

# Interpolating polynomials and divided differences

Notes by G.J.O. Jameson

*Distinct points: the Lagrange form*

We shall take it as known that a polynomial of degree  $n$  has at most  $n$  distinct zeros (a proof is given in Lemma 1 below). Given  $n + 1$  distinct real numbers  $x_j$  and any numbers  $\alpha_j$  ( $0 \leq j \leq n$ ), there is a unique polynomial  $p$  of degree at most  $n$  satisfying  $p(x_j) = \alpha_j$  ( $0 \leq j \leq n$ ). The polynomial is unique, since if  $p_1$  and  $p_2$  were two such polynomials, then  $p_1 - p_2$  would be zero at each  $x_j$ : since it has degree at most  $n$ , it can only be zero. Existence can be deduced from the fact that the matrix with entries  $x_j^k$  ( $0 \leq j \leq n, 0 \leq k \leq n$ ) is non-singular, but it is easy to describe an explicit construction, as follows.

First, let

$$q(x) = (x - x_0)(x - x_1) \dots (x - x_n).$$

Note that  $q(x_j) = 0$  for all  $j$  and  $q(x)$  is of the form  $x^{n+1} + c_n x^n + \dots + c_0$ , so  $q^{(n+1)}(x) = (n+1)!$  for all  $x$ .

For each  $j$ , write  $N_j = \{0, 1, \dots, n\} \setminus \{j\}$ , and let

$$q_j(x) = \frac{q(x)}{x - x_j} = \prod_{k \in N_j} (x - x_k),$$

$$r_j(x) = \frac{q_j(x)}{q_j(x_j)} = \prod_{k \in N_j} \frac{x - x_k}{x_j - x_k}.$$

Then  $r_j$  is a polynomial of degree  $n$  and we have

$$r_j(x_k) = \begin{cases} 1 & \text{if } k = j, \\ 0 & \text{if } k \neq j. \end{cases}$$

So the required polynomial satisfying  $p(x_j) = \alpha_j$  for all  $j$  is:

$$p(x) = \sum_{j=0}^n \alpha_j r_j(x). \tag{1}$$

So, given a function  $f$ , there is a unique polynomial  $p$  of degree at most  $n$  such that  $p(x_j) = f(x_j)$  for each  $j$ . It is called the “polynomial interpolating  $f$  at  $x_0, x_1, \dots, x_n$ ”. Expression (1), with  $\alpha_j = f(x_j)$ , is called the *Lagrange form* of the interpolating polynomial.

Note that since  $q(x) = (x - x_j)q_j(x)$ , we have  $q'(x_j) = q_j(x_j)$ .

If  $n = 2$ , then of course  $p(x)$  is the linear function  $ax + b$  agreeing with  $f$  at  $x_0$  and  $x_1$ . Note that  $a = [f(x_1) - f(x_0)]/(x_1 - x_0)$ .

To find the polynomial in a particular case, it is usually simpler to solve for the coefficients, as in the next example. (A very effective alternative method will be described below.)

*Example 1.* To interpolate  $f(x) = 2^x$  at 0, 1, 2: let the required polynomial be  $a + bx + cx^2$ . Equating values at  $x = 0, 1$  and 2, we get the equations

$$a = 1, \quad a + b + c = 2, \quad a + 2b + 4c = 4,$$

hence  $b = c = \frac{1}{2}$ , so the polynomial is  $1 + \frac{1}{2}x + \frac{1}{2}x^2$ . Note that  $p(\frac{1}{2}) = 1\frac{3}{8}$ , while  $f(\frac{1}{2}) = \sqrt{2}$ .

If  $f(x) = x^k$ , where  $0 \leq k \leq n$ , then of course  $p(x)$  is also  $x^k$ . In terms of the functions  $r_j(x)$ , this says:

PROPOSITION 1. For  $0 \leq k \leq n$ , we have  $\sum_{j=0}^n x_j^k r_j(x) = x^k$ . In particular,  $\sum_{j=0}^n r_j(x) = 1$  for all  $x$ .

COROLLARY. 
$$\sum_{j=0}^n \frac{1}{q_j(x_j)} = 0.$$

*Proof.* This is the coefficient of  $x^n$  in  $\sum_{j=0}^n r_j(x)$  ( $= 1$ ). □

*Example 2.* Let  $f(x) = x^{n+1}$ . Then, with notation as above, the required  $p(x)$  is  $f(x) - q(x)$ , since this has degree at most  $n$  (the  $x^{n+1}$  term cancels) and agrees with  $f(x)$  at each  $x_j$ .

It is clear from (1) that the leading term of  $p(x)$  is  $a_n x^n$ , where

$$a_n = \sum_{j=0}^n \frac{f(x_j)}{q_j(x_j)}. \tag{2}$$

We shall see below that the leading coefficient  $a_n$  has particular significance.

### Repeated points

In this section, we show how, for functions with enough derivatives, the notion of interpolating polynomials can be extended to the situation where there are repetitions among the points  $x_j$ . This version will be used in the application to Simpson's rule. However, the reader who is so inclined could omit, or defer, this section after glancing at Lemma 1 and its corollaries (suitably simplified by ignoring the orders of the zeros).

Suppose, then, that the *distinct* elements among  $x_0, x_1, \dots, x_n$  are listed as  $y_0, y_1, \dots, y_r$ , and that  $k_j$  of the  $x_i$ 's are equal to  $y_j$ , so that  $\sum_{j=0}^r k_j = n + 1$ . Let  $f$  be a function having at least  $K - 1$  derivatives, where  $K = \max k_j$ . By a "polynomial interpolating  $f$  at  $x_0, x_1, \dots, x_n$ " we mean a polynomial  $p$  of degree at most  $n$  such that for each  $j$ ,  $p^{(k)}(y_j) = f^{(k)}(y_j)$  for  $0 \leq k \leq k_j - 1$ , requiring agreement of the function and the first  $k_j - 1$  derivatives at  $y_j$ . (The term *osculating polynomial* is sometimes used for  $p$ .) We now establish existence and uniqueness of such a polynomial.

Let  $f$  be a function having at least  $k$  derivatives. We say that  $f$  has a zero of order (or *multiplicity*)  $k$  at the point  $a$  if  $f^{(j)}(a) = 0$  for  $0 \leq j \leq k - 1$  and  $f^{(k)}(a) \neq 0$ . Note that  $f'$  then has a zero of order  $k - 1$  at  $a$  (this even works for  $k = 1$  if a "zero of order 0" is taken to mean a point that is not a zero!)

Given an interval  $I$ , we denote by  $Z(f, I)$  the number of zeros of  $f$  in  $I$ , *counted with their orders*. We will just write  $Z(f)$  for  $Z(f, \mathbb{R})$ . By Rolle's theorem, we have:

LEMMA 1. *For any function  $f$  (having enough derivatives) and any interval  $I$ , we have  $Z(f', I) \geq Z(f, I) - 1$ .*

*Proof.* Let  $f$  have a zero of order  $k_r$  at  $a_r$  ( $1 \leq r \leq n$ ), so that  $\sum_{r=1}^n k_r = Z(f, I)$ . Then  $f'$  has a zero of order  $k_r - 1$  at  $a_r$  (with the above comment about order 0): these add up to

$$\sum_{r=1}^n (k_r - 1) = Z(f, I) - n.$$

By Rolle's theorem,  $f'$  also has at least  $n - 1$  zeros in the gaps between the points  $x_r$ . Together, these two facts give  $Z(f', I) \geq Z(f, I) - 1$ .  $\square$

COROLLARY 1. *If  $Z(f, I) \geq n$ , then there exists  $\xi \in I$  such that  $f^{(n-1)}(\xi) = 0$ .*

COROLLARY 2. *Let  $p$  be a polynomial of degree  $n$ . Then  $Z(p) \leq n$ .*

*Proof.* Let the leading term be  $a_n x^n$ . If  $Z(p) \geq n + 1$ , then there exists  $\xi$  such that  $p^{(n)}(\xi) = 0$ . But this is not true, since  $p^{(n)}(x) = n!a_n$  for all  $x$ .  $\square$

Uniqueness of the interpolating polynomial now follows as before: if  $p_1$  and  $p_2$  were two such polynomials, then we would have  $Z(p_1 - p_2) \geq \sum_{j=0}^r k_j = n + 1$ . We now turn our attention to existence. The next Lemma follows at once from Leibniz's rule for the higher derivatives of a product:

LEMMA 2. *Let  $f(x) = (x - a)^k g(x)$ , where  $g$  has at least  $k$  derivatives. Then  $f$  has a zero of order at least  $k$  at  $a$ , and  $f^{(k)}(a) = k!g(a)$ .  $\square$*

In particular, if (as above)  $x_0, x_1, \dots, x_n$  consist of points  $y_j$  repeated  $k_j$  times (for  $0 \leq j \leq r$ ) and

$$q(x) = \prod_{i=0}^n (x - x_i) = \prod_{j=0}^r (x - y_j)^{k_j},$$

then  $q(x)$  has a zero of order  $k_j$  at  $y_j$ .

**THEOREM 2.** *Suppose that  $x_i$  ( $0 \leq i \leq n$ ) are points of an interval  $I$  (possibly with repetitions), and that  $f$  has at least  $K - 1$  derivatives on  $I$ , with  $K$  as above. Then there is a unique polynomial interpolating  $f$  at the points  $x_i$ .*

*Proof.* We prove the statement by induction on  $n$ . It is trivial for  $n = 0$  (and indeed for  $n = 1$ ). Assume it is correct for a certain  $n$ , and let points  $x_i$  ( $0 \leq i \leq n + 1$ ) be given, consisting of points  $y_j$  repeated  $k_j$  times as above. Let  $p(x)$  be the polynomial interpolating  $f$  at  $x_0, x_1, \dots, x_n$ , and let  $q(x) = \prod_{i=0}^n (x - x_i)$ . Let

$$p_1(x) = p(x) + a_{n+1}q(x),$$

where  $a_{n+1}$  is to be chosen (we use this notation because  $a_{n+1}$  is the coefficient of  $x^{n+1}$ ). Then  $p_1^{(k)}(y_j) = f^{(k)}(y_j)$  for  $0 \leq k \leq k_j - 1$ . We distinguish two cases.

*Case 1:*  $x_{n+1}$  different from all  $y_j$ . Then  $q(x_{n+1}) \neq 0$ , so we can choose  $a_{n+1}$  to ensure that  $p_1(x_{n+1}) = f(x_{n+1})$ .

*Case 2:*  $x_{n+1} = y_j$ , say, so  $k_j + 1$  of the extended list of  $x_i$ 's equal  $y_j$ . Since  $q^{(k_j)}(y_j) \neq 0$ , we can choose  $a_{n+1}$  to ensure that  $p_1^{(k_j)}(y_j) = f^{(k_j)}(y_j)$ , which is what is required.  $\square$

We cannot offer an explicit expression for  $p(x)$  corresponding to (1). However, Example 2 still applies, because the required derivatives agree.

The most extreme case, of course, is when *all* the points  $x_j$  coincide. It is then elementary that the interpolating polynomial is the truncated Taylor expansion

$$p_n(x) = \sum_{k=0}^n \frac{f^{(k)}(x_0)}{k!} (x - x_0)^k,$$

since  $p_n^{(k)}(x_0) = f^{(k)}(x_0)$  for  $0 \leq k \leq n$ .

*Example 3.* To interpolate  $f(x) = 2^x$  at 0, 1, 1, 2 (equally, at 0, 1, 2, 1). We saw in Example 1 that the polynomial interpolating  $f$  at 0, 1, 2 is  $p(x) = 1 + \frac{1}{2}x + \frac{1}{2}x^2$ . Let  $q(x) = x(x - 1)(x - 2)$ . Then the required polynomial is  $p_1(x) = p(x) + a_3q(x)$ , with  $a_3$  chosen so that  $p_1'(1) = f'(1) = 2 \log 2$ . Now  $p'(1) = \frac{3}{2}$  and  $q'(1) = -1$ , so  $a_3 = \frac{3}{2} - 2 \log 2$ . Note that  $q(\frac{1}{2}) = \frac{1}{2}(-\frac{1}{2})(-\frac{3}{2}) = \frac{3}{8}$ , hence  $p_1(\frac{1}{2}) \approx 1.4176$ .

*Estimation of the leading coefficient and the error*

We now apply Rolle's theorem (more exactly, Corollary 1 of Lemma 1) to give estimations for the leading coefficient  $a_n$  and the "error"  $f(x) - p(x)$ .

**THEOREM 3.** *Suppose that the points  $x_0, x_1, \dots, x_n$  (some possibly repeated) lie in an interval  $I$  and that  $f$  is  $n$  times differentiable on  $I$ . Let  $p(x)$  be the polynomial interpolating  $f$  at the points  $x_j$ , and let the leading term of  $p(x)$  be  $a_n x^n$ . Then there exists a point  $\xi$  in  $I$  such that*

$$a_n = \frac{f^{(n)}(\xi)}{n!}.$$

*Proof.* We have  $Z(f - p, I) \geq n + 1$ : this is obvious when the points  $x_j$  are distinct, and follows from our definition of the interpolating polynomial when there are repetitions. By Corollary 1 of Lemma 1, it follows that there exists  $\xi \in I$  such that  $f^{(n)}(\xi) - p^{(n)}(\xi) = 0$ . But  $p^{(n)}(x) = n!a_n$  for all  $x$ , so  $a_n = f^{(n)}(\xi)/n!$ .  $\square$

The case  $n = 1$  (with  $x_1 \neq x_0$ ) equates to the mean-value theorem, since  $a_1 = [f(x_1) - f(x_0)]/(x_1 - x_0)$ .

The estimation of  $f(x) - p(x)$  is derived by a slight elaboration of the same reasoning:

**THEOREM 4.** *Suppose that the points  $x_0, x_1, \dots, x_n$  (some possibly repeated) lie in an interval  $I$  and that  $f$  is  $n + 1$  times differentiable on  $I$ . Let  $p(x)$  be the polynomial interpolating  $f$  at the points  $x_j$ , and let  $q(x) = \prod_{j=0}^n (x - x_j)$ . Then, given a point  $x$  in  $I$ , there exists  $\xi$  in  $I$  such that*

$$f(x) - p(x) = \frac{1}{(n + 1)!} q(x) f^{(n+1)}(\xi).$$

*Proof.* Choose a point  $x^*$  in  $I$ , different from all the  $x_j$ . We will show that the given statement applies with  $x = x^*$ . Define

$$G(x) = f(x) - p(x) - kq(x),$$

with  $k$  chosen so that  $G(x^*) = 0$ , in other words,  $f(x^*) - p(x^*) = kq(x^*)$ . We need to show that  $k = f^{(n+1)}(\xi)/(n + 1)!$  for some  $\xi$ . Now  $Z(G, I) \geq n + 2$ . In the case where the points  $x_j$  are distinct, this is simply because  $G$  is zero at  $x_0, \dots, x_n$  and  $x^*$ . When there are repetitions (for readers interested in this case), it follows from the fact in the notation of Theorem 2, that  $q(x)$ , and hence also  $G(x)$ , has a zero of order at least  $k_j$  at  $y_j$  for each  $j$ . By Corollary 1, it follows that there is a point  $\xi$  in  $I$  such that  $G^{(n+1)}(\xi) = 0$ . But  $p^{(n+1)}(x) = 0$  (since  $p$  has degree at most  $n$ ) and  $q^{(n+1)}(x) = (n + 1)!$  for all  $x$ . So  $0 = G^{(n+1)}(\xi) = f^{(n+1)}(\xi) - (n + 1)!k$ , hence  $k = f^{(n+1)}(\xi)/(n + 1)!$ , as required.  $\square$

It makes sense for  $q(x)$  to appear in the estimation, since of course the error is 0 at each  $x_j$ . Example 2 is a special case of Theorem 4, since if  $f(x) = x^{n+1}$ , then  $f^{(n+1)}(\xi) = (n+1)!$  for all  $\xi$ .

When all the points  $x_j$  coincide, Theorem 4 reproduces the Lagrange form of Taylor's theorem.

*Example 4.* Revisit Examples 1 and 3. In Example 1,  $f^{(3)}(x) = 2^x(\log 2)^3$ , which is between  $(\log 2)^3$  and  $4(\log 2)^3$  for  $x$  in  $[0, 2]$ . Also,  $q(\frac{1}{2}) = \frac{3}{8}$ . So the bounds for  $f(\frac{1}{2}) - p(\frac{1}{2})$  given by Theorem 4 are  $\frac{1}{16}(\log 2)^3 \approx 0.021$  and  $\frac{1}{4}(\log 2)^3 \approx 0.083$ . As we saw, the actual value is  $\approx 0.039$ .

In Example 3, we apply  $f^{(4)}(x) = 2^x(\log 2)^4$ , and  $q(x)$  is  $x(x-1)^2(x-2)$ , so  $q(\frac{1}{2}) = \frac{3}{16}$ . The bounds are  $\frac{1}{128}(\log 2)^4 \approx 0.0018$  and  $\frac{1}{32}(\log 2)^4 \approx 0.0072$ . The actual value is  $\approx 0.0034$ .

*Application: error estimates for the trapezium rule and Simpson's rule*

(This section could be deferred.) The trapezium rule estimates  $\int_a^b f$  by  $T(f) = \frac{1}{2}(b-a)[f(a) + f(b)]$ , the integral of the linear function  $p(x)$  interpolating  $f$  at  $a$  and  $b$ . From Theorem 4, we can derive the following bounds for its error:

**PROPOSITION 5.** *Suppose that  $m \leq f''(x) \leq M$  for  $x \in [a, b]$ , and let  $T(f)$  be as above. Then*

$$\frac{1}{12}m(b-a)^3 \leq T(f) - \int_a^b f \leq \frac{1}{12}M(b-a)^3.$$

*Proof.* Let  $m \leq f''(x) \leq M$  for  $x \in [a, b]$ , and let  $p(x)$  be as above. Apply Theorem 4 with  $n = 1$ :  $q(x)$  is  $(x-a)(x-b)$ , so (reversing signs) we obtain

$$\frac{1}{2}m(x-a)(b-x) \leq p(x) - f(x) \leq \frac{1}{2}M(x-a)(b-x)$$

for  $x \in [a, b]$ . Writing  $b-a = h$ , we have

$$\int_a^b (x-a)(b-x) dx = \int_0^h y(h-y) dy = \left[ \frac{1}{2}hy^2 - \frac{1}{3}y^3 \right]_0^h = \frac{1}{6}h^3.$$

The stated inequalities follow. □

Note that when  $f''(x) \geq 0$  (so that  $f$  is convex) on  $[a, b]$ , this result reproduces the geometrically obvious fact that  $T(f) \geq \int_a^b f$ .

An alternative method for Proposition 5 is as follows: integration by parts shows that  $T(f) - \int_a^b f = \int_a^b (x-c)f'(x) dx$ , where  $c = \frac{1}{2}(a+b)$ . The result is then derived by integrating

by parts the other way round. Further development of this approach leads to the powerful process known as Euler-Maclaurin summation.

For the discussion of Simpson's rule, we denote the interval in question by  $I = [a - h, a + h]$ . The integral  $\int_I f$  is approximated by  $S(f, I) =: \frac{h}{3}[f(a - h) + 4f(a) + f(a + h)]$ . This equals the integral exactly if  $f$  is a quadratic or cubic polynomial, as one can easily check. So  $S(f, I)$  is the integral of (a) the quadratic interpolating  $f$  at  $a - h, a, a + h$ , or (b) the cubic interpolating  $f$  at  $a - h, a, a, a + h$ . (It is *not* the integral of the cubic interpolating  $f$  at four equally spaced points!)

In general, one would expect the error estimate derived from (b) to be sharper. This is indeed the one presented as the standard result in most books. The statement is as follows:

PROPOSITION 6. *Let  $I = [a - h, a + h]$ . Suppose that  $m \leq f^{(4)}(x) \leq M$  for  $x \in I$ , and let  $S(f, I)$  be as above. Then*

$$\frac{1}{90}mh^5 \leq S(f, I) - \int_I f \leq \frac{1}{90}Mh^5.$$

*Proof.* By considering  $f_1(x) = f(x - a)$ , we may assume that  $a = 0$ , so that  $I = [-h, h]$ . Then  $S(f, I) = \int_I p$ , where  $p$  is the cubic interpolating  $f$  at  $-h, 0, 0, h$ . Then

$$q(x) = (x + h)x^2(x - h) = x^2(x^2 - h^2).$$

Note that  $q(x) \leq 0$  on  $I$ . By Theorem 4,

$$\frac{1}{24}mx^2(h^2 - x^2) \leq p(x) - f(x) \leq \frac{1}{24}Mx^2(h^2 - x^2)$$

for  $x \in I$ . The statement now follows from the fact that

$$\int_{-h}^h x^2(h^2 - x^2) dx = \frac{2}{3}h^5 - \frac{2}{5}h^5 = \frac{4}{15}h^5. \quad \square$$

We remark that some books give proofs that are decidedly more complicated!

Let us at least mention the estimate derived from (a). For this, we assume that  $m' \leq f^{(3)}(x) \leq M'$  on  $I$ , and we have  $q(x) = x(x^2 - h^2)$ , which is positive on  $(-h, 0)$  and negative on  $(0, h)$ . By considering these intervals separately, one finds that

$$\left| S(f, I) - \int_I f \right| \leq \frac{1}{24}(M' - m')h^4.$$

*Example 5.* Let  $J = \int_0^2 2^x dx$ . This equals  $3/\log 2 \approx 4.3281$ . The approximation  $S$  given by Simpson's rule is  $\frac{1}{3}(1 + 4 \times 2 + 4) = 4\frac{1}{3}$ , so  $S - J \approx 0.0052$ . In the notation of Proposition 6,  $M = 4(\log 2)^4$  and  $m = (\log 2)^4$ , so the upper and lower estimates for

$S - J$  are  $\frac{M}{90} \approx 0.0103$  and  $\frac{m}{90} \approx 0.0026$ . Meanwhile, the estimates derived from (a) are  $\pm \frac{1}{8}(\log 2)^3 \approx \pm 0.0416$ . (However, there are cases where one of the estimates derived from (a) is actually better; the reader can verify that this occurs for the integral  $\int_1^3 (1/x) dx$ .)

### *Newton's form of the polynomial and divided differences*

Let  $p(x) = c_0 + c_1x + \cdots + c_nx^n$  be any polynomial of degree  $n$ , and let  $x_0, x_1, \dots, x_n$  be any list of points (possibly with repetitions). Then  $p(x)$  can be expressed in the form

$$p(x) = a_0 + a_1(x - x_0) + a_2(x - x_0)(x - x_1) + \dots + a_n(x - x_0)(x - x_1)\dots(x - x_{n-1}). \quad (3)$$

To demonstrate this, first equate coefficients of  $x^n$  to get  $a_n = c_n$ . Next, equate coefficients of  $x^{n-1}$ : we get  $c_{n-1} = a_{n-1} - a_n(x_0 + \cdots + x_{n-1})$ , which determines  $a_{n-1}$ . Continuing in the same way, we see that each  $a_k$  exists and is uniquely determined. At the opposite end, the value at  $x_0$  shows that  $a_0 = p(x_0)$ . Equating values at  $x_1$ , we then have  $a_0 + a_1(x_1 - x_0) = p(x_1)$ , which determines  $a_1$  if  $x_1 \neq x_0$ .

The expression (3) is called *Newton's form* of the polynomial. Clearly, it depends on a chosen order for the points  $x_j$ , and  $x_n$  does not appear at all.

We now consider expressing interpolating polynomials in this form. For now, we restrict the discussion to the case where the points  $x_j$  are distinct. As already mentioned,  $a_n$  is the coefficient of  $x^n$ , which is given by (2). We can actually derive an expression of this type for each  $a_k$ , because of the highly desirable property described in the next result.

**PROPOSITION 7.** *Let  $p(x)$  be the polynomial interpolating  $f$  at distinct points  $x_0, x_1, \dots, x_n$ , expressed as in (3). and let  $p_k(x)$  be the sum of the first  $k + 1$  terms (that is, as far as the term with coefficient  $a_k$ ). Then  $p_k(x)$  is the polynomial interpolating  $f$  at  $x_0, x_1, \dots, x_k$ .*

*Proof.* The polynomial  $p_k(x)$  has degree at most  $k$  and, for  $0 \leq j \leq k$ , we have  $p_k(x_j) = p(x_j) = f(x_j)$ , since all the subsequent terms in  $p(x)$  have  $(x - x_j)$  as a factor.  $\square$

It follows that each  $a_k$  in (3) is defined by (2), with  $k$  replacing  $n$ . We restate this more carefully. Our original notation  $N_j, q_j$  presupposed a fixed, unstated  $n$ . We now need to adopt the following more precise notation. Let distinct points  $x_0, x_1, \dots, x_k$  (with  $k \geq 1$ ) be given. For  $0 \leq j \leq k$ , let  $N_{k,j} = \{0, 1, \dots, k\} \setminus \{j\}$  and  $q_{k,j}(x) = \prod_{r \in N_{k,j}} (x - x_r)$ . What we have shown is that when  $p(x)$  is expressed as in (3), the coefficient  $a_k$  equates to the *divided difference*  $f[x_0, x_1, \dots, x_k]$ , defined by

$$f[x_0, x_1, \dots, x_k] = \sum_{j=0}^k \frac{f(x_j)}{q_{k,j}(x_j)}. \quad (4)$$



Also, for a single point,  $f[x_0]$  is defined to be  $f(x_0)$ . Note that  $f[x_0, x_1] = a_1 = [f(x_1) - f(x_0)]/(x_1 - x_0)$ .

Of course, Theorem 3 applies:  $f[x_0, x_1, \dots, x_k] = f^{(k)}(\xi)/k!$  for some  $\xi \in I$ .

With this notation in place, we can give a pleasantly simple expression for the difference between  $f(x)$  and  $p(x)$ :

PROPOSITION 8. *Let  $p(x)$  be the polynomial interpolating  $f(x)$  at distinct points  $x_0, x_1, \dots, x_n$ , and let  $q(x) = \prod_{j=0}^n (x - x_j)$ . Then for  $x$  different from all the  $x_j$ ,*

$$f(x) - p(x) = f[x_0, x_1, \dots, x_n, x]q(x).$$

*Proof.* Let  $x$  be given and write  $x = x_{n+1}$ . Let  $p_{n+1}(x)$  be the polynomial interpolating  $f$  at  $x_0, x_1, \dots, x_n, x_{n+1}$ . By Proposition 7,

$$p_{n+1}(x) = p(x) + f[x_0, \dots, x_n, x_{n+1}]q(x)$$

(for all  $x$ ). Apply this with  $x = x_{n+1}$ : since  $p_{n+1}(x_{n+1}) = f(x_{n+1})$ , we obtain the desired statement.  $\square$

This gives a second proof of Theorem 4 (at least for distinct points).

Both from (4), and from the fact that the interpolating polynomial does not depend on the order in which the points are listed, we have:

PROPOSITION 9. *If  $(y_0, y_1, \dots, y_n)$  is a permutation of  $(x_0, x_1, \dots, x_n)$ , then*

$$f[y_0, y_1, \dots, y_n] = f[x_0, x_1, \dots, x_n]. \quad \square$$

*Example 6.* Let  $f(x) = x^k$ , and let  $x_0, x_1, \dots, x_n$  be given, with  $n \geq k$ . Then the interpolating polynomial is  $x^k$  itself. So  $f[x_0, x_1, \dots, x_n]$  equals 0 if  $n > k$  and 1 if  $n = k$ .

We now show how divided differences of order  $n$  can be derived from those of order  $n - 1$ . This is very useful for actual calculation. Numerous different proofs can be found in the literature. We present two of them.

LEMMA 3. *Let points  $x_0, x_1, \dots, x_{n-2}, y, z$  be given (where  $n \geq 2$ ). Denote as follows the polynomials interpolating  $f$  at the points stated:*

$$\begin{aligned} p_y(x): & \quad \text{points } x_0, x_1, \dots, x_{n-2}, y; \\ p_z(x): & \quad \text{points } x_0, x_1, \dots, x_{n-2}, z; \\ p_{y,z}(x): & \quad \text{points } x_0, x_1, \dots, x_{n-2}, y, z. \end{aligned}$$

Then

$$p_{y,z}(x) = \frac{(x-z)p_y(x) - (x-y)p_z(x)}{y-z}.$$

*Proof.* The stated polynomial has degree  $n$  and agrees with  $f$  at each  $x_j$ ,  $y$  and  $z$ .  $\square$

PROPOSITION 10. *We have*

$$f[x_0, x_1, \dots, x_{n-1}, x_n] = \frac{f[x_0, \dots, x_{n-2}, x_n] - f[x_0, \dots, x_{n-2}, x_{n-1}]}{x_n - x_{n-1}}. \quad (5)$$

*Note.* Because of the symmetry of divided differences, we can present (5) in various alternative ways, for example

$$f[x_0, x_1, \dots, x_n] = \frac{f[x_1, x_2, \dots, x_n] - f[x_0, x_1, \dots, x_{n-1}]}{x_n - x_0}. \quad (6)$$

*Proof 1.* In Lemma 3, take  $y = x_n$  and  $z = x_{n-1}$ . Equating the coefficients of  $x^n$ , we obtain the stated identity.  $\square$

*Proof 2.* We prove the statement in the form (6). Express  $p(x)$  as in (3). But also, taking the points in reverse order, we can write

$$p(x) = b_0 + b_1(x - x_n) + \dots + b_{n-1}(x - x_n) \dots (x - x_2) + b_n(x - x_n)(x - x_{n-1}) \dots (x - x_1).$$

We have  $a_n = b_n = f[x_0, x_1, \dots, x_n]$ , also

$$a_{n-1} = f[x_0, x_1, \dots, x_{n-1}], \quad b_{n-1} = f[x_n, x_{n-1}, \dots, x_1].$$

But, equating the coefficients of  $x^{n-1}$ , we have

$$a_{n-1} - a_n(x_0 + x_1 + \dots + x_{n-1}) = b_{n-1} - b_n(x_n + x_{n-1} + \dots + x_1),$$

hence (noting that  $b_n = a_n$ )  $a_n(x_n - x_0) = b_{n-1} - a_{n-1}$ .  $\square$

Some writers prefer to use (5), together with  $f[x_0] = f(x_0)$ , as a recursive *definition* of divided differences.

We can use Proposition 10 to calculate divided differences successively. For example,

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

The calculations can be tabulated in the following way:

$$\begin{array}{ccccccc} x_0 & f(x_0) & & & & & \\ & & f[x_0, x_1] & & & & \\ x_1 & f(x_1) & & f[x_0, x_1, x_2] & & & \\ & & f[x_1, x_2] & & f[x_0, x_1, x_2, x_3] & & \\ x_2 & f(x_2) & & f[x_1, x_2, x_3] & & & \\ & & f[x_2, x_3] & & & & \\ x_3 & f(x_3) & & & & & \end{array}$$

*Example 7.* To interpolate  $f(x) = 2^x$  at 0, 1, 2, 3. We construct a table as above. The columns labelled  $f_2, f_3, f_4$  are the divided differences over sets of 2, 3 and 4 points respectively.

$x_j$	$f(x_j)$	$f_2$	$f_3$	$f_4$
0	1			
1	2	1		
2	4	2	$\frac{1}{2}$	
3	8	4	1	$\frac{1}{6}$

The successive divided differences involving  $x_0$  are seen on the top sloping line: 1, 1,  $\frac{1}{2}$ ,  $\frac{1}{6}$ , so the Newton form for the polynomial (with the points in this order) is

$$1 + x + \frac{1}{2}x(x - 1) + \frac{1}{6}x(x - 1)(x - 2),$$

which one can rewrite (if desired) as  $1 + \frac{5}{6}x + \frac{1}{6}x^3$ . The method of solving for coefficients would have been distinctly more laborious! The first three terms give the polynomial  $1 + \frac{1}{2}x + \frac{1}{2}x^2$  interpolating  $2^x$  at the points 0, 1, 2, as previously found in Example 1. Also, ignoring the point 0, we can read off the polynomial interpolating  $2^x$  at 1, 2, 3:  $2 + 2(x - 1) + (x - 1)(x - 2)$ .

We finish this section with a pleasant result on substitution of divided differences.

PROPOSITION 11. *Let the points  $x_i$  ( $0 \leq i \leq k$ ) and  $y_j$  ( $0 \leq j \leq r$ ) be distinct, and let  $g(x) = f[y_0, y_1, \dots, y_r, x]$ . Then*

$$g[x_0, x_1, \dots, x_k] = f[y_0, \dots, y_r, x_0, \dots, x_k].$$

*Proof.* Induction on  $k$ . The case  $k = 0$  is the definition of  $g$ . Assume the statement true for  $k - 1$  (i.e. for sets of  $k$  points). By Proposition 10 and the induction hypothesis,

$$\begin{aligned} g[x_0, x_1, \dots, x_k] &= \frac{g[x_1, x_2, \dots, x_k] - g[x_0, x_1, \dots, x_{k-1}]}{x_k - x_0} \\ &= \frac{f[y_0, \dots, y_r, x_1, x_2, \dots, x_k] - f[y_0, \dots, y_r, x_0, x_1, \dots, x_{k-1}]}{x_k - x_0} \\ &= f[y_0, \dots, y_r, x_0, \dots, x_k]. \quad \square \end{aligned}$$

### *Repeated points*

We have already defined what we mean by the interpolating polynomial in the case when points are repeated, and shown that it exists (Theorem 2). We now show how the

notion of divided differences and the construction of the Newton form can be adapted, without too much trouble, to deal with this case.

First consider the extreme case when all the points coincide:  $x_j = x_0$  for  $0 \leq j \leq n$ . Suppose that  $f^{(n)}$  exists and is continuous at  $x_0$ . It is clear from Proposition 3 that if we define

$$f[x_0, x_0, \dots, x_0] = \frac{f^{(n)}(x_0)}{n!}, \quad (7)$$

then we will have extended the definition of the divided difference in a way that makes it continuous at the point  $(x_0, x_0, \dots, x_0)$ . Furthermore, the resulting Newton form

$$f[x_0] + f[x_0, x_0](x - x_0) + \dots + f[x_0, x_0, \dots, x_0](x - x_0)^n = \sum_{k=0}^n \frac{f^{(k)}(x_0)}{k!} (x - x_0)^k,$$

is the polynomial interpolating  $f$  at  $x_0, \dots, x_0$ .

Now suppose that the sequence  $x_0, x_1, \dots, x_n$  comprises  $k_0$  repetitions of  $y_0$  followed by  $k_1$  repetitions of  $y_1$ , up to  $k_r$  repetitions of  $y_r$ . For the moment, it is important to keep the repeated terms together. Having defined expressions of the form  $f[y_j, \dots, y_j]$  by (7), we now complete a difference table as before. In other words, having defined divided differences of length  $k$ , those of length  $k + 1$  are defined by

$$f[x_0, x_1, \dots, x_k] = \frac{f[x_1, x_2, \dots, x_k] - f[x_0, x_1, \dots, x_{k-1}]}{x_k - x_0} \quad (8)$$

whenever the  $x_j$  do not all coincide, so that  $x_k \neq x_0$ . It is clear that divided differences, defined this way, are *continuous* functions of the variables, given continuity of enough derivatives of  $f$ . Also, an easy induction shows that  $f[x_k, x_{k-1}, \dots, x_0] = f[x_0, x_1, \dots, x_k]$ .

LEMMA 4. *Proposition 7 still applies when there are repetitions.*

*Proof.* Let  $p(x)$ , expressed as in (3), be the polynomial interpolating  $f$  at  $x_0, x_1, \dots, x_n$ , and let  $p_{n-1}(x)$  be formed from  $p(x)$  by leaving out the last term  $a_n q_n(x)$ , where  $q_n(x) = (x - x_0) \dots (x - x_{n-1})$ . We show that  $p_{n-1}(x)$  interpolates  $f$  at  $x_0, x_1, \dots, x_{n-1}$ . Suppose that  $k_j$  of the terms  $x_0, x_1, \dots, x_n$  equal  $y_j$ . Then  $f - p$  has a zero of order at least  $k_j$  at  $y_j$ . If  $x_n \neq y_j$ , then  $q_n$  has a zero of order at least  $k_j$  at  $y_j$ , and hence  $f - p_{n-1}$  does so. If  $x_n = y_j$ , then  $q_n$ , and hence  $f - p_{n-1}$ , has a zero of order at least  $k_j - 1$  at  $y_j$ . In both cases, this is what is required.  $\square$

PROPOSITION 12. *Let  $x_0, x_1, \dots, x_n$  be any list of points, possibly with repetitions, in an interval  $I$ , and suppose that  $f$  has continuous  $n$ th derivative on  $I$ . Then the polynomial interpolating  $f$  at these points is given by (3), with  $a_k = f[x_0, x_1, \dots, x_k]$ , as just defined.*

*Proof.* We prove the statement by induction. It is trivial for  $n = 0$  (and almost trivial for  $n = 1$ ). Assume that it is correct for  $n - 1$ , and let  $x_0, x_1, \dots, x_n$  be given. The case where  $x_j = x_0$  for all  $j$  has been established above, so we suppose that this is not the case. Let  $p(x)$ , expressed as in (3), be the polynomial interpolating  $f$  at  $x_0, x_1, \dots, x_n$ . We have to show that  $a_n = f[x_0, x_1, \dots, x_n]$ .

We do this by adapting the second proof of Proposition 10. In the notation used there, we have again  $a_n = b_n$  and, by Lemma 4 and the induction hypothesis,

$$a_{n-1} = f[x_0, x_1, \dots, x_{n-1}],$$

$$b_{n-1} = f[x_n, x_{n-1}, \dots, x_1] = f[x_1, x_2, \dots, x_n].$$

Exactly as before, we have  $a_n(x_n - x_0) = b_{n-1} - a_{n-1}$ . By (8), it follows that  $a_n = f[x_0, x_1, \dots, x_n]$ .  $\square$

*Example 8.* To interpolate  $f(x) = 1/x$  at 1, 2, 2, 2, 3. Then  $f[2, 2] = f'(2) = -\frac{1}{4}$  and  $f[2, 2, 2] = \frac{1}{2}f''(2) = \frac{1}{8}$ . Entering these values in the table, we obtain:

$x_j$	$f(x_j)$	$f_2$	$f_3$	$f_4$	$f_5$
1	1				
2	$\frac{1}{2}$	$-\frac{1}{2}$			
2	$\frac{1}{2}$	$-\frac{1}{4}$	$\frac{1}{4}$		
2	$\frac{1}{2}$	$-\frac{1}{4}$	$\frac{1}{8}$	$-\frac{1}{8}$	
2	$\frac{1}{2}$	$-\frac{1}{4}$	$\frac{1}{12}$	$-\frac{1}{24}$	$\frac{1}{24}$
3	$\frac{1}{3}$	$-\frac{1}{6}$			

So the polynomial is

$$1 - \frac{1}{2}(x - 1) + \frac{1}{4}(x - 1)(x - 2) - \frac{1}{8}(x - 1)(x - 2)^2 + \frac{1}{24}(x - 1)(x - 2)^3.$$

The reader may care to repeat Example 3 in this style.

For Proposition 12, we did not need full-scale symmetry of divided differences, only reversal of the order. However, a simple continuity argument shows that full-scale symmetry still applies:

**PROPOSITION 13.** *Proposition 9 still applies when there are repetitions: if  $(y_0, y_1, \dots, y_n)$  is a permutation of  $(x_0, x_1, \dots, x_n)$ , then*

$$f[y_0, y_1, \dots, y_n] = f[x_0, x_1, \dots, x_n].$$

*Proof.* Given a point  $\mathbf{x} = (x_0, x_1, \dots, x_n)$  of  $\mathbb{R}^{n+1}$ , there are clearly points  $\mathbf{x}^{(k)}$  of  $\mathbb{R}^{n+1}$  that converge to  $\mathbf{x}$  as  $k \rightarrow \infty$  and have all components distinct. Let  $\mathbf{y}^{(k)}$  be formed from  $\mathbf{x}^{(k)}$  by the corresponding permutation. By Proposition 9, we have (with obvious notation)  $f[\mathbf{y}^{(k)}] = f[\mathbf{x}^{(k)}]$  for each  $k$ . Since the divided difference is a continuous function on  $\mathbb{R}^{n+1}$ , we have  $f[\mathbf{y}] = \lim_{k \rightarrow \infty} f[\mathbf{y}^{(k)}]$  and similarly for  $\mathbf{x}$ , so  $f[\mathbf{y}] = f[\mathbf{x}]$ .  $\square$

In the same way, Propositions 8 and 11 can be freed from the requirement that the points are distinct.

### *The integral expression for divided differences*

(The reader is free to defer this section, or leave it out.) There is an explicit expression for divided differences in the form of a repeated integral. This is of interest in theory (especially in the case of repeated points), but distinctly less useful for actual computation than the process described above.

PROPOSITION 14. *Suppose that  $x_j$  ( $0 \leq j \leq n$ ) are points of an interval  $I$  (possibly repeated), and that  $f$  has continuous  $n$ th derivative on  $I$ . Then  $f[x_0, x_1, \dots, x_n]$  equals*

$$\int_0^1 dt_1 \int_0^{t_1} dt_2 \cdots \int_0^{t_{n-1}} f^{(n)}[x_0 + (x_1 - x_0)t_1 + \cdots + (x_n - x_{n-1})t_n] dt_n. \quad (9)$$

*Proof.* We prove the statement for the case when the points are distinct. The case where there are repeated points then follows by continuity as in Proposition 13, since it is clear that the integral defines a continuous function of the variables  $x_j$ . First, we prove the case  $n = 1$ . The stated integral is then

$$I_1 = \int_0^1 f'[x_0 + (x_1 - x_0)t_1] dt_1.$$

The substitution  $x_0 + (x_1 - x_0)t_1 = u$  gives

$$I_1 = \frac{1}{x_1 - x_0} \int_{x_0}^{x_1} f'(u) du = \frac{f(x_1) - f(x_0)}{x_1 - x_0} = f[x_0, x_1].$$

Assume the statement correct for sets of  $n$  points. Take  $x_0, x_1, \dots, x_n$  and let  $I_n$  be the integral stated. For the integration with respect to  $t_n$ , substitute

$$x_0 + (x_1 - x_0)t_1 + \cdots + (x_{n-1} - x_{n-2})t_{n-1} + (x_n - x_{n-1})t_n = u.$$

The limits of integration for  $u$  are

$$u_0 = x_0 + (x_1 - x_0)t_1 + \cdots + (x_{n-1} - x_{n-2})t_{n-1},$$

$$u_1 = x_0 + (x_1 - x_0)t_1 + \cdots + (x_n - x_{n-2})t_{n-1}.$$

The transformed integral is

$$\frac{1}{x_n - x_{n-1}} \int_{u_0}^{u_1} f^{(n)}(u) du = \frac{1}{x_n - x_{n-1}} [f^{(n-1)}(u_1) - f^{(n-1)}(u_0)].$$

By the induction hypothesis and Proposition 10, we now have

$$I_n = \frac{f[x_0, \dots, x_{n-2}, x_n] - f[x_0, \dots, x_{n-2}, x_{n-1}]}{x_n - x_{n-1}} = f[x_0, x_1, \dots, x_{n-1}, x_n]. \quad \square$$

This gives an alternative proof of Proposition 3, since

$$\int_0^1 dt_1 \int_0^{t_1} dt_2 \cdots \int_0^{t_{n-1}} dt_n = \frac{1}{n!}.$$

There is also an explicit expression for divided differences (with repeated points) in terms of partial derivatives. We state it without proof: *Let the list  $x_0, x_1, \dots, x_n$  comprise  $k_j + 1$  repetitions of  $y_j$  for  $0 \leq j \leq r$ . Then*

$$f[x_0, x_1, \dots, x_n] = \frac{1}{k_0! \cdots k_r!} \frac{\partial^{k_0}}{\partial y_0^{k_0}} \frac{\partial^{k_r}}{\partial y_r^{k_r}} f[y_0, y_1, \dots, y_r].$$

*Equally spaced points: forward differences*

Suppose that the points  $x_j$  are equally spaced, so that (for some  $h > 0$ ),  $x_j = x_0 + jh$  for each  $j$ . The divided differences can then be expressed in terms of the *forward difference* operator  $\Delta$ , defined as follows (for a chosen  $h$ ):

$$(\Delta f)(x) = f(x + h) - f(x),$$

and  $\Delta^n f = \Delta(\Delta^{n-1} f)$ , so that

$$(\Delta^n f)(x) = (\Delta^{n-1} f)(x + h) - (\Delta^{n-1} f)(x).$$

Hence, for example,  $(\Delta^2 f)(x) = f(x + 2h) - 2f(x + h) + f(x)$ . Clearly,  $(\Delta f)(x_0) = hf[x_0, x_1]$ , where  $x_1 = x_0 + h$ .

**PROPOSITION 15.** *Let  $x_j = x + jh$  ( $0 \leq j \leq n$ ). Then*

$$(\Delta^n f)(x) = n!h^n f[x_0, x_1, \dots, x_n].$$

*Proof.* By induction. The case  $n = 1$  is immediate, as above. Assume the statement holds for  $n - 1$ . Then, by Proposition 10,

$$\begin{aligned} (\Delta^n f)(x) &= (\Delta^{n-1} f)(x_1) - (\Delta^{n-1} f)(x_0) \\ &= (n-1)!h^{n-1}(f[x_1, x_2, \dots, x_n] - f[x_0, x_1, \dots, x_{n-1}]) \\ &= (n-1)!h^{n-1}nhf[x_0, x_1, \dots, x_n] \\ &= n!h^n f[x_0, x_1, \dots, x_n]. \end{aligned}$$

PROPOSITION 16.  $(\Delta^n f)(x) = \sum_{j=0}^n (-1)^j \binom{n}{j} f(x_j)$ , where  $x_j = x + jh$ .

*Proof.* By Proposition 15,

$$(\Delta^n f)(x) = n!h^n \sum_{j=0}^n \frac{f(x_j)}{q_j(x_j)}.$$

Now for fixed  $j$ ,  $\prod_{i < j} (x_j - x_i) = j!h^j$  and  $\prod_{k > j} (x_j - x_k) = (-1)^{n-j} (n-j)!h^{n-j}$ , so

$$q_j(x_j) = h^n (-1)^{n-j} j! (n-j)!.$$

The stated equality follows. □

Propositions 16 and 3 give at once:

PROPOSITION 17. *If  $f$  is  $n$  times differentiable on  $[x, x + nh]$ , then there exists  $\xi$  in  $(x, x + nh)$  such that  $(\Delta^n f)(x) = h^n f^{(n)}(\xi)$ .* □

*Alternative direct proof.* By induction. The case  $n = 1$  is the mean-value theorem. Assume the statement true for a certain  $n$ . Then  $\Delta^{n+1} f = \Delta^n g$ , where  $g = \Delta f$ , so  $g(x) = f(x+h) - f(x)$ . By the induction hypothesis, there exists  $\eta$  in  $[x, x+nh]$  such that  $(\Delta^n g)(x) = g^{(n)}(\eta) = f^{(n)}(\eta+h) - f^{(n)}(\eta)$ . By the mean-value theorem again, this equals  $f^{(n+1)}(\xi)$  for some  $\xi$  in  $(\eta, \eta+h)$ . □

One can give a direct proof of Proposition 16 in similar fashion.

The integral expression for forward differences is pleasantly simple: each integration is now on the fixed interval  $[0, h]$ :

PROPOSITION 18. *Suppose that  $f$  is  $n$  times differentiable on  $[x, x + nh]$ . Then*

$$(\Delta^n f)(x) = \int_0^h dt_1 \int_0^h dt_2 \dots \int_0^h f^{(n)}(x + t_1 + \dots + t_n) dt_n.$$

*Proof.* The case  $n = 1$  is correct, since it says

$$\int_0^h f'(x + t_1) dt_1 = f(x+h) - f(x).$$

Assuming the statement correct for  $n$ , we have

$$(\Delta^{n+1} f)(x) = (\Delta^n f)(x+h) - (\Delta^n f)(x) = \int_0^h dt_1 \int_0^h dt_2 \dots \int_0^h G(t_n) dt_n,$$

where

$$\begin{aligned} G(t_n) &= f^{(n)}(x+h+t_1+\dots+t_n) - f^{(n)}(x+t_1+\dots+t_n) \\ &= \int_0^h f^{(n+1)}(x+t_1+\dots+t_n+t_{n+1}) dt_{n+1}. \end{aligned}$$



Substituting this, we obtain the required formula for the case  $n + 1$ . □

### *Uniform approximation and Chebyshev polynomials*

This section of our notes has a slightly more advanced flavour, but most of it should still be accessible to readers with a basic grounding in Real Analysis (at one point, we use a standard result from Complex Analysis).

Interpolating polynomials of degree  $n$  are determined by the choice of the points  $x_j$ . How can we evaluate how well  $p(x)$  approximates to  $f(x)$ , and how can we choose the points  $x_j$  to make this approximation as good as possible?

Different measures of approximation are possible, but we will adopt the simple-minded one of uniform approximation, measured by the maximum absolute value on the given interval  $[a, b]$ . More exactly, define

$$\|f\|_\infty = \sup\{|f(x)| : a \leq x \leq b\}$$

(this is standard notation, the “supremum norm”). The degree of approximation to  $f(x)$  by  $p(x)$  is then measured by  $\|f - p\|_\infty$ . As the reader may know, one says that  $p_n \rightarrow f$  as  $n \rightarrow \infty$  *uniformly* on  $[a, b]$  if  $\|f - p_n\|_\infty \rightarrow 0$  as  $n \rightarrow \infty$ .

Our results will require the existence of higher derivatives, so they have no application, for example, to the theorem of Weierstrass stating that all continuous functions can be uniformly approximated by polynomials (for this, we recommend the elegant proof by Bernstein polynomials). For functions represented by power series, approximating polynomials are already provided by truncations of the series; we shall see that in some cases, better approximations are delivered by interpolating polynomials of the same degree.

Recall that we have an expression for  $f(x) - p(x)$  from Theorem 4:

$$f(x) - p(x) = \frac{1}{(n+1)!} f^{(n+1)}(\xi) q(x)$$

for some  $\xi$  in  $[a, b]$ , where  $q(x) = \prod_{j=0}^n (x - x_j)$ . In Example 2, we saw that when  $f(x)$  is  $x^{n+1}$ , the error is exactly  $q(x)$ . So our problem is essentially to choose the points  $x_j$  so as to minimise  $\|q\|_\infty$ . For the moment, we take  $[a, b]$  to be  $[-1, 1]$  and replace  $n + 1$  by  $n$ . One might expect the objective to be achieved by taking equally spaced points, but this is not the case! Furthermore, it is rather remarkable that the desired  $q(x)$  can be identified explicitly: a beautiful piece of reasoning shows that it is the *Chebyshev polynomial*  $T_n(x)$ . These polynomials are defined, for each  $n$ , by the identity

$$\cos nt = 2^{n-1} T_n(\cos t).$$

For example, since  $\cos 3t = 4 \cos^3 t - 3 \cos t$ , we have  $T_3(x) = x^3 - \frac{3}{4}x$ . For present purposes, we do not need to know anything about these polynomials except that they exist, and that  $T_n$  is monic with degree  $n$  (easily proved by induction, using the identity  $\cos(n+1)t + \cos(n-1)t = 2 \cos nt \cos t$ ).

Every  $x$  in  $[-1, 1]$  is expressible as  $\cos t$ , and  $|\cos nt| \leq 1$  for all  $t$ , so  $\|T_n\|_\infty = 2^{-(n-1)}$ . (Of course, this says nothing about the behaviour of  $T_n$  outside  $[-1, 1]$ .) Also, since  $\cos k\pi = (-1)^k$  and  $\cos(k + \frac{1}{2})\pi = 0$  for integers  $k$ , we have:

$$\text{if } y_k = \cos \frac{k\pi}{n}, \text{ then } T_n(y_k) = (-1)^k 2^{-(n-1)},$$

$$\text{if } x_j = \cos \frac{(j + \frac{1}{2})\pi}{n}, \text{ then } T_n(x_j) = 0, \text{ hence } T_n(x) = (x - x_0)(x - x_1) \dots (x - x_{n-1}).$$

**THEOREM 19.** *Let  $\|q\|_\infty = \sup\{|q(x)| : -1 \leq x \leq 1\}$ . Among all monic polynomials  $q$  of degree  $n$ ,  $\|q\|_\infty$  is least when  $q = T_n$ , and then  $\|q\|_\infty = 2^{-(n-1)}$ . So if  $q(x)$  is expressed as  $\prod_{j=0}^{n-1} (x - x_j)$ , then  $\|q\|_\infty$  is least when  $x_j = \cos[(j + \frac{1}{2})\pi/n]$ .*

*Proof.* Write  $2^{-(n-1)} = \alpha$ , so that  $T_n(y_k) = (-1)^k \alpha$  for  $0 \leq k \leq n$ . Note that the points  $y_k$  are in decreasing order, with  $y_0 = 1$ ,  $y_n = -1$ . Suppose that  $|q(y_k)| < \alpha$  for each  $k$ . Then  $T_n(y_k) - q(y_k)$  is strictly positive for *even*  $k$  (since then  $T_n(y_k) = \alpha$ ) and strictly negative for *odd*  $k$ . By the intermediate value theorem, it follows that  $T_n - q$  has a zero in each open interval  $(y_{k+1}, y_k)$ , hence at least  $n$  zeros in total. But this is impossible, since  $T_n - q$  is a polynomial of degree at most  $n - 1$  (the  $x^n$  term cancels). So in fact  $|q(y_k)| \geq \alpha$  for some  $k$ , hence  $\|q\| \geq \alpha = \|T_n\|_\infty$ .  $\square$

An obvious restatement of this result is: if  $f_n(x) = x^n$ , then the best uniform approximation to  $f_n(x)$  on  $[-1, 1]$  by a polynomial  $p_{n-1}$  of degree at most  $n - 1$  is given by  $p_{n-1}(x) = x^n - T_n(x)$ , and then  $\|f_n - p_{n-1}\| = 2^{-(n-1)}$ .

To transfer Theorem 19 to a general interval  $[a, b]$ , perform the substitution  $x = \frac{1}{2}(b - a)t + \frac{1}{2}(a + b)$ : when  $x$  goes from  $a$  to  $b$ ,  $t$  goes from  $-1$  to  $1$ . If  $x_j$  is the point corresponding to  $t_j = \cos[(j + \frac{1}{2})\pi/n]$ , then  $x - x_j = \frac{1}{2}(b - a)(t - t_j)$ , so

$$\prod_{j=0}^{n-1} (x - x_j) = \frac{1}{2^n} (b - a)^n T_n(t) :$$

denote this polynomial by  $\tilde{T}_n(x)$ : we call it the “transferred Chebyshev polynomial”, and the points  $x_j$  the “Chebyshev points”. The conclusion is:

**COROLLARY.** *Let  $\|q\|_\infty = \sup\{|q(x)| : a \leq x \leq b\}$ . Among all monic polynomials  $q$  of degree  $n$ ,  $\|q\|_\infty$  is least when  $q = \tilde{T}_n$ , and then  $\|q\|_\infty = (b - a)^n / 2^{2n-1}$ .*

Inserted into Theorem 4 (still with  $n + 1$  replaced by  $n$ ), this gives at once:

**THEOREM 20.** *Suppose that  $|f^{(n)}(x)| \leq M_n$  on  $[a, b]$ , and let  $p_{n-1}$  be the polynomial interpolating  $f$  at the Chebyshev points  $x_j$  ( $0 \leq j \leq n - 1$ ). Let  $\|f\|_\infty = \sup\{|f(x)| : a \leq x \leq b\}$ . Then*

$$\|f - p_{n-1}\|_\infty \leq \frac{(b-a)^n M_n}{2^{2n-1} n!}. \quad \square$$

We give two examples comparing this with polynomials obtained by truncating power series.

*Example 9.* Consider  $f(x) = e^x$  on  $[0, 1]$ . Let  $e_n(x)$  be the truncation  $\sum_{r=0}^n \frac{x^r}{r!}$  of the series. Clearly,  $\|f - e_n\|_\infty > 1/(n+1)!$  (it is also less than  $1/[n!n]$ ). Let  $p_n$  defined as in Theorem 20. Since  $f^{(n)}(x) = e^x$  for all  $n$ , we have  $M_n = e$ , hence  $\|f - p_n\|_\infty \leq e/[2^{2n+1}(n+1)!]$ .

*Example 10.* Let  $f(x) = 1/(1+x)$ . The geometric series  $\sum_{n=0}^\infty (-1)^n x^n$  converges to  $f(x)$  only for  $|x| < 1$ , and if  $s_{n-1}(x) = \sum_{r=0}^{n-1} (-1)^r x^r$ , then  $|f(x) - s_{n-1}(x)| = x^n/(1+x)$ . So on the interval  $[0, a]$ , we have  $\|f - s_{n-1}\|_\infty = a^n/(1+a)$ . Now  $|f^{(n)}(x)| = n!/(1+x)^{n+1}$ , so  $M_n = n!$  and  $\|f - p_{n-1}\|_\infty = a^n/2^{2n-1}$ . Note that this tends to zero if  $a < 4$ . If we modify to the interval  $[1, a+1]$ , we have  $M_n = n!/2^{n+1}$ , hence  $\|f - p_{n-1}\|_\infty = a^n/2^{3n}$ , which tends to zero if  $a < 8$ .

To address the question of uniform convergence in general, we record the following obvious consequence of Theorem 20.

**COROLLARY.** *In the situation of Theorem 20, if for some  $M$ , we have  $M_n \leq Mn!/r^n$  for all  $n \geq 1$ , where  $r > \frac{1}{4}(b-a)$ , then  $p_n \rightarrow f$  uniformly on  $[a, b]$  as  $n \rightarrow \infty$ .*

*Proof.* Then

$$\|f - p_{n-1}\|_\infty \leq \frac{M(b-a)^n}{2^{2n-1} r^n} = 2M \left( \frac{b-a}{4r} \right)^n.$$

Since  $b-a < 4r$ , this tends to 0 as  $n \rightarrow \infty$ . □

Real Analysis does not offer a pleasant estimation of bounds for higher derivatives, but the following standard result from Complex Analysis provides exactly the type of bound required:

**LEMMA 5.** *Suppose that  $f$  is analytic on  $\{z : |z - z_0| < R\}$ . Suppose that  $r < R$  and  $|f(z)| \leq M$  for  $|z - z_0| = r$ . Then  $|f^{(n)}(z_0)| \leq Mn!/r^n$  for all  $n \geq 1$ . □*

With this, we can formulate a convergence theorem based on the nature of  $f$  as a

complex function. Given a real interval  $I = [a, b]$ , let  $E_r(I)$  be the set of points in the complex plane at distance no more than  $r$  from some point of  $I$ . This is comprised of the rectangle  $\{x + iy : a \leq x \leq b, |y| \leq r\}$  together with semicircles centred at  $a$  and  $b$ . We conclude:

**THEOREM 21.** *Suppose that  $f$  is analytic on a complex region containing  $E_r(I)$ , where  $I = [a, b]$  and  $r > \frac{1}{4}(b - a)$ . Let  $p_n$  be defined as in Theorem 20. Then  $p_n \rightarrow f$  uniformly on  $I$  as  $n \rightarrow \infty$ .  $\square$*

This is not the strongest possible theorem of this type, but it is what follows naturally from this approach.

For comparison, we now outline the corresponding results for *equally spaced* points  $x_j$  (we revert to  $n + 1$  points). We start with an estimation of  $|q(x)|$ .

**LEMMA 6.** *Let  $x_j = x_0 + jh$  ( $0 \leq j \leq n$ ), with  $x_0 = a$  and  $x_n = b$ , so that  $h = \frac{1}{n}(b - a)$ . Let  $q_n(x) = \prod_{j=0}^n (x - x_j)$ . Then*

$$\|q_n\|_\infty \leq \frac{n!}{n^{n+1}}(b - a)^{n+1}.$$

*Proof.* The substitution  $x = x_0 + th$  gives  $q_n(x) = h^{n+1}Q_n(t)$ , where  $Q_n(t) = t(t - 1) \dots (t - n)$  and  $0 \leq t \leq n$ . Let  $k \leq t \leq k + 1$  (where  $0 \leq k \leq n - 1$ ). Clearly,

$$\prod_{j=0}^k (t - j) \leq \prod_{j=0}^k (k + 1 - j) = (k + 1)!,$$

$$\prod_{j=k+1}^n (j - t) \leq \prod_{j=k+1}^n (j - k) = (n - k)!.$$

If  $k < n - 1$ , then  $(n - k)! = 2.3 \dots (n - k) \leq (k + 2)(k + 3) \dots n$ . Hence  $(k + 1)!(n - k)! \leq n!$  for each  $k$  (with equality when  $k$  is 0 or  $n - 1$ ). So  $|Q_n(t)| \leq n!$  for  $0 \leq t \leq n$ . Substituting  $h = (b - a)/n$ , we obtain the stated bound for  $\|q_n\|_\infty$ .  $\square$

Inserting Lemma 6 into Theorem 4, we obtain:

**THEOREM 22.** *Suppose that  $|f^{(n)}(x)| \leq M_n$  on  $[a, b]$ , and let  $p_n$  be the polynomial interpolating  $f$  at  $n + 1$  equally spaced points  $x_j$  in  $[a, b]$ . Then*

$$\|f - p_n\|_\infty \leq \frac{(b - a)^{n+1} M_{n+1}}{(n + 1)n^{n+1}}. \quad \square$$

We now apply Stirling's formula:  $n! \sim cn^{n+\frac{1}{2}}e^{-n}$  as  $n \rightarrow \infty$ , where  $c = (2\pi)^{1/2}$ .

COROLLARY. In Theorem 22, if for some  $M$ , we have  $M_n \leq Mn!/r^n$  for all  $n \geq 1$ , where  $r > \frac{1}{e}(b-a)$ , then  $p_n \rightarrow f$  uniformly on  $[a, b]$  as  $n \rightarrow \infty$ .

*Proof.* With this assumption, we have

$$\begin{aligned} \|f - p_n\|_\infty &\leq \frac{M(b-a)^{n+1}n!}{(nr)^{n+1}} \\ &\sim \frac{M(b-a)^{n+1}}{(nr)^{n+1}} cn^{n+\frac{1}{2}}e^{-n} \quad \text{by Stirling's formula} \\ &= \frac{Mc}{n^{1/2}} \left(\frac{b-a}{er}\right)^{n+1}. \quad \square \end{aligned}$$

Compare the Corollary of Theorem 20: we see that the factor  $\frac{1}{4}$  has been replaced by  $\frac{1}{e}$ . Clearly, Theorem 21 applies with this modification.

*Note.* Using the more accurate estimate  $(k+1)!(n-k)!$  instead of  $n!$  in Lemma 6, one can show that convergence will occur at the mid-point of the interval under the weaker condition  $r > \frac{1}{2e}(b-a)$ . We omit the details.

*An example of non-convergence*

It was shown by Runge in 1901 that for the function  $f(x) = 1/(1+x^2)$ , the sequence of interpolating polynomials for equally spaced points (on long enough intervals) fails to converge even pointwise to  $f(x)$ . We finish with a proof of this fact. It involves some fairly detailed estimations, but the following version offers at least a modest degree of simplification compared with some accounts.

The estimate in Theorem 4 will not suffice. Recall Proposition 8:  $f(x) - p(x) = f[x_0, x_1, \dots, x_n, x]q(x)$ .

LEMMA 7. Let  $f(x) = 1/(1+x^2)$  and let  $x_j$  ( $-n \leq j \leq n$ ) be distinct points with  $x_{-j} = -x_j$  for each  $j$ . Then

$$f[x_0, x_{-1}, x_1, \dots, x_{-n}, x_n, x] = (-1)^{n+1}xf(x) \prod_{j=1}^n \frac{1}{1+x_j^2}.$$

*Proof.* We show first that

$$f[x_{-1}, x_1, \dots, x_{-n}, x_n, x] = (-1)^n f(x) \prod_{j=1}^n \frac{1}{1+x_j^2}. \quad (10)$$

First we consider the case  $n = 1$ . Since  $x_{-1} = -x_1$  and  $f(x_{-1}) = f(x_1)$ , we have, by (2):

$$f[x_{-1}, x_1, x] = \frac{f(x)}{x^2 - x_1^2} + \frac{f(x_1)}{2x_1(x_1 - x)} + \frac{f(x_1)}{2x_1(x_1 + x)}$$

$$\begin{aligned}
&= \frac{f(x) - f(x_1)}{x^2 - x_1^2} \\
&= \frac{1}{x^2 - x_1^2} \left( \frac{1}{1 + x^2} - \frac{1}{1 + x_1^2} \right) \\
&= -\frac{1}{(1 + x^2)(1 + x_1^2)}.
\end{aligned}$$

Assume now that (10) holds for a certain  $n$ , and denote the LHS by  $g(x)$ . By Proposition 11 and the case  $n = 1$ , we then have

$$\begin{aligned}
f[x_{-1}, x_1, \dots, x_{-n}, x_n, x_{-(n+1)}, x_{n+1}, x] &= g[x_{-(n+1)}, x_{n+1}, x] \\
&= (-1)^n \prod_{j=1}^n \frac{1}{1 + x_j^2} f[x_{-(n+1)}, x_{n+1}, x] \\
&= (-1)^{n+1} \prod_{j=1}^{n+1} \frac{1}{1 + x_j^2} f(x).
\end{aligned}$$

This establishes (10), by induction. Also, since  $x_0 = 0$  and  $f(x_0) = 1$ , we have

$$f[x_0, x] = \frac{f(x) - 1}{x} = -\frac{x}{1 + x^2} = -xf(x).$$

By Proposition 10 again,

$$\begin{aligned}
f[x_0, x_{-1}, x_1, \dots, x_{-n}, x_n, x] &= g[x_0, x] \\
&= (-1)^n \prod_{j=1}^n \frac{1}{1 + x_j^2} f[x_0, x] \\
&= (-1)^{n+1} xf(x) \prod_{j=1}^n \frac{1}{1 + x_j^2}. \quad \square
\end{aligned}$$

**PROPOSITION 23.** *Let  $f(x) = 1/(1 + x^2)$ . Let  $p_n$  be the polynomial interpolating  $f$  at  $n + 1$  equally spaced points through  $[-a, a]$ . If  $a > 12\frac{1}{2}$ , then the sequence  $[p_n(\frac{1}{2}a)]$  does not converge.*

*Note.* The condition of Theorem 21 (with 4 replaced by  $e$ ) is not satisfied, because  $f$  has a pole at  $i$ , at distance 1 from the interval.

*Proof.* Let  $h = a/n$  and  $x_j = jh$  for  $-n \leq j \leq n$ . By Proposition 8 and Lemma 7,

$$|f(x) - p_{2n}(x)| = |x|f(x)|q_{2n}(x)| \prod_{j=1}^n \frac{1}{1 + j^2h^2},$$

where  $q_{2n}(x) = \prod_{j=-n}^n (x - jh)$ . Write  $A_n = |q_{2n}(\frac{1}{2}a)|$  and  $B_n = \prod_{j=1}^n (1 + j^2h^2)$ . Our statement will follow if we can show that  $A_n/B_n \rightarrow \infty$  when  $n$  tends to infinity through odd

values (note: when  $n$  is even,  $A_n = 0$ ). To do this, we will estimate  $\log A_n$  from below and  $\log B_n$  from above by comparison with the corresponding integrals.

Let  $n = 2r - 1$ , so that  $\frac{1}{2}a = (r - \frac{1}{2})h$ . Then

$$\begin{aligned}\log A_n &= \sum_{j=-2r+1}^{2r-1} \log |r - \frac{1}{2} - j| + (2n + 1) \log h \\ &= \sum_{k=1}^r \log(k - \frac{1}{2}) + \sum_{k=1}^{3r-1} \log(k - \frac{1}{2}) + (4r - 1) \log h.\end{aligned}$$

Now  $\log x$  is a concave function (the second derivative is negative) and for any concave function  $g$ , one has  $\int_{k-1}^k g \leq g(k - \frac{1}{2})$  (this is geometrically obvious, and easy to prove formally from the mean-value theorem). So

$$\sum_{k=1}^r \log(k - \frac{1}{2}) \geq \int_0^r \log x \, dx = r \log r - r,$$

Hence also

$$\begin{aligned}\sum_{k=1}^{3r-1} \log(k - \frac{1}{2}) &\geq \sum_{k=1}^{3r} \log(k - \frac{1}{2}) - \log 3r \\ &\geq 3r \log 3r - 3r - \log 3r \\ &\geq 3r \log r + 3(\log 3 - 1)r - \log r - \log 3.\end{aligned}$$

Further, since  $h = a/(2r - 1)$ ,

$$(4r - 1) \log h > (4r - 1)(\log a - \log 2r) = 4r \log a - 4r(\log r + \log 2) + \log 2r - \log a.$$

Together, these inequalities give

$$\begin{aligned}\log A_n &> 4r \log a + (3 \log 3 - 4 \log 2 - 4)r - c_1 \\ &= 4r(\log a - 1) + (3 \log 3 - 4 \log 2)r - c_1\end{aligned}\tag{11}$$

for a certain constant  $c_1$ .

We turn to the estimation of  $\log B_n$ . It equals  $\sum_{j=1}^n g(jh)$ , where  $g(x) = \log(1 + x^2)$ . Since  $g$  is an increasing function, standard integral comparison gives  $h \sum_{j=1}^{n-1} g(jh) \leq I$ , where  $I = \int_0^a g(x) \, dx$ . Integrating by parts, with 1 as one factor, we find

$$\begin{aligned}I &= a \log(1 + a^2) - 2 \int_0^a \left(1 - \frac{1}{1 + x^2}\right) dx \\ &< a \log(1 + a^2) - 2a + \pi.\end{aligned}$$

Hence

$$\log B_n \leq \frac{I}{h} + g(a) = \frac{nI}{a} + g(a) < (n+1)\log(1+a^2) - 2n + \frac{\pi n}{a}.$$

Now

$$\log(1+a^2) - 2\log a = \int_{a^2}^{1+a^2} \frac{1}{t} dt < \frac{1}{a^2},$$

hence (with  $n = 2r - 1$ ) we have

$$\begin{aligned} \log B_n &< 2n(\log a - 1) + \left(\frac{\pi}{a} + \frac{1}{a^2}\right)n + c_2 \\ &< 4r(\log a - 1) + 2\left(\frac{\pi}{a} + \frac{1}{a^2}\right)r + c_2 \end{aligned} \tag{12}$$

for a certain constant  $c_2$ . By (11) and (12), it is clear that we will have  $\log A_{2r-1} - \log B_{2r-1} \rightarrow \infty$  as  $r \rightarrow \infty$  provided that

$$2\left(\frac{\pi}{a} + \frac{1}{a^2}\right) < 3\log 3 - 4\log 2 \approx 0.5232,$$

which is true for  $a \geq 12\frac{1}{2}$ . □

Clearly, this is quite a delicate estimation. One can show that, by contrast, convergence does occur at 0.

*updated 12 May 2017*