x) = \sum_{k=0}^{n} q_k\ell_k(x) = \sum_{k=0}^{n} (x_k - u)\ell_k(x).
\]
Setting \( u = x_c \) leads to a contradiction.

### 3.3 The Lagrange formula

**Remark 3.4.** The proof of Theorem 3.4 gives one way to construct an interpolating polynomial. However, it requires solving a linear system. Can we construct interpolation polynomials explicitly? The answer is yes and the Lagrange formula is such a well-known one.

**Definition 3.8.** To interpolate given values \( f_0, f_1, \ldots, f_n \) at distinct points \( x_0, x_1, \ldots, x_n \), the **Lagrangian formula** is
\[
p_n(x) = \sum_{k=0}^{n} f_k\ell_k(x), \quad (3.8)
\]
where the fundamental polynomials for pointwise interpolation (or elementary Lagrange interpolation polynomials) \( \ell_k(x) \) is
\[
\ell_k(x) = \prod_{i \neq k}^{n} \frac{x-x_i}{x_k-x_i}, \quad (3.9)
\]
In particular, for \( n = 0, \ell_0 = 1 \).

**Example 3.5.** For \( i = 0, 1, 2 \), we are given \( x_i = 1, 2, 4 \) and \( f(x_i) = 8, 1, 5 \), respectively. The Lagrangian formula generates \( p_2(x) = 3x^2 - 16x + 21 \).

### 3.4 The Newton formula

**Remark 3.8.** The Lagrange formula has one drawback. If we desire to pass from degree \( n \) to degree \( n + 1 \), we must determine an entirely new set of fundamental polynomials. In the Newton formula, constructing the degree-\((n + 1)\) interpolating polynomial only entails adding one more term to the degree-\(n\) interpolating polynomial.

For this purpose, we express the unique interpolating polynomial \( p_n \in \mathbb{P}_n \) as
\[
p_n(x) = a_0 + a_1(x-x_0) + a_2(x-x_0)(x-x_1) + \cdots + a_n \prod_{i=0}^{n-1}(x-x_i),
\]
which motivates the following definition.
Definition 3.11 (Divided difference and the Newton formula). The Newton formula for interpolating the values \( f_0, f_1, \ldots, f_n \) at distinct points \( x_0, x_1, \ldots, x_n \) is
\[
p_n(x) = a_0 + \sum_{k=1}^{n} \left\{ \begin{array}{l}
\frac{k!}{\prod_{i=0}^{k-1} (x-x_i)} f_i \end{array} \right\},
\]
where the \( k \)th divided difference \( a_k \) is defined as the coefficient of \( x^k \) in \( p_k(f;x) \) and is denoted by \( f[x_0, x_1, \ldots, x_k] \) or \([x_0, x_1, \ldots, x_k]f\). In particular, \( f[x_0] = f(x_0) \).

Remark 3.9. With the notation \( f[x_0, x_1, \ldots, x_k] \), we emphasize that a divided difference be a function of the given interpolating nodes. On the other hand, the notation \([x_0, x_1, \ldots, x_k]f\) emphasizes that, for any given set of nodes \( \{x_0, x_1, \ldots, x_k\} \), the divided difference is a functional that takes a function and returns a number.

Corollary 3.12. Suppose \((i_0, i_1, i_2, \ldots, i_k)\) is a permutation of \((0, 1, 2, \ldots, k)\). Then
\[
f[x_0, x_1, \ldots, x_k] = f[x_{i_0}, x_{i_1}, \ldots, x_{i_k}].
\]

Proof. The interpolating polynomial does not depend on the numbering of the interpolating nodes. The rest of the proof follows from the uniqueness of the interpolating polynomial in Theorem 3.4.

Corollary 3.13. The \( k \)th divided difference can be expressed as
\[
f[x_0, x_1, \ldots, x_k] = \sum_{i=0}^{k} \frac{f_i}{\prod_{j \neq i=0}^{k} (x_i - x_j)} = \sum_{i=0}^{k} \frac{f_i}{\pi_k(x_i)}
\]
where \( \pi_k(x) \) is defined in (3.10).

Proof. The uniqueness of interpolating polynomials in Theorem 3.4 implies that the two polynomials in (3.8) and (3.14) are the same. Then the first equality follows from (3.9) and Definition 3.11, while the second equality follows from Proposition 3.9.

Theorem 3.14. Divided differences satisfy the recursion
\[
f[x_0, x_1, \ldots, x_k] = \frac{f[x_1, x_2, \ldots, x_k] - f[x_0, x_1, \ldots, x_{k-1}]}{x_k - x_0}.
\]

Proof. By Definition 3.11, \( f[x_1, x_2, \ldots, x_k] \) is the coefficient of \( x^{k-1} \) in a degree \((k-1)\) interpolating polynomial, say, \( P_2(x) \). Similarly, let \( P_1(x) \) be the interpolating polynomial whose coefficient of \( x^{k-1} \) is \( f[x_0, x_1, \ldots, x_{k-1}] \). Construct a polynomial
\[
P(x) = P_1(x) + \frac{x - x_0}{x_k - x_0} (P_2(x) - P_1(x)).
\]
Clearly \( P(x_0) = P_1(x_0) \). Furthermore, the interpolation condition implies \( P_2(x_i) = P_1(x_i) \) for \( i = 1, 2, \ldots, k-1 \). Hence \( P(x_i) = P_1(x_i) \) for \( i = 1, 2, \ldots, k-1 \). Lastly, \( P(x_k) = P_2(x_k) \). Therefore, \( P(x) \) as above is the interpolating polynomial for given values at the \( k + 1 \) points. The rest follows from the definitions of \( P(x) \) and the \( k \)th divided difference.

Remark 3.10. Theorem 3.14 can be used to generate a table of divided differences.

Definition 3.15. The \( k \)th divided difference \((k > 0)\) on the table of divided differences
\[
\begin{array}{cccc}
x_0 & f[x_0] \\
x_1 & f[x_1] & f[x_0, x_1] \\
x_2 & f[x_2] & f[x_1, x_2] & f[x_0, x_1, x_2] \\
x_3 & f[x_3] & f[x_2, x_3] & f[x_1, x_2, x_3] & f[x_0, x_1, x_2, x_3] \\
\vdots & \vdots & \vdots & \vdots & \vdots
\end{array}
\]
is calculated as the difference of the entry immediately to the left and the one above it, divided by the difference of the \( x \)-value horizontal to the left and the one corresponding to the \( f \)-value found by going diagonally up.

Example 3.11. Derive the interpolating polynomial via the Newton formula for the function \( f \) with given values as follows. Then estimate \( f(\frac{3}{2}) \).

\[
\begin{array}{cccc}
x & f(x) \\
0 & 6 \\
1 & -3 & 9 & 3 \\
2 & -6 & -3 & 3 \\
3 & 9 & 15 & 2
\end{array}
\]

By Definition 3.15, we can construct the following table of divided difference,

\[
\begin{array}{cccc}
x & f(x) \\
0 & 6 \\
1 & -3 & 3 & 15 \\
2 & -3 & 1 & 2
\end{array}
\]


Remark 3.14. Example 3.12 shows one advantage of the Newton formula over the Lagrange formula: it is easier to preform the interpolation incrementally.

Theorem 3.16. For distinct points \( x_0, x_1, \ldots, x_n \) and an arbitrary \( x \), we have
\[
f(x) = f[x_0] + f[x_0, x_1](x - x_0) + \cdots + f[x_0, x_1, \ldots, x_n] \prod_{i=0}^{n-1} (x - x_i) + f[x_0, x_1, \ldots, x_n, x] \prod_{i=0}^{n-1} (x - x_i).
\]

Proof. Take another point \( z \neq x \). The Newton formula applied to \( x_0, x_1, \ldots, x_n, z \) yields an interpolating polynomial
\[
Q(x) = f[x_0] + f[x_0, x_1](x - x_0) + \cdots + f[x_0, x_1, \ldots, x_n, z] \prod_{i=0}^{n-1} (x - x_i).
\]
The interpolation condition \( Q(z) = f(z) \) yields
\[
 f(z) = Q(z) = f[x_0] + f[x_0, x_1](z - x_0) + \cdots + f[x_0, x_1, \ldots, x_n] \prod_{i=0}^{n-1} (z - x_i) + f[x_0, x_1, \ldots, x_n, z] \prod_{i=0}^{n} (z - x_i).
\]
Replacing the dummy variable \( z \) with \( x \) yields (3.20).

The above argument assumes \( x \neq x_i \). Now consider the case of \( x = x_j \) for some fixed \( j \). Rewrite (3.20) as
\[
 f(x) = p_n(f; x) + R(x) \text{ where } R(x) \text{ is clearly the last term in (3.20).}
\]
We need to show
\[
\forall j = 0, 1, \ldots, n, \quad p_n(f; x_j) + R(x_j) - f(x_j) = 0,
\]
which clearly holds because \( R(x_j) = 0 \) and the interpolation condition at \( x_j \) dictates \( p_n(f; x_j) = f(x_j) \).

**Remark 3.14** (An alternative proof of Theorem 3.16). The simplest nontrivial case of Theorem 3.14 leads to
\[
 f[x] = f[x_0] + (x - x_0)f[x, x_0]. \tag{3.21}
\]

**Theorem 3.14 and Corollary 3.12 yield**
\[
 f[x, x_0, \ldots, x_{n-1}] = f[x_0, \ldots, x_n] + (x - x_0)f[x, x_0, x], \tag{3.22}
\]
Set \( n = 1 \) in (3.22) and we have
\[
 f[x, x_0] = f[x_0, x_1] + (x - x_1)f[x, x_0, x_1]. \tag{3.23}
\]
Substitute (3.23) into (3.21) and we have
\[
 f[x] = f[x_0] + (x - x_0)f[x_0, x] + (x - x_0)(x - x_1)f[x, x_0, x]. \tag{3.24}
\]
Repeating the above steps \( n - 1 \) more times yields (3.20).

**Corollary 3.17.** Suppose \( f \in \mathcal{C}[a, b] \) and \( f^{(n+1)}(x) \) exists at each point of \((a, b)\). If \( a = x_0 < x_1 < \cdots < x_n = b \) and \( x \in [a, b] \), then
\[
 f[x_0, x_1, \ldots, x_n, x] = \frac{1}{(n + 1)!} f^{(n+1)}(\xi(x)) \tag{3.25}
\]
where \( \xi \) depends on \( x \) and \( \xi(x) \in (a, b) \).

**Proof.** This follows from Theorems 3.16 and 3.6.

**Corollary 3.18.** If \( x_0 < x_1 < \cdots < x_n \) and \( f \in \mathcal{C}^n[x_0, x_n] \), we have
\[
 \lim_{x_n \to x_0} f[x_0, x_1, \ldots, x_n] = \frac{1}{n!} f^{(n)}(x_0). \tag{3.26}
\]

**Proof.** Set \( x = x_0 \) in Corollary 3.17, replace \( n + 1 \) by \( n \), and we have \( \xi \to x_0 \) as \( x_n \to x_0 \) since each \( x_i \to x_0 \).

**Remark 3.15.** We can think of Theorem 3.16 as a generalization of Taylor’s expansion.

The rest of this subsection concerns a useful specialization of the Newton formula.

**Definition 3.19.** For \( n \in \mathbb{N}^+ \), the \( n \)th forward difference associated with a sequence of values \( \{f_0, f_1, \ldots\} \) is
\[
 \Delta f_i = f_{i+1} - f_i, \quad \Delta^{n+1} f_i = \Delta \Delta^n f_i = \Delta^n f_{i+1} - \Delta^n f_i, \tag{3.27}
\]
and the \( n \)th backward difference is
\[
 \nabla f_i = f_{i-1} - f_i, \quad \nabla^{n+1} f_i = \nabla \nabla^n f_i = \nabla^n f_{i-1} - \nabla^n f_i. \tag{3.28}
\]

**Theorem 3.20.** The forward difference and backward difference are related as
\[
 \forall n \in \mathbb{N}^+, \quad \Delta^n f_i = \nabla^n f_{i+n}. \tag{3.29}
\]

**Proof.** An easy induction.

**Remark 3.16.** In light of Theorem 3.20, hereafter we only study forward differences since similar conclusions on backward differences can be deduced by Theorem 3.20.

**Theorem 3.21.** The forward difference can be expressed explicitly as
\[
 \Delta^n f_i = \sum_{k=0}^{n} (-1)^{n-k} \binom{n}{k} f_{i+k}. \tag{3.30}
\]

**Proof.** For \( n = 1 \), (3.30) reduces to \( \Delta f_i = f_{i+1} - f_i \). The rest of the proof is an induction utilizing the identity
\[
 \binom{n}{k} + \binom{n}{k+1} = \binom{n+1}{k}. \tag{3.31}
\]
Suppose (3.30) holds. For the inductive step, we have
\[
 \Delta^{n+1} f_i = \Delta [\Delta^n f_i] = \Delta \left( \sum_{k=0}^{n} (-1)^{n-k} \binom{n}{k} f_{i+k} \right)
\]
\[
 = \sum_{k=0}^{n} (-1)^{n-k} \binom{n}{k} f_{i+k+1} - \sum_{k=0}^{n} (-1)^{n-k} \binom{n}{k} f_{i+k}
\]
\[
 = \sum_{k=1}^{n+1} (-1)^{n+1-k} \binom{n}{k} f_{i+k} + (-1)^{n+1} f_i
\]
\[
 = \sum_{k=0}^{n+1} (-1)^{n+1-k} \binom{n+1}{k} f_{i+k},
\]
where the second line follows from (3.31), the third line from splitting one term out of each sum and replacing the dummy variable in the first sum, and the fourth line from (3.31) and the fact that \((-1)^{n+1} f_i \) and \( f_{i+n+1} \) contribute to the first and last terms, respectively.

**Remark 3.17.** In Definition 3.19 and Theorems 3.20 and 3.21, there are no restrictions on the locus of the values \( f_i \). However, for uniformly spaced grids, a divided difference reduces to a forward difference; this is the original motivation for introducing Definition 3.19.
Theorem 3.22. On a grid \( x_i = x_0 + ih \) with uniform spacing \( h \), the sequence of values \( f_i = f(x_i) \) satisfies
\[
\forall n \in \mathbb{N}^+, \quad f[x_0, x_1, \ldots, x_n] = \frac{\Delta^n f_0}{n! h^n}.
\] (3.32)

Proof. Of course (3.32) can be proven by induction. Here we provide a more informative proof. For \( \pi_n(x) \) defined in (3.10), we have \( \pi'(x_k) = \prod_{i=0, i \neq k}^n (x_k - x_i) \). It follows from \( x_k - x_i = (k - i)h \) that
\[
\pi'(x_k) = \prod_{i=0, i \neq k}^n (k - i)h = h^n k! (n - k)! (-1)^{n-k}.
\] (3.33)

Then we have
\[
f[x_0, x_1, \ldots, x_n] = \sum_{k=0}^n \frac{f_k}{\pi'(x_k)} = \frac{1}{h^n n!} \sum_{k=0}^n (-1)^{n-k} \binom{n}{k} \Delta^n f_0 = \frac{\Delta^n f_0}{h^n n!},
\]
where the first step follows from Corollary 3.13, the second from (3.33), and the last from Theorem 3.21. \( \square \)

Theorem 3.23. (Newton’s forward difference formula). Suppose \( p_n(f; x) \in \mathbb{P}_n \) interpolates \( f(x) \) on a uniform grid \( x_i = x_0 + ih \) at \( x_0, x_1, \ldots, x_n \) with \( f_i = f(x_i) \). Then
\[
\forall s \in \mathbb{R}, \quad p_n(f; x_0 + sh) = \sum_{k=0}^n \binom{s}{k} \Delta^k f_0,
\] (3.34)
where \( \Delta^0 f_0 = f_0 \) and
\[
\binom{s}{k} = \frac{s(s - 1) \cdots (s - k + 1)}{k!}.
\] (3.35)

Proof. Set \( f(x) = p_n(f; x) \) in Theorem 3.16, apply Theorem 3.22, and we have
\[
p(x) = \sum_{k=0}^n \frac{\Delta^k f_0}{k! h^k} \prod_{i=0}^{k-1} (x - x_i);
\]
the remainder is zero because any \((n + 1)\)th divided difference applied to a degree \( n \) polynomial is zero. The proof is completed by \( x = x_0 + sh \), \( x_i = x_0 + ih \), and (3.35). \( \square \)

Remark 3.18. One may consider (3.35) as a generalization of \( n \) in \( \binom{n}{k} \) from integers to real numbers.

Remark 3.19. Unlike Theorem 3.22, Theorem 3.23 does not have \( h \) in its formula. This is due to the change of variable \( x = x_0 + sh \).

3.5 The Neville-Aitken algorithm

In Example 3.11, we first derived the interpolating polynomial \( p(x) \) of the function \( f(x) \) and then evaluated \( p(x) \) to estimate \( f(x) \) at a given point, say \( x_0 \). The Neville-Aitken algorithm admits a direct estimation of \( f(x_0) \) without constructing \( p(x) \).

Theorem 3.24. Denote \( p_0^{[i]} = f(x_i) \) for \( i = 0, 1, \ldots, n \). For all \( k = 0, 1, \ldots, n - 1 \) and \( i = 0, 1, \ldots, n - k - 1 \), define
\[
p_{k+1}^{[i]}(x) = \frac{(x - x_i) p_k^{[i+1]}(x) - (x - x_{i+k+1}) p_k^{[i]}(x)}{x_{i+k+1} - x_i}.
\] (3.36)

Then each \( p_k^{[i]} \) is the interpolating polynomial for the function \( f(x) \) at the points \( x_i, x_{i+1}, \ldots, x_{i+k} \). In particular, \( p_n^{[i]} \) is the interpolating polynomial of degree \( n \) for the function \( f(x) \) at the points \( x_0, x_1, \ldots, x_n \).

Proof. The induction basis clearly holds for \( n = 0 \) because of the definition \( p_0^{[i]} = f(x_i) \). Suppose that \( p_k^{[i]} \) is the interpolating polynomial of degree \( k \) for the function \( f(x) \) at the points \( x_i, x_{i+1}, \ldots, x_{i+k} \). Then we have
\[
\forall j = i + 1, i + 2, \ldots, i + k, \quad p_{k+1}^{[j]}(x_j) = p_{k+1}^{[j]}(x_j) = f(x_j),
\]
which, together with (3.36), implies
\[
\forall j = i + 1, i + 2, \ldots, i + k, \quad p_{k+1}^{[j]}(x_j) = f(x_j).
\]
In addition, (3.36) and the induction hypothesis yield
\[
p_{k+1}^{[i]}(x_i) = p_{k+1}^{[i]}(x_i) = f(x_i),
\]
\[
p_{k+1}^{[i]}(x_{i+k+1}) = p_{k+1}^{[i]}(x_{i+k+1}) = f(x_{i+k+1}).
\]
The proof is completed by the last three equations and the uniqueness of interpolating polynomials. \( \square \)

Example 3.20. To estimate \( f(x) \) for \( x = 3 \) directly from the table in Example 3.11, we construct a table by repeating (3.36) with \( x_i = i \) for \( i = 0, 1, 2, 3 \).

<table>
<thead>
<tr>
<th>( i )</th>
<th>( x_i )</th>
<th>( f(x_i) )</th>
<th>( p_0^{[i]}(x) )</th>
<th>( p_1^{[i]}(x) )</th>
<th>( p_2^{[i]}(x) )</th>
<th>( p_3^{[i]}(x) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>6</td>
<td>-5 ( \frac{4}{3} )</td>
<td>-5 ( \frac{2}{3} )</td>
<td>-6</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>( \frac{1}{2} )</td>
<td>-3</td>
<td>- ( \frac{2}{3} )</td>
<td>- ( \frac{27}{4} )</td>
<td>-6</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>- ( \frac{1}{2} )</td>
<td>-6</td>
<td>- ( \frac{27}{4} )</td>
<td>-6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>- ( \frac{3}{2} )</td>
<td>9</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The result is the same as that in Example 3.11. In contrast, the calculation and layout of the two tables are distinct.

3.6 Hermite interpolation

Definition 3.25. Given distinct points \( x_0, x_1, \ldots, x_k \) in \( [a, b] \), non-negative integers \( m_0, m_1, \ldots, m_k \), and a function \( f \in C^M[a, b] \) where \( M = \max m_i \), the Hermite interpolation problem seeks to find a polynomial \( p \) of the lowest degree such that
\[
\forall i = 0, 1, \ldots, k, \forall \mu = 0, 1, \ldots, m_i, \quad p^{(\mu)}(x_i) = f_i^{(\mu)}(x_i),
\] (3.38)
where \( f_i^{(\mu)}(x) \) is the value of the \( \mu \)th derivative of \( f(x) \) at \( x_i \); in particular, \( f_i^{(0)} = f(x_i) \).

Definition 3.26. The \( n \)th divided difference at \( n + 1 \) “confluent” (i.e. identical) points is defined as
\[
f[x_0, x_0, \ldots, x_0] = \frac{1}{n!} f^{(n)}(x_0),
\] (3.39)
where \( x_0 \) is repeated \( n + 1 \) time on the left-hand side.
Remark 3.21. With Definition 3.26, we can build a table of divided difference for Hermite interpolation, in the same way as we did in Newton interpolation. The only difference here is to apply (3.39) whenever possible; see the examples on pages 98–99 on NAG2012.

Theorem 3.27. For the Hermite interpolation problem in Definition 3.25, denote \( N = k + \sum_i m_i \). Denote by \( p_N(f; x) \) the unique element of \( \mathbb{P}_N \) for which (3.38) holds. Suppose \( f^{(N+1)}(x) \) exists in \((a, b)\). Then

\[
  f(x) - p_N(f; x) = \frac{f^{(N+1)}(\xi)}{(N+1)!} \prod_{i=0}^k (x - x_i)^{m_i+1}.
\]

Proof. The proof is similar to that of Theorem 3.6. Pay attention to the difference caused by the multiple roots of the polynomial \( \prod_{i=0}^k (x - x_i)^{m_i+1} \).

\( \square \)

3.7 The Chebyshev polynomials

Remark 3.22. The points \( x_0, x_1, \ldots, x_n \) in Theorem 3.4 are usually given \emph{a priori}, e.g., as uniformly distributed over the interval \([x_0, x_n]\). As \( n \) increases, the degree of the interpolating polynomial also increases. Ideally we would like to have

\[
  \forall f \in C[x_0, x_n], \forall x \in [x_0, x_n], \lim_{n \to +\infty} p_n(f; x) = f(x).
\]

(3.41)

However, this is not true for polynomial interpolation on equally spaced points. The famous Runge’s example illustrates the violent oscillations at the end of the interval.

The above plot is created by interpolating

\[
  f(x) = \frac{1}{1 + x^2}
\]

on \( x_i = -5 + 10 \cdot \frac{i}{n}, i = 0, 1, \ldots, n \) with \( n = 2, 4, 6, 8 \).

Remark 3.23. How did the Runge phenomenon happen? Despite the large errors in the above plot, the theorem of Cauchy remainder still holds:

\[
  R_n(f; x) := f(x) - p_n(f; x) = \frac{f^{(n+1)}(\xi)}{(n+1)!} \prod_{i=0}^n (x - x_i).
\]

To understand the Runge phenomenon, we ask the question: which of the three terms on the right-hand side is the culprit responsible for large errors? Clearly it is not \((n+1)!\), it is not \( \prod_{i=0}^n (x - x_i) \) either: we can choose \( x_n - x_0 < 1 \) so that each of its factors is smaller than one. So the culprit term must be \( f^{(n+1)}(\xi) \). To illustrate this point, we consider a circle with radius \( R \). For \( \theta \in (0, \frac{\pi}{2}) \), the local neighborhood of a point \( p \) on the circle can be expressed as the graph of a function

\[
  y = H(x) = R \cos \theta - \sqrt{R^2 - (x + R \sin \theta)^2}.
\]

(3.43)

The values of factorials and derivatives of \( H(x) \) are listed in the following table.

<table>
<thead>
<tr>
<th>( n )</th>
<th>( n! )</th>
<th>( H^{(n)}(0) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>2.83</td>
</tr>
<tr>
<td>3</td>
<td>6</td>
<td>12</td>
</tr>
<tr>
<td>4</td>
<td>24</td>
<td>1.02e+02</td>
</tr>
<tr>
<td>5</td>
<td>120</td>
<td>1.20e+03</td>
</tr>
<tr>
<td>6</td>
<td>720</td>
<td>1.83+04</td>
</tr>
<tr>
<td>7</td>
<td>5040</td>
<td>3.43e+05</td>
</tr>
<tr>
<td>8</td>
<td>40320</td>
<td>7.58e+06</td>
</tr>
<tr>
<td>9</td>
<td>362880</td>
<td>1.94e+08</td>
</tr>
<tr>
<td>10</td>
<td>3.63e+06</td>
<td>5.61e+09</td>
</tr>
</tbody>
</table>

Why does \( H^{(n)} \) quickly outweigh \((n+1)!\)? Because the chain rule implies a \emph{potentially} exponential growth of the number of terms in the derivatives of the function \( H \).

Remark 3.24. We now have a serious problem: the function to be interpolated is usually fixed and the value of \( H^{(n)} \) will grow fast as \( n \) increases. However, if we have the freedom to choose positions of the interpolation points, we can further reduce the remainder. More precisely, how can we choose \( x_0, x_1, \ldots, x_n \) to minimize the quantity

\[
  \max_{x \in [a,b]} |(x - x_0)(x - x_1) \cdots (x - x_n)|.
\]

Such \( x_i \)'s are the zeros of the Chebyshev polynomial.

Definition 3.28. The \emph{Chebyshev polynomial} of degree \( n \) of the first kind is a polynomial \( T_n : [-1, 1] \to \mathbb{R} \)

\[
  T_n(x) = \cos(n \arccos(x)).
\]

(3.44)

Remark 3.25. In estimating the remainder in Corollary 3.7, we cannot do much about \( \max |f^{(n+1)}(x)| \) as \( f \) is usually given \emph{a priori}. However, if we have the freedom to choose positions of the interpolation points, we can further reduce the remainder. More precisely, how can we choose \( x_0, x_1, \ldots, x_n \) to minimize the quantity

\[
  \max_{x \in [a,b]} |(x - x_0)(x - x_1) \cdots (x - x_n)|.
\]

Such \( x_i \)'s are the zeros of the Chebyshev polynomial.

Remark 3.26. First, \( |T_n(x)| \leq 1 \). Second, (3.44) indeed defines a polynomial of degree \( n \). Set \( x = \cos \theta \). Then

\[
  e^{i\theta} = \cos \theta + i \sin \theta
\]

\[
  \Rightarrow (e^{i\theta})^n = e^{in\theta} = \cos(n\theta) + i \sin(n\theta) = (x + i \sqrt{1 - x^2})^n
\]

\[
  \Rightarrow \cos n\theta = x^n + \binom{n}{2} x^{n-2} (x^2 - 1) + \cdots
\]

27
where the first line is the Euler’s formula and the last line follows from the binomial theorem. The first five Chebyshev polynomials are

\[
\begin{align*}
T_0(x) &= 1 \\
T_1(x) &= x \\
T_2(x) &= 2x^2 - 1 \\
T_3(x) &= 4x^3 - 3x \\
T_4(x) &= 8x^4 - 8x^2 + 1
\end{align*}
\]

Aside from the above explicit formula, Chebyshev polynomials can also be generated by a recurrence relation.

**Theorem 3.29.**

\[\forall n \in \mathbb{N}^+, \quad T_{n+1}(x) = 2xT_n(x) - T_{n-1}(x). \quad (3.45)\]

**Proof.** By trigonometric identities, we have

\[
\begin{align*}
\cos(n + 1)\theta &= \cos n\theta \cos \theta - \sin n\theta \sin \theta, \\
\cos(n - 1)\theta &= \cos n\theta \cos \theta + \sin n\theta \sin \theta.
\end{align*}
\]

Adding up the two equations and setting \(\cos \theta = x\) complete the proof. \(\square\)

**Corollary 3.30.** The coefficient of \(x^n\) in \(T_n\) is \(2^{n-1}\) for each \(n > 0\).

**Proof.** Use (3.45) and \(T_1 = x\) in an induction.

**Theorem 3.31.** \(T_n(x)\) has simple zeros at the \(n\) points

\[x_k = \cos \frac{2k - 1}{2n} \pi, \quad (3.46)\]

where \(k = 1, 2, \ldots, n\). For \(x \in [-1, 1]\) and \(n \in \mathbb{N}^+, \ T_n(x)\) has extreme values at the \(n + 1\) points

\[x_k' = \cos \frac{k}{n} \pi, \quad k = 0, 1, \ldots, n, \quad (3.47)\]

where it assumes the alternating values \((-1)^k\).

**Proof.** (3.44) and (3.46) yield

\[
T_n(x_k) = \cos(n \arccos \left(\frac{2k - 1}{2n} \pi\right)) = \cos \left(\frac{2k - 1}{2} \pi\right) = 0.
\]

Differentiate (3.44) and we have

\[
T'_n(x_k) = \frac{n}{\sqrt{1 - x_k^2}} \sin(n \arccos x).
\]

Then each \(x_k\) must be a zero since

\[
T'_n(x_k) = \frac{n}{\sqrt{1 - x_k^2}} \sin \left(\frac{2k - 1}{2} \pi\right) \neq 0.
\]

In contrast, \(\forall k = 1, 2, \ldots, n - 1,\)

\[
T'_n(x_k') = n \left(1 - \cos^2 \frac{kn}{n}\right)^{-\frac{1}{2}} \sin(k \pi) = 0.
\]

For \(k = 0, n, \ T_n(x_k')\) attains its extreme values since \(T_n(x_0') = 1, \ T_n(x_n') = -1, \) and by (3.44) \(|T_n(x)| \leq 1\). Clearly these are the only extrema of \(T_n(x)\) on \([-1, 1]\). \(\square\)

**Remark 3.27.** According to Theorem 3.31, the zeros of Chebyshev polynomials are the projections onto the real line of equally spaced points on the unit circle. In the following plot, \(n = 4, \) the \(n + 1\) squares represent the extreme points, the stars represent the points \(\exp(i \frac{2\pi}{2n} k)\) on the unit circle and their projection onto the horizontal axis.

**Exercise 3.28.** Write a matlab code to reproduce the above plot.

**Remark 3.29.** For a polynomial of degree no more than \(n\) defined over the interval \([-1, 1]\), the maximum numbers of its zeros and extrema are \(n\) and \(n + 1\), respectively. As proved in Theorem 3.31 and illustrated in the above plot, Chebyshev polynomials achieve both maximums. This fact, together with its trigonometric definition, leads to the central result of this subsection.

**Theorem 3.32 (Chebyshev).** Denote by \(\tilde{P}_n\) the class of all polynomials of degree \(n\) with leading coefficient 1. Then

\[
\forall p \in \tilde{P}_n, \quad \max_{x \in [-1, 1]} \left|\frac{T_n(x)}{2^n}\right| \leq \max_{x \in [-1, 1]} |p(x)|. \quad (3.48)
\]

**Proof.** By Theorem 3.31, \(T_n(x)\) assumes its extrema \(n + 1\) times at the points \(x_k'\) defined in (3.47). Suppose (3.48) does not hold. Then Corollary 3.30 implies that

\[
\exists p \in \tilde{P}_n \text{ s.t. } \max_{x \in [-1, 1]} |p(x)| < \frac{1}{2^{n-1}}. \quad (3.49)
\]

Consider the polynomial \(Q(x) = \frac{1}{2^n} T_n(x) - p(x)\).

\[
Q(x_k') = \frac{(-1)^k}{2^{n-1}} - p(x_k'), \quad k = 0, 1, \ldots, n.
\]

By (3.49), \(Q(x)\) has alternating signs at these \(n + 1\) points. Hence \(Q(x)\) must have \(n\) zeros. However, by the construction of \(Q(x)\), the degree of \(Q(x)\) is at most \(n - 1\). Therefore, \(Q(x) \equiv 0\) and \(p(x) = \frac{1}{2^n} T_n(x)\), which implies \(\max |p(x)| = \frac{1}{2^{n-1}}\). This is a contradiction to (3.49). \(\square\)

**Corollary 3.33.**

\[
\max_{x \in [-1, 1]} |x^n + a_1 x^{n-1} + \cdots + a_n| \geq \frac{1}{2^n - 1}. \quad (3.50)
\]

**Corollary 3.34.** Suppose polynomial interpolation is performed for \(f\) on the \(n + 1\) zeros of \(T_{n+1}(x)\) as in Theorem 3.31. The Cauchy remainder in Theorem 3.6 satisfies

\[
|R_n(f; x)| \leq \frac{1}{2^n(n + 1)!} \max_{x \in [-1, 1]} |f^{(n+1)}(x)|. \quad (3.51)
\]

**Proof.** Theorem 3.6, Corollary 3.30, and Theorem 3.31 yield

\[
|R_n(f; x)| = \left|\frac{f^{(n+1)}(x)}{(n + 1)!} \sum_{i=0}^{n} (x - x_i)\right| = \left|\frac{f^{(n+1)}(x)}{2^n(n + 1)!}\right| |T_{n+1}|.
\]

Definition 3.28 completes the proof as \(|T_{n+1}| \leq 1\). \(\square\)