Interpolation *

Hector D. Ceniceros

1 Approximation Theory

Given $f \in C[a, b]$, we would like to find a "good" approximation to it by "simpler functions", i.e. functions in a given class (or family) Φ. For example,$\Phi = \mathbb{P}_n = \{$all polynomials of degree $\leq n\}$.

A natural problem is that of finding the *best approximation* to f by functions in Φ. But how do we measure the accuracy of any approximation? that is, what norm[1] do we use? We have several choices for norms of functions. The most commonly used are:

1. The max or infinity norm: $\|f\|_\infty = \sup_{x\in[a,b]} |f(x)|$.

2. The 2-norm: $\|f\|_2 = (\int_a^b f^2(x)dx)^{\frac{1}{2}}$.

3. The p-norm: $\|f\|_p = (\int_a^b f^p(x)dx)^{\frac{1}{p}}$.

Later, we will need to consider weighted norms: for some positive function $\omega(x)$ in $[a, b]$ (it could be zero on a finite number of points) we define

$$\|f\|_{\omega,2} = \left(\int_a^b \omega(x) f^2(x) dx \right)^{\frac{1}{2}}. \tag{1}$$

---

[1]A norm $\|\cdot\|$ is a real valued function on a vector space such that (1) $\|f\| > 0, f \not\equiv 0$, (2) $\|\lambda f\| = |\lambda|\|f\|$ for any λ scalar, and (3) $\|f + g\| \leq \|f\| + \|g\|$.

Then, by best approximation in Φ we mean a function $p \in \Phi$ such that

$$\|f - p\| \leq \|f - q\|, \forall q \in \Phi.$$

Computationally, it is often more efficient to seek not the best approximation but one that is sufficiently accurate and fast converging to f. The central building block for this approximation is the problem of interpolation.

2 Interpolation

Let us focus on the case of approximating a given function by a polynomial of degree at most n. Then the *interpolation problem* can be stated as follows: Given $n+1$ distinct points, $x_0, x_1, ..., x_n$ called *nodes* and corresponding values $f(x_0), f(x_1), ..., f(x_n)$, find a polynomial of degree at most n, $P_n(x)$, which satisfies (the interpolation property)

$$\begin{gathered} P_n(x_0) = f(x_0) \\ P_n(x_1) = f(x_1) \\ \vdots \\ P_n(x_n) = f(x_n). \end{gathered}$$

Let us represent such polynomial as $P_n(x) = a_0 + a_1 x + \cdots + a_n x^n$. Then, the interpolation property means

$$P_n(x_0) = f(x_0), P_n(x_1) = f(x_1), \cdots, P_n(x_n) = f(x_n),$$

which implies

$$\begin{gathered} a_0 + a_1 x_0 + \cdots + a_n x_0^n = f(x_0) \\ a_0 + a_1 x_1 + \cdots + a_n x_1^n = f(x_1) \\ \vdots \\ a_0 + a_1 x_n + \cdots + a_n x_n^n = f(x_n). \end{gathered}$$

This is a linear system of $n+1$ equations in $n+1$ unknowns (the polynomial coefficients $a_0, a_1, \ldots, a_n$). In matrix form:

$$\begin{bmatrix} 1 & x_0 & x_0^2 \cdots x_0^n \\ 1 & x_1 & x_1^2 \cdots x_1^n \\ \vdots & & \\ 1 & x_n & x_n^2 \cdots x_n^n \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_n \end{bmatrix} = \begin{bmatrix} f(x_0) \\ f(x_1) \\ \vdots \\ f(x_n) \end{bmatrix} \tag{2}$$

Does this linear system have a solution? Is this solution unique? The answer is yes to both. Here is a simple proof. Take $f \equiv 0$, then $P_n(x_j) = 0$,for $j = 0, 1, ..., n$ but P_n is a polynomial of degree $\leq n$, it cannot have $n+1$ zeros unless $P_n(x) \equiv 0$, which implies $a_0 = a_1 = \cdots = a_n = 0$. That is, the homogenous problem associated with (2) has only the trivial solution. Therefore, (2) has a unique solution.

In general the values to interpolate might not come from a function. They are just data supplied to us. We will often write $(x_0, f_0), (x_1, f_1)$, etc., to emphasize this more general setting.

Example 1. *As an illustration let us consider interpolation by a linear polynomial, $P_1(x)$. Suppose we are given (x_0, f_0) and (x_1, f_1). We have written $P_1(x)$ explicitly in the Introduction. We write it now in a different form:*

$$P_1(x) = \frac{x - x_1}{x_0 - x_1} f_0 + \frac{x - x_0}{x_1 - x_0} f_1 \tag{3}$$

Clearly, this polynomial has degree at most 1 and satisfies the interpolation property:

$$P_1(x_0) = f_0, \tag{4}$$

$$P_1(x_1) = f_1. \tag{5}$$

Example 2. *Given $(x_0, f_0), (x_1, f_1), (x_2, f_2)$ let us construct $P_2(x)$, the polynomial of degree at most 2 which interpolates these points. The way we have written $P_1(x)$ in (3) is suggestive of how to explicitly write $P_2(x)$:*

$$P_2(x) = \frac{(x - x_1)(x - x_2)}{(x_0 - x_1)(x_0 - x_2)} f_0 + \frac{(x - x_0)(x - x_2)}{(x_1 - x_0)(x_1 - x_2)} f_1 + \frac{(x - x_0)(x - x_1)}{(x_2 - x_0)(x_1 - x_1)} f_2.$$

If we define

$$l_0^{(2)}(x) = \frac{(x - x_1)(x - x_2)}{(x_0 - x_1)(x_0 - x_2)}, \tag{6}$$

$$l_1^{(2)}(x) = \frac{(x - x_0)(x - x_2)}{(x_1 - x_0)(x_1 - x_2)}, \tag{7}$$

$$l_2^{(2)}(x) = \frac{(x - x_0)(x - x_1)}{(x_2 - x_0)(x_1 - x_1)}, \tag{8}$$

then we simply have

$$P_2(x) = l_0^{(2)}(x) f_0 + l_1^{(2)}(x) f_1 + l_2^{(2)}(x) f_2. \tag{9}$$

Note that each of the polynomials (6), (7), and (8) are exactly of degree 2 and they satisfy $l_j^{(2)}(x_k) = \delta_{jk}$ [2]*. Therefore, it follows that* $P_2(x)$ *given by (9) satisfies the interpolation property*

$$P_2(x_0) = f_0, \tag{10}$$
$$P_2(x_1) = f_1, \tag{11}$$
$$P_2(x_2) = f_2. \tag{12}$$

We can now write down the polynomial (of degree at most n) which interpolates $n+1$ given values, $(x_0, f_0), \ldots, (x_n, f_n)$, where the interpolation nodes $x_0, \ldots, x_n$ are assumed distinct.

Define

$$\begin{aligned} l_j^{(n)}(x) &= \frac{(x - x_0)\cdots(x - x_{j-1})(x - x_{j+1}\cdots(x - x_n)}{(x_j - x_0)\cdots(x_j - x_{j-1})(x_j - x_{j+1}\cdots(x_j - x_n)} \\ &= \prod_{k=0,k\neq j}^{n} \frac{(x - x_k)}{(x_j - x_k)}, \qquad \text{for } j = 0, 1, ..., n. \end{aligned} \tag{13}$$

These are called the *elementary Lagrange polynomials* of degree n. Note that $l_j^{(n)}(x_k) = \delta_{jk}$. Therefore

$$P_n(x) = l_0^{(n)}(x)f_0 + l_1^{(n)}(x)f_1 + \cdots + l_n^{(n)}(x)f_n = \sum_{j=0}^{n} l_j^{(n)}(x)f_j \tag{14}$$

interpolates the given data, i.e., it satisfies the interpolation property $P_n(x_j) = f_j$ for $j = 0, 1, 2, \ldots, n$. Relation (14) is called the *Lagrange form* of the interpolating polynomial. The following result summarizes our discussion.

Theorem 1. *Given the* $n + 1$ *values* $(x_0, f_0), \ldots, (x_n, f_n)$*, for* $x_0, x_1, ..., x_n$ *distinct. There is a unique polynomial of degree at most* n*,* $P_n(x)$*, such that* $P_n(x_j) = f_j$ *for* $j = 0, 1, \ldots, n$*.*

Proof. $P_n(x)$ in (14) is of degree at most n and interpolates the data. Uniqueness follows from the fundamental algebra : suppose there is another polynomial $Q_n(x)$ of degree at most n such that $Q_n(x_j) = f_j$ for $j = 0, 1, \ldots, n$. Consider $W(x) = P_n(x) - Q_n(x)$. This is a polynomial of degree at most n and $W(x_j) = P_n(x_j) - Q_n(x_j) = f_j - f_j = 0$ for $j = 0, 1, 2, \ldots, n$, which is impossible unless $W(x) \equiv 0$ which implies $Q_n = P_n$. □

[2] δ_{jk} is the Kronecker delta, i.e. $\delta_{jk} = 0$ if $k \neq j$ and 1 if $k = j$.

3 Connection to Best Approximation

We can view interpolation as a linear operator. Suppose that we have $n+1$ distinct nodes $x_0, x_1, \ldots, x_n$ contained in an interval $[a, b]$. Let f and g be two continuous functions in $[a, b]$ and α and β two scalars. Then, the interpolating polynomial for $\alpha f(x) + \beta g(x)$ is $P(x) = \alpha P_n(x) + \beta Q_n(x)$ where $P_n(x)$ and $Q_n(x)$ are the interpolating polynomials of f and g, respectively. This follows immediately from (14). Also, note that if f is a polynomial of degree at most n, its interpolating polynomial is itself, i.e. $P_n(x) = f(x)$.

Now suppose that $P_n^*(x)$ is the best polynomial approximation of f in the max or uniform norm, i.e.

$$\min_{p \in \mathbb{P}_n} \|f - P\|_\infty = \|f - P_n^*\|_\infty, \tag{15}$$

where $\mathbb{P}_n = \{\text{all polynomials of degree} \leq n\}$. Let $P_n(x)$ be the interpolating polynomials of f at $x_0, x_1, \ldots, x_n$. Then,

$$\|f - P_n\|_\infty = \|f - P_n^* - (P_n - P_n^*)\|_\infty \leq \|f - P_n^*\|_\infty + \|P_n - P_n^*\|_\infty.$$

But $P_n(x) - P_n^*(x)$ is a polynomial of degree at most n which interpolates $f - P_n^*$, that is

$$P_n(x) - P_n^*(x) = \sum_{j=0}^{n} l_j^{(n)}(x)(f(x_j) - P_n^*(x_j)).$$

Therefore,

$$\|P_n - P_n^*\|_\infty \leq \Lambda_n \|f - P_n^*\|_\infty \tag{16}$$

where

$$\Lambda_n = \max_{a \leq x \leq b} \sum_{j=0}^{n} |l_i^{(n)}(x)| \tag{17}$$

is called the *Lebesgue Constant*. Using this in (16) we obtain

$$\|f - P_n\|_\infty \leq (1 + \Lambda_n)\|f - P_n^*\|_\infty. \tag{18}$$

This inequality connects the interpolation error $\|f - P_n\|_\infty$ with the best approximation error $\|f - P_n^*\|_\infty$. Let us see if can extract more information from this connection.

There is a fundamental result in approximation theory, which states that any continuous function can be approximated uniformly, with arbitrary accuracy by a polynomial. This is the celebrated Weierstrass Theorem.

Theorem 2. *Weierstrass Theorem. Let f be a continuous function in $[a,b]$. Given $\epsilon > 0$ there is a polynomial P such that*

$$\|f - P\|_\infty < \epsilon.$$

Weierstrass theorem implies that as we increase the degree the best approximation polynomial converges uniformly to f, that is $\|f - P_n^*\|_\infty \to 0$ as $n \to \infty$. However, because the Lebesgue constant Λ_n is not bounded in n, we cannot conclude that $\|f - P_n\|_\infty$ as $n \to \infty$, i.e. that the interpolating polynomial, as we add more and more nodes, converges uniformly to f. That depends on the regularity of f and on the distribution of the nodes. We will discuss this further later.

4 Barycentric Formula

The Lagrange form of the interpolating polynomial is not convenient for computations. If we want to increase the degree of the polynomial we cannot reuse the work done in getting and evaluating a lower degree one. However, we can obtain a very efficient formula by rewriting the interpolating polynomial as follows.

Let

$$w(x) = (x - x_0)(x - x_1) \cdots (x - x_n). \tag{19}$$

The numerator in $l_j^{(n)}(x)$ is $w(x)/(x - x_j)$. Thus, we can write

$$l_j^{(n)}(x) = w(x)\frac{\lambda_j^{(n)}}{x - x_j}, \quad \lambda_j^{(n)} = \frac{1}{\prod_{\substack{k=0 \\ k \neq j}}^{n}(x_j - x_k)}. \tag{20}$$

Therefore

$$P_n(x) = \sum_{j=0}^{n} w(x)\frac{\lambda_j^{(n)}}{x - x_j} f(x_j) = w(x)\sum_{j=0}^{n} \frac{\lambda_j^{(n)}}{x - x_j} f(x_j). \tag{21}$$

Now, note that from (14) with $f(x) \equiv 1$ it follows that

$$\sum_{j=0}^{n} l_j^{(n)}(x) = 1 \tag{22}$$

and diving (21) by (22) we get the so-called *Barycentric Formula* for interpolation:

$$P_n(x) = \frac{\displaystyle\sum_{j=0}^{n} \frac{\lambda_j^{(n)}}{x - x_j} f(x_j)}{\displaystyle\sum_{j=0}^{n} \frac{\lambda_j^{(n)}}{x - x_j}}, \qquad \text{for } x \neq x_j,\ j = 0, 1, \ldots, n. \tag{23}$$

For $x = x_j$, $j = 0, 1, \ldots, n$, the interpolation property should be used: $P_n(x_j) = f(x_j)$.

The numbers $\lambda_j^{(n)}$ depend only on the nodes $x_0, x_1, ..., x_n$ and not on given values $f(x_0), f(x_1), ..., f(x_n)$. We can precompute them efficiently as follows:

$\lambda_0^{(0)} = 1$;
for $m = 1 : n$
    for $j = 0 : m - 1$
        $\lambda_j^{(m)} = \frac{\lambda_j^{(m-1)}}{x_j - x_m}$;
    end
    $\lambda_m^{(m)} = \frac{1}{\prod_{k=0}^{m-1} (x_m - x_k)}$;
end

If we want to add one more point $(x_{n+1}, f(x_{n+1}))$ we just extend the m-loop to $n + 1$ to generate $\lambda_0^{(n+1)}, \lambda_1^{(n+1)}, \cdots, \lambda_{n+1}^{(n+1)}$.

For equidistributed points, $x_j = x_0 + jh$, $j = 0, 1, \ldots, n$ we have

$$\begin{aligned}
\lambda_j^{(n)} &= \frac{1}{(x_j - x_0)\cdots(x_j - x_{j-1})(x_j - x_{j+1})\cdots(x_j - x_n)} \\
&= \frac{1}{(jh)[(j-1)h]\cdots(h)(-h)(-2h)\cdots(j-n)h} \\
&= \frac{1}{(-1)^{n-j}h^n[j(j-1)\cdots 1][1\cdot 2\cdots(n-j)]} \\
&= \frac{(-1)^{j-n}}{h^n n!}\frac{n!}{j!(n-j)!} \\
&= \frac{(-1)^{j-n}}{h^n n!}\binom{n}{j} \\
&= \frac{(-1)^n}{h^n n!}(-1)^j\binom{n}{j}.
\end{aligned}$$

We can omit the factor $\frac{(-1)^n}{h^n n!}$ because it cancels out in the Barycentric Formula. Thus, for equidistributed points we can use

$$\lambda_j^{(n)} = (-1)^j\binom{n}{j}, \quad j = 0, 1, \ldots n. \tag{24}$$

5 Newton's Form and Divided Differences

There is another representation of the interpolating polynomial which is both very efficient computationally and very convenient in the derivation of numerical methods based on interpolation. The idea of this representation, due to Newton, is to use successively lower order polynomials for constructing $P_n(x)$.

Suppose we have gotten $P_{n-1}(x)$, that is the interpolating poly of f at $x_0, ..., x_{n-1}$ and we would like to obtain $P_n(x)$, the interpolating poly of f at $x_0, ..., x_n$ by reusing $P_{n-1}(x)$. $P_n(x) - P_{n-1}(x) = R(x)$, where $R(x)$ is a polynomial of degree at most n. Moreover, for $j = 0, \ldots, n-1$

$$r(x_j) = P_n(x_j) - P_{n-1}(x_j) = f(x_j) - f(x_j) = 0. \tag{25}$$

Therefore, $R(x)$ can be factored as

$$R(x) = c_n(x - x_0)(x - x_1)\cdots(x - x_{n-1}). \tag{26}$$

The constant c_n is called the *n-th divided difference* of f with respect to $x_0, x_1, ..., x_n$, and is usually denoted as $f[x_0, \ldots, x_n]$. Thus, we have

$$P_n(x) = P_{n-1}(x) + f[x_0, \ldots, x_n](x - x_0)(x - x_1)\cdots(x - x_{n-1}) \tag{27}$$

By the same argument, we have

$$P_{n-1}(x) = P_{n-2}(x) + f[x_0, \ldots, x_{n-1}](x - x_0)(x - x_1)\cdots(x - x_{n-2}), \tag{28}$$

etc. So we arrive at Newton's Form of $P_n(x)$:

(29)
$$P_n(x) = f[x_0] + f[x_0, x_1](x - x_0) + \ldots + f[x_0, \ldots, x_n](x - x_0)\cdots(x - x_{n-1}).$$

Note that for $n = 1$

$$\begin{aligned} P_1(x) &= f[x_0] + f[x_0, x_1](x - x_0) \\ P_1(x_0) &= f[x_0] = f(x_0) \\ P_1(x_1) &= f[x_0] + f[x_0, x_1](x_1 - x_0) = f(x_1) \end{aligned}$$

Therefore

$$f[x_0] = f(x_0) \tag{30}$$

$$f[x_0, x_1] = \frac{f(x_1) - f(x_0)}{x_1 - x_0}. \tag{31}$$

and

$$P_1(x) = f(x_0) + \frac{f(x_1) - f(x_0)}{x_1 - x_0}(x - x_0). \tag{32}$$

Define $f[x_j] = f(x_j)$ for $j = 0, 1, ...n$. The following identity will allow us to compute all the required divided differences.

Theorem 3.

$$f[x_0, x_1, ..., x_k] = \frac{f[x_1, x_2, ..., x_k] - f[x_0, x_1, ..., x_{k-1}]}{x_k - x_0}. \tag{33}$$

Proof. Let P_{k-1} be the interpolating polynomial of degree at most $k-1$ of f at $x_1, \ldots, x_k$ and Q_{k-1} the interpolating polynomial of degree at most $k-1$ of f at $x_0, \ldots, x_{k-1}$. Then

$$P(x) = P_{k-1}(x) + \frac{x - x_k}{x_k - x_0}[P_{k-1}(x) - Q_{k-1}(x)]. \tag{34}$$

is a polynomial of degree at most k and for $j = 1, 2, ...k-1$

$$P_k(x_j) = f(x_j) + \frac{x_j - x_k}{x_k - x_0}[f(x_j) - f(x_j)] = f(x_j).$$

Moreover, $P_k(x_0) = Q_{k-1}(x_0) = f(x_0)$ and $P_k(x_k) = P_{k-1}(x_k) = f(x_k)$. Therefore, P is the interpolation poly of f at $x_0, ..., x_k$. The leading order coefficient of $P_k(x)$ is $f[x_0, ..., x_k]$ and equating this with the leading order coefficient of P, $\frac{f[x_1,...,x_k] - f[x_0,x_1,...x_{k-1}]}{x_k - x_0}$, gives (33). $\square$

To use (33) to obtain the divided difference we construct a table. Setting $f_j = f(x_j)$ this divided difference table is built column by column as illustrated below for $n = 3$.

x_j	0th order	1th order	2th order	3th order
x_0	f_0			
		$f[x_0, x_1]$		
x_1	f_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_2		$f[x_1, x_2, x_3]$	
		$f[x_2, x_3]$		
x_3	f_3			

Example 3. *Let $f(x) = 1 + x^2$ and $x_j = j - 1$ for $j = 0, \ldots, 3$. Then*

x_j	*0th order*	*1th order*	*2th order*	*3th order*
0	*1*			
		$\frac{2-1}{1-0} = 1$		
1	2		$\frac{3-1}{2-0} = 1$	
		$\frac{6-2}{2-1} = 3$		*0*
2	5		$\frac{5-3}{3-1} = 1$	
		$\frac{10-5}{3-2} = 5$		
3	10			

so $P_3(x) = 1 + 1(x-0) + 1(x-0)(x-1) + 0(x-0)(x-1)(x-2) = 1 + x^2$.

After computing the divided differences, we need to evaluate P_n at a given point x. This can be done efficiently by suitably factoring it. For example, for $n = 3$ we have

$$\begin{aligned} P_3(x) &= c_0 + c_1(x - x_0) + c_2(x - x_0)(x - x_1) + c_3(x - x_0)(x - x_1)(x - x_2) \\ &= c_0 + (x - x_0)\{c_1 + (x - x_1)[c_2 + (x - x_2)c_3]\} \end{aligned}$$

For general n we can use the following *Horner*-like scheme to get $p = P_n(x)$:

```
p = c_n;
for k = n − 1 : 0
    p = c_k + (x − x_k) ∗ p;
end
```

Note that

$$P_k(x) = P_{k-1}(x) + f[x_0, ..., x_k](x - x_0) \cdots (x - x_{k-1}) \tag{35}$$

where $P_{k-1}(x)$ is a polynomial of degree $\leq k - 1$. Then, the leading order coefficient of $P_k(x)$ is $f[x_0, ..., x_k]$ and consequently $P_k^{(k)}(x) = k! f[x_0, ..., x_k]$ and thus

$$f[x_0, ..., x_k] = \frac{1}{k!} P_k^{(k)}(x). \tag{36}$$

6 Cauchy's Remainder

In the Introduction we proved that if x_0, x_1, and x are in $[a, b]$ then

$$f(x) - P_1(x) = \frac{1}{2} f''(\xi(x))(x - x_0)(x - x_1),$$

where $\xi(x) \in (a, b)$. The general result about the interpolation error is the following theorem:

Theorem 4. *Let* $f \in C^{n+1}[a, b]$, $x_0, x_1, ..., x_n, x$ *be contained in* $[a, b]$, *and* $P_n(x)$ *be the interpolation polynomial of degree* $\leq n$ *of* f *at* $x_0, ..., x_n$ *then*

$$f(x) - P_n(x) = \frac{1}{(n+1)!} f^{(n+1)}(\xi(x))(x - x_0)(x - x_1) \cdots (x - x_n), \tag{37}$$

where $\min\{x_0, \ldots, x_n, x\} < \xi(x) < \max\{x_0, \ldots, x_n, x\}$.

Proof. The right hand side of (37) is known as the Cauchy Remainder and the following proof is due to Cauchy.

For x equal to one of the nodes x_j the result is trivially true. Take x fixed not equal to any of the nodes and define

$$\phi(t) = f(t) - P_n(t) - [f(x) - P_n(x)]\frac{(t-x_0)(t-x_1)\cdots(t-x_n)}{(x-x_0)(x-x_1)\cdots(x-x_n)}. \tag{38}$$

Clearly, $\phi \in c^{n+1}[a,b]$ and vanishes at $t = x_0, x_1, ..., x_n, x$. That is, phi has at least $n+2$ zeros. Applying Rolle's Theorem $n+1$ times we conclude that there exists a point $\xi(x) \in (a,b)$ such that $\phi^{(n+1)}(\xi(x)) = 0$. Therefore,

$$0 = \phi^{(n+1)}(\xi(x)) = f^{(n+1)}(\xi(x)) - [f(x) - P_n(x)]\frac{(n+1)!}{(x-x_0)(x-x_1)\cdots(x-x_n)}$$

from which (37) follows. Note that the repeated application of Rolle's theorem implies that $\xi(x)$ is between $\min\{x_0, x_1, ..., x_n, x\}$ and $\max\{x_0, x_1, ..., x_n, x\}$.

$\square$

We can use the same argument to relate divided differences to the derivatives of f. Let $P_{n+1}(t)$ be the interpolating polynomial of f at $x_0, \ldots, x_n, x$. Then $f(t) - P_{n+1}(t)$ vanishes at $t = x_0, \ldots, x_n, x$. By repeated application of Rolle's theorem there is $\xi(x)$ between $\min\{x_0, x_1, ..., x_n, x\}$ and $\max\{x_0, x_1, ..., x_n, x\}$ such that

$$0 = f^{n+1}(\xi(x)) - P_{n+1}^{n+1}(\xi) = f^{n+1}(\xi(x)) - (n+1)! f[x_0, \ldots, x_n, x].$$

Therefore

$$f[x_0, ..., x_n, x] = \frac{f^{(n+1)}(\xi(x))}{(n+1)!}. \tag{39}$$

Similarly, if P_k is the interpolating polynomial of f at $x_0, x_1, ..., x_k$, then

$$f[x_0, ..., x_k] = \frac{1}{k!} f^{(k)}(\xi), \tag{40}$$

where $\min\{x_0, \ldots, x_k\} < \xi < \max\{x_0, \ldots, x_k\}$. Suppose that we now let $x_1, ..., x_k \to x_0$. Then $\xi \to x_0$ and

$$\lim_{x_1, \ldots, x_k \to x_0} f[x_0, ..., x_k] = \frac{1}{k!} f^{(k)}(x_0). \tag{41}$$

We can use this relation to define a divided difference where there are "coincident" nodes. For example $f[x_0, x_1]$ when $x_0 = x_1$ by $f[x_0, x_0] = f'(x_0)$, etc. This is going to be very useful for the following interpolation problem.

7 Hermite Interpolation

The Hermite interpolation problem is: given values of f and some of its derivatives at the nodes $x_0, x_1, ..., x_n$, find the interpolating polynomial of smallest degree interpolating those values. This polynomial is called the *Hermite Interpolation Polynomial* and can be obtained with a minor modification to the Newton's form representation.

For example: Suppose we look for a polynomial of P of lowest degree which satisfies the interpolation conditions:

$$\begin{aligned} P(x_0) &= f(x_0), \\ P'(x_0) &= f'(x_0), \\ P(x_1) &= f(x_1), \\ P'(x_1) &= f'(x_1). \end{aligned}$$

We can view this problem as a limiting case of polynomial interpolation of f at two pairs of coincident nodes, x_0, x_0, x_1, x_1 and we can use Newton's Interpolation form to obtain P. The table of divided differences, in view of (41), is

$$\begin{array}{c|cccc} x_0 & f(x_0) & & & \\ x_0 & f(x_0) & f'(x_0) & & \\ x_1 & f(x_1) & f[x_0, x_1] & f[x_0, x_0, x_1] & \\ x_1 & f(x_1) & f'(x_1) & f[x_0, x_1, x_1] & f[x_0, x_0, x_1, x_1] \end{array} \tag{42}$$

and

$$P(x) = f(x_0) + f'(x_0)(x - x_0) + f[x_0, x_0, x_1](x - x_0)^2 + f[x_0, x_0, x_1, x_1](x - x_0)^2(x - x_1).$$

Example 4. *Let $f(0) = 1$, $f'(0) = 0$ and $f(1) = \sqrt{2}$. Find the Hermite Interpolation Polynomial.*

We construct the table of divided differences as follows:

$$\begin{array}{c|ccc} 0 & \textcircled{1} & & \\ 0 & 1 & \textcircled{0} & \\ 1 & \sqrt{2} & \sqrt{2} - 1 & \textcircled{\sqrt{2} - 1} \end{array} \tag{43}$$

and therefore

$$P(x) = 1 + 0(x - 0) + (\sqrt{2} - 1)(x - 0)^2 = 1 + (\sqrt{2} - 1)x^2.$$