Divided Differences

Carl de Boor

21 December 2004

"Est enim fere ex pulcherrimis quæ solvere desiderem."
(It is among the most beautiful I could desire to solve.) [Newton 1676]

Abstract. Starting with a novel definition of divided differences, this essay derives and discusses the basic properties of, and facts about, (univariate) divided differences.

MSC: 41A05, 65D05, 41-02, 41-03

1 Introduction and basic facts

While there are several ways to think of divided differences, including the one suggested by their very name, the most efficient way is as the coefficients in a Newton form. This form provides an efficient representation of Hermite interpolants.

Let $\Pi \subset (\mathbb{F} \to \mathbb{F})$ be the linear space of polynomials in one real ($\mathbb{F} = \mathbb{R}$) or complex ($\mathbb{F} = \mathbb{C}$) variable, and let $\Pi_{<n}$ denote the subspace of all polynomials

Surveys in Approximation Theory
Volume 1, 2005. pp. 46–69.

ISSN 1555-578X

of degree $< n$. The **Newton form** of $p \in \Pi$ with respect to the sequence $t = (t_1, t_2, \ldots)$ of **centers** t_j is its expansion

$$p =: \sum_{j=1}^{\infty} w_{j-1,t}\, c(j) \tag{1}$$

in terms of the **Newton polynomial**s

$$w_i \ := \ w_{i,t} \ := \ (\cdot - t_1) \cdots (\cdot - t_i), \quad i = 0, 1, \ldots \ . \tag{2}$$

Each $p \in \Pi$ does, indeed, have exactly one such expansion for any given t since $\deg w_{j,t} = j$, all j, hence $(w_{j-1,t} : j \in \mathbb{N})$ is a **graded basis for** Π in the sense that, for each n, $(w_{j-1,t} : j = 1{:}n)$ is a basis for $\Pi_{<n}$.

In other words, the column map

$$W_t : \mathbb{F}_0^{\mathbb{N}} \to \Pi : c \mapsto \sum_{j=1}^{\infty} w_{j-1,t} c(j) \tag{3}$$

(from the space $\mathbb{F}_0^{\mathbb{N}}$ of scalar sequences with finitely many nonzero entries to the space Π) is 1-1 and onto, hence invertible. In particular, for each $n \in \mathbb{N}$, the coefficient $c(n)$ in the Newton form (1) for p depends linearly on p, i.e., $p \mapsto c(n) = (W_t^{-1} p)(n)$ *is a well-defined linear functional on* Π*, and vanishes on* $\Pi_{<n-1}$. More than that, since all the (finitely many nontrivial) terms in (1) with $j > n$ have $w_{n,t}$ as a factor, we can write

$$p \ = \ p_n + w_{n,t} q_n, \tag{4}$$

with q_n a polynomial we will look at later (in Example 6), and with

$$p_n \ := \ \sum_{j=1}^{n} w_{j-1,t} c(j)$$

a polynomial of degree $< n$. This makes p_n necessarily the remainder left by the division of p by $w_{n,t}$, hence well-defined for every n, hence, by induction, we obtain another proof that the Newton form (1) itself is well-defined.

In particular, p_n depends only on p and on

$$t_{1:n} := (t_1, \ldots, t_n),$$

therefore the same is true of its leading coefficient, $c(n)$. This is reflected in the (implicit) definition

$$p \ =: \ \sum_{j=1}^{\infty} w_{j-1,t}\, \Delta(t_{1:j}) p, \quad p \in \Pi, \tag{5}$$

in which the coefficient $c(j)$ in the Newton form (1) for p is denoted

$$\Delta(t_{1:j})p \;=\; \Delta(t_1,\ldots,t_j)p \;:=\; ((W_t)^{-1}p)(j) \tag{6}$$

and called the **divided difference of** p **at** $t_1,\ldots,t_j$. It is also called a divided difference of **order** $j-1$, and the reason for all this terminology will be made clear in a moment.

Since W_t is a continuous function of t, so is W_t^{-1}, hence so is $\Delta(t_{1:j})$ (see Proposition 21 for proof details). Further, since $w_{j,t}$ is symmetric in $t_1,\ldots,t_j$, so is $\Delta(t_{1:j})$. Also, $\Delta(t_{1:j}) \perp \Pi_{<j}$ (as mentioned before).

In more practical terms, we have

Proposition 7. *The sum*

$$p_n \;=\; \sum_{j=1}^{n} w_{j-1,t}\,\Delta(t_{1:j})p$$

of the first n terms in the Newton form (1) for p is the **Hermite interpolant to** p **at** $t_{1:n}$, *i.e., the unique polynomial r of degree $< n$ that* **agrees with** p **at** $t_{1:n}$ *in the sense that*

$$D^i r(z) = D^i p(z), \quad 0 \le i < \mu_z := \#\{j \in [1\,..\,n] : t_j = z\}, \quad z \in \mathbb{F}. \tag{8}$$

Proof: One readily verifies by induction on the nonnegative integer μ that, for any $z \in \mathbb{F}$, any polynomial f **vanishes** μ**-fold at** z iff f has $(\cdot - z)^\mu$ as a factor, i.e.,

$$D^i f(z) = 0 \ \text{ for } \ i = 0,\ldots,\mu-1 \quad\Longleftrightarrow\quad f \in (\cdot - z)^\mu \Pi. \tag{9}$$

Since $p - p_n = w_{n,t}q_n$, this implies that $r = p_n$ does, indeed satisfy (8).

Also, p_n is the only such polynomial since, by (9), for any polynomial r satisfying (8), the difference $p_n - r$ must have w_n as a factor and, if r is of degree $< n$, then this is possible only when $r = p_n$. □

Example 1. For $n = 1$, we get that

$$\Delta(t_1) : p \mapsto p(t_1),$$

i.e., $\Delta(\tau)$ can serve as a (nonstandard) notation for the linear functional of evaluation at τ. □

Example 2. For $n = 2$, p_n is the polynomial of degree < 2 that matches p at $t_{1:2}$. If $t_1 \neq t_2$, then we know p_2 to be writable in 'point-slope form' as

$$p_2 = p(t_1) + (\cdot - t_1)\frac{p(t_2) - p(t_1)}{t_2 - t_1},$$

while if $t_1 = t_2$, then we know p_2 to be

$$p_2 = p(t_1) + (\cdot - t_1)Dp(t_1).$$

Hence, altogether,

$$\Delta(t_{1:2})p \;=\; \begin{cases} \dfrac{p(t_2) - p(t_1)}{t_2 - t_1}, & t_1 \neq t_2; \\ Dp(t_1), & \text{otherwise}. \end{cases} \tag{10}$$

Thus, for $t_1 \neq t_2$, $\Delta(t_{1:2})$ is a quotient of differences, i.e., a *divided difference.* □

Example 3. Directly from the definition of the divided difference,

$$\Delta(t_{1:j})w_{i-1,t} = \delta_{ji}, \tag{11}$$

therefore (remembering that $\Delta(t_{1:j}) \perp \Pi_{<j-1}$)

$$\Delta(t_{1:j})()^{j-1} \;=\; 1, \tag{12}$$

with

$$()^k : \mathbb{F} \to \mathbb{F} : z \mapsto z^k$$

a handy if nonstandard notation for the power functions. □

Example 4. If t is a constant sequence, $t = (\tau, \tau, \ldots)$ say, then

$$w_{j,(\tau,\tau,\ldots)} = (\cdot - \tau)^j,$$

hence the **Taylor expansion**

$$p = \sum_{n=0}^{\infty} (\cdot - \tau)^n D^n p(\tau)/n! \tag{13}$$

is the Newton form for the polynomial p with respect to the sequence $(\tau, \tau, \ldots)$. Therefore,

$$\Delta(\tau^{[n+1]})p := \Delta(\underbrace{\tau, \ldots, \tau}_{n+1\text{ terms}})p = D^n p(\tau)/n!, \quad n = 0, 1, \ldots. \tag{14}$$

□

Example 5. If $\ell : t \mapsto at + b$, then $(\ell(z) - \ell(t_i)) = a(z - t_i)$, hence

$$a^{n-1}\,\Delta(\ell(t_{1:n}))p = \Delta(t_{1:n})(p \circ \ell). \tag{15}$$

□

Example 6. Consider the polynomial q_n introduced in (4):

$$p = p_n + w_{n,t}q_n.$$

Since $p(t_{n+1}) = p_{n+1}(t_{n+1})$ and $p_{n+1} = p_n + w_{n,t}\,\Delta(t_{1:n+1})p$, we have

$$w_{n,t}(t_{n+1})q_n(t_{n+1}) = w_{n,t}(t_{n+1})\,\Delta(t_{1:n+1})p,$$

therefore

$$q_n(t_{n+1}) = \Delta(t_{1:n+1})p,$$

at least for any t_{n+1} for which $w_{n,t}(t_{n+1}) \neq 0$, hence for every $t_{n+1} \in \mathbb{F}$, by the continuity of q_n, and the continuity of $\Delta(t_{1:n}, \cdot)p$, i.e., of $\Delta(t_{1:n+1})p$ as a function of t_{n+1}. It follows that

$$q_n = \Delta(t_{1:n}, \cdot)p$$

and

$$p = p_n + w_{n,t}\,\Delta(t_{1:n}, \cdot)p, \tag{16}$$

the **standard error formula for Hermite interpolation**. More than that, by the very definition, (4), of q_n, we now know that

$$\Delta(t_{1:n}, \cdot)p \;=\; q_n = (p - p_n)/w_{n,t} = \sum_{j>n} \frac{w_{j-1,t}}{w_{n,t}}\,\Delta(t_{1:j})p, \tag{17}$$

and we recognize the sum here as a Newton form with respect to the sequence $(t_j : j > n)$. This provides us with the following **basic divided difference identity**:

$$\Delta(t_{n+1:j})\,\Delta(t_{1:n}, \cdot) = \Delta(t_{1:j}), \quad j > n. \tag{18}$$

□

For the special case $n = j - 2$, the basic divided difference identity, (18), reads

$$\Delta(t_{j-1:j})\,\Delta(t_{1:j-2}, \cdot) = \Delta(t_{1:j}),$$

or, perhaps more suggestively,

$$\Delta(t_{j-1}, \cdot)\,\Delta(t_{1:j-2}, \cdot) = \Delta(t_{1:j-1}, \cdot),$$

hence, by induction,

$$\Delta(t_{j-1}, \cdot)\,\Delta(t_{j-2}, \cdot)\,\cdots\,\Delta(t_1, \cdot) \;=\; \Delta(t_{1:j-1}, \cdot). \tag{19}$$

In other words, $\Delta(t_{1:j})$ is obtainable by forming difference quotients $j-1$ times. This explains our calling $\Delta(t_{1:j})$ a ‘divided difference of order $j - 1$’.

2 Continuity and smoothness

The column map

$$W_t : \mathbb{F}_0^{\mathbb{N}} \to \Pi : c \mapsto \sum_{j=1}^{\infty} w_{j-1,t}c(j)$$

introduced in (3) *is continuous as a function of* t*, hence so is its inverse*, as follows directly from the identity

$$A^{-1} - B^{-1} = A^{-1}(B - A)B^{-1}, \tag{20}$$

valid for any two invertible maps A, B (with the same domain and target). Therefore, also each $\Delta(t_{1:j})$ is a continuous function of t, all of this in the pointwise sense. Here is the formal statement and its proof.

Proposition 21. *For any* $p \in \Pi$,

$$\lim_{s\to t}(\Delta(s_{1:j})p : j \in \mathbb{N}) = (\Delta(t_{1:j})p : j \in \mathbb{N}).$$

Proof: Let $p \in \Pi_{<n}$. Then $t \mapsto \Delta(t_{1:k})p = 0$ for $k > n$, hence trivially continuous. As for $k \le n$, let

$$W_{t,n} := \mathbb{F}^n \to \Pi_{<n} : c \mapsto \sum_{j=1}^{n} w_{j-1,t}c(j)$$

be the restriction of W_t to $\mathbb{F}^n$, as a linear map to $\Pi_{<n}$. Then, in whatever norms we might choose on $\mathbb{F}^n$ and $\Pi_{<n}$, $W_{t,n}$ is bounded and invertible, hence boundedly invertible uniformly in $t_{1:n}$ as long as $t_{1:n}$ lies in some bounded set. Therefore, with (20), since $\lim_{s\to t} W_{s,n} = W_{t,n}$, also

$$\lim_{s\to t}(\Delta(s_{1:j})p : j = 1{:}n) = (W_{t,n})^{-1}p = (\Delta(t_{1:j})p : j = 1{:}n).$$

$\square$

This continuity is very useful. For example, it implies that it is usually sufficient to check a proposed divided difference identity by checking it only for pairwise distinct arguments.

As another example, we used the continuity earlier (in Example 6) to prove that $\Delta(t_{1:n}, \cdot)p$ is a polynomial. This implies that $\Delta(t_{1:n}, \cdot)p$ is differentiable, and, with that, (18) and (14) even provide the following formula for the derivatives.

Proposition 22.

$$D^k \Delta(t_{1:j}, \cdot)p = k!\,\Delta(t_{1:j}, [\cdot]^{k+1})p, \quad k \in \mathbb{N}.$$

3 Refinement

Already Cauchy [Cauchy 1840] had occasion to use the simplest nontrivial case of the following fact.

Proposition 23. *For any n-sequence t and any $1 \le \sigma(1) < \cdots < \sigma(k) \le n$,*

$$\Delta(t_{\sigma(1:k)}) = \sum_{j=\sigma(1)-1}^{\sigma(k)-k} \alpha(j)\, \Delta(t_{j+1:j+k}),$$

with $\alpha = \alpha_{t,\sigma}$ positive in case t is strictly increasing.

Proof: Since $\Delta(t_{1:n})$ is symmetric in the t_j, (18) implies

$$(t_n - t_1)(\Delta(t_{1:n\backslash m}) - \Delta(t_{2:n})) \;=\; (t_1 - t_m)(\Delta(t_{2:n}) - \Delta(t_{1:n-1})),$$

with

$$t_{1:n\backslash m} := t_{1:m-1,m+1:n} := (t_1, \ldots, t_{m-1}, t_{m+1}, \ldots, t_n).$$

On rearranging the terms, we get

$$(t_n - t_1)\,\Delta(t_{1:n\backslash m}) \;=\; (t_n - t_m)\,\Delta(t_{2:n}) + (t_m - t_1)\,\Delta(t_{1:n-1}),$$

and this proves the assertion for the special case $k = n - 1$, and even gives an explicit formula for α in this case.

From this, the general case follows by induction on $n-k$, with α computable as a convolution of sequences which, by induction, are positive in case t is strictly increasing (since this is then trivially so for $k = n - 1$), hence then α itself is positive. □

My earliest reference for the general case is [Popoviciu 1933].

4 Divided difference of a product: Leibniz, Opitz

The map

$$P := P_{n,t} : \Pi \to \Pi : p \mapsto p_n,$$

of Hermite interpolation at $t_{1:n}$, is the linear projector P on Π with

$$\operatorname{ran} P = \Pi_{<n}, \quad \operatorname{ran}(\operatorname{id} - P) = \operatorname{null} P = w_{t,n}\Pi.$$

In particular, the nullspace of P is an *ideal* if, as we may, we think of Π as a ring, namely the ring with multiplication defined pointwise,

$$(pg)(z) := p(z)g(z), \quad z \in \mathbb{F}.$$

In other words, the nullspace of P is a linear subspace closed also under pointwise multiplication. This latter fact is (see [de Boor 2003b]) equivalent to the identity

$$P(pq) = P(p(Pq)), \quad p, q \in \Pi. \tag{24}$$

For $p \in \Pi$, consider the map

$$M_p : \Pi_{<n} \to \Pi_{<n} : f \mapsto P(pf).$$

Then M_p is evidently linear and, also evidently, so is the resulting map

$$M : \Pi \to L(\Pi_{<n}) : p \mapsto M_p$$

on Π to the space of linear maps on $\Pi_{<n}$. More than that, since, by (24),

$$M_{pq} f = P(pqf) = P(pP(qf)) = M_p M_q f, \quad p \in \Pi_{<n},\ p, q \in \Pi,$$

M is a ring homomorphism, from the ring Π into the ring $L(\Pi_{<n})$ in which composition serves as multiplication. The latter ring is well known not to be commutative while, evidently, $\operatorname{ran} M$ is a commutative subring.

It follows, in particular, that

$$M_p = p(M_{()^1}), \quad p \in \Pi,$$

hence

$$\widehat{M_p} = p(\widehat{M_{()^1}}), \quad p \in \Pi,$$

for the *matrix representation*

$$\widehat{M_p} := V M_p V^{-1}$$

of M_p with respect to any particular basis V of $\Pi_{<n}$. Look, in particular, at the matrix representation with respect to the Newton basis

$$V := [w_{j-1,t} : j = 1{:}n]$$

for $\Pi_{<n}$. Since

$$()^1 w_{j-1,t} = t_j w_{j-1,t} + w_{j,t}, \quad j = 1, 2, \ldots,$$

therefore evidently

$$M_{()^1} w_{j-1,t} = P(()^1 w_{j-1,t}) = t_j w_{j-1,t} + (1 - \delta_{j,n}) w_{j,t}, \quad j = 1{:}n.$$

Consequently, the matrix representation for $M_{()^1}$ with respect to the Newton basis V is the bidiagonal matrix

$$\widehat{M_{()^1}} \;=\; A_{n,t} \;:=\; \begin{bmatrix} t_1 & & & & \\ 1 & t_2 & & & \\ & 1 & t_3 & & \\ & & \ddots & \ddots & \\ & & & 1 & t_n \end{bmatrix}.$$

On the other hand, for any $p \in \Pi$ and $j = 1{:}n$,

$$\left(\sum_{i=j}^{n} (w_{i-1,t}/w_{j-1,t})\, \Delta(t_{j:i})p \right) w_{j-1,t}$$

is a polynomial of degree $< n$ and, for pairwise distinct t_i, it agrees with $pw_{j-1,t}$ at $t_{1:n}$ since the sum describes the polynomial of degree $\le n-j$ that matches p at $t_{j:n}$ while both functions vanish at $t_{1:j-1}$. Consequently, with the convenient agreement that

$$\Delta(t_{j:i}) := 0, \qquad j > i,$$

we conclude that

$$P(pw_{j-1,t}) = \sum_{i=1}^{n} w_{i-1,t}\, \Delta(t_{j:i})p, \quad j = 1{:}n,$$

at least when the t_i are pairwise distinct. In other words, the jth column of the matrix $\widehat{M_p} = V^{-1} M_p V$ (which represents M_p with respect to the Newton basis V for $\Pi_{<n}$) has the entries

$$(\Delta(t_{j:i})p : i = 1{:}n) = (0, \ldots, 0, p(t_j), \Delta(t_j, t_{j+1})p, \ldots, \Delta(t_{j:n})p).$$

By the continuity of the divided difference (see Proposition 21), this implies

Proposition 25: Opitz formula. *For any* $p \in \Pi$,

$$p(A_{n,t}) \;=\; (\Delta(t_{j:i})p : i, j = 1{:}n). \tag{26}$$

The remarkable identity (26) is due to G. Opitz; see [Opitz 1964] which records a talk announced but not delivered. Opitz calls the matrices $p(A_{n,t})$ *Steigungsmatrizen* ('difference-quotient matrices'). Surprisingly, Opitz explicitly excludes the possibility that some of the t_j might coincide. [Bulirsch et al. 1968] ascribe (26) to Sylvester, but I have been unable to locate anything like this formula in Sylvester's collected works.

Example 7. For the monomial $()^k$, Opitz' formula gives

$$\Delta(t_{1:n})()^k = (A_{n,t})^k(n,1) = \sum_{\nu\in\{1:n\}^k} A_{n,t}(n,\nu_k)A_{n,t}(\nu_k,\nu_{k-1})\cdots A_{n,t}(\nu_1,1),$$

and, since $A_{n,t}$ is bidiagonal, the νth summand is zero unless the sequence $(1,\nu_1,\ldots,\nu_k,n)$ is increasing, with any strict increase no bigger than 1, in which case the summand equals t^α, with $\alpha_j - 1$ the multiplicity with which j appears in the sequence ν, $j = 1{:}n$. This confirms that $\Delta(t_{1:n})()^k = 0$ for $k < n-1$ and proves that

$$\Delta(t_{1:n})()^k = \sum_{|\alpha|=k-n-1} t^\alpha, \quad k \ge n-1. \tag{27}$$

□

My first reference for (27) is [Steffensen 1927: p.19f]. To be sure, once (27) is known, it is easily verified by induction, using the Leibniz formula, to be derived next.

Since, for any square matrix A and any polynomials p and q,

$$(pq)(A) = p(A)q(A),$$

it follows, in particular, that

$$\Delta(t_{1:n})(pq) = \widehat{M_{pq}}(n,1) = \widehat{M}_p(n,:)\widehat{M}_q(:,1),$$

hence

Corollary 28: Leibniz formula. *For any* $p,q\in\Pi$,

$$\Delta(t_{1:n})(pq) = \sum_{j=1:n} \Delta(t_{j:n})p\ \Delta(t_{1:j})q. \tag{29}$$

On the other hand, the Leibniz formula implies that, for any $p,q\in\Pi$,

$$(\Delta(t_{j:i})p : i,j=1{:}n)(\Delta(t_{j:i})q : i,j=1{:}n) \;=\; (\Delta(t_{j:i})(pq) : i,j=1{:}n),$$

hence, that, for any $p\in\Pi$,

$$p((\Delta(t_{j:i})()^1 : i,j=1{:}n)) \;=\; (\Delta(t_{j:i})p : i,j=1{:}n).$$

In other words, we can also view Opitz' formula as a corollary to Leibniz' formula.

My first reference for the Leibniz formula is [Popoviciu 1933], though Steffensen later devotes an entire paper, [Steffensen 1939], to it and this has become the standard reference for it despite the fact that Popoviciu, in response, wrote his own overview of divided differences, [Popoviciu 1940], trying, in vain, to correct the record.

The (obvious) name 'Leibniz formula' for it appears first in [de Boor 1972].

Induction on m proves the following

Corollary 30: General Leibniz formula. *For* $f:\mathbb{F}^m\to\mathbb{F}$,

$$\Delta(t_1,\ldots,t_k)f(\cdot,\ldots,\cdot) = \sum_{1=i(0)\le\cdots\le i(m)=k} \left(\otimes_{j=1}^m \Delta(t_{i(j-1)},\ldots,t_{i(j)})\right) f.$$

5 Construction of Newton form via a divided difference table

Divided difference table. *Assume that the sequence $(t_1, \ldots, t_n)$ has all its multiplicities (if any)* **clustered**, *meaning that, for any $i < j$, $t_i = t_j$ implies that $t_i = t_{i+1} = \cdots = t_j$. Then, by (18) and (14),*

$$\Delta(t_{i:j})p \ = \ \begin{cases} \dfrac{\Delta(t_{i+1:j})p - \Delta(t_{i:j-1})p}{t_j - t_i}, & t_i \neq t_j; \\ D^{j-i}p(t_i)/(j-i)! & \text{otherwise,} \end{cases} \qquad 1 \le i \le j \le n.$$

Hence, it is possible to fill in all the entries in the **divided difference table**

$$\begin{array}{cccccc}
\Delta(t_1)p & & & & & \\
 & \Delta(t_{1:2})p & & & & \\
\Delta(t_2)p & & \Delta(t_{1:3})p & & & \\
 & \Delta(t_{2:3})p & & \cdot & & \\
\Delta(t_3)p & & \cdot & & \Delta(t_{1:n-1})p & \\
 & \cdot & & \cdot & & \Delta(t_{1:n})p \\
\cdot & & \Delta(t_{n-3:n-1})p & & \Delta(t_{2:n})p & \\
 & \Delta(t_{n-2:n-1})p & & \cdot & & \\
\Delta(t_{n-1})p & & \Delta(t_{n-2:n})p & & & \\
 & \Delta(t_{n-1:n})p & & & & \\
\Delta(t_n)p & & & & &
\end{array}$$

column by column from left to right, using one of the n pieces of information

$$y(j) := D^{\mu_j}p(t_j), \quad \mu_j := \#\{i < j : t_i = t_j\}; \quad j = 1, \ldots, n, \tag{31}$$

in the leftmost column or else whenever we would otherwise be confronted with $0/0$.

After construction of this divided difference table, the top diagonal of the table provides the coefficients $(\Delta(t_{1:j})p : j = 1, \ldots, n)$ for the Newton form (with respect to centers $t_1, \ldots, t_{n-1}$) of the polynomial of degree $< n$ that matches p at $t_{1:n}$, i.e., the polynomial p_n. More than that, for any sequence $(i_1, \ldots, i_n)$ in which, for each j, $\{i_1, \ldots, i_j\}$ consists of consecutive integers in $[1\,..\,n]$, the above divided difference table provides the coefficients in the Newton form for the above r, but with respect to the centers $(t_{i_j} : j = 1{:}n)$.

Now note that the only information about p entering this calculation is the scalar sequence y described in (31). Hence we now know the following.

Proposition 32. *Let $(t_1, \ldots, t_n)$ have all its multiplicities (if any) clustered, and let $y \in \mathbb{F}^n$ be arbitrary. For $j = 1, \ldots, n$, let $c(j)$ be the first entry in the jth column in the above divided difference table as constructed in the described manner from y.*

Then

$$r := \sum_{j=1}^{n} w_{j-1,t} c(j)$$

is the unique polynomial of degree $< n$ *that satisfies the* **Hermite interpolation conditions**

$$D^{\mu_j} r(t_j) = y(j), \quad \mu_j := \#\{i < j : t_i = t_j\}; \quad j = 1, \ldots, n. \tag{33}$$

6 Evaluation of a Newton form via Horner's method

Horner's method. *Let* $c(j) := \Delta(t_{1:j}) r$ *for* $j = 1, \ldots, n > \deg r$, $z \in \mathbb{F}$, *and*

$$\begin{aligned} \widehat{c}(n) &:= c(n), \\ \widehat{c}(j) &:= c(j) + (z - t_j)\widehat{c}(j+1), \quad j = n{-}1, n{-}2, \ldots, 1. \end{aligned}$$

Then $\widehat{c}(1) = r(z)$. *More than that,*

$$r = \sum_{j=1}^{n} w_{j-1,\widehat{t}}\, \widehat{c}(j),$$

with

$$\widehat{t} \;:=\; (z, t_1, t_2, \ldots).$$

Proof: The first claim follows from the second, or else directly from the fact that Horner's method is nothing but the evaluation, from the inside out, of the nested expression

$$r(z) = c(1) + (z - t_1)(c(2) + \cdots + (z - t_{n-2})(c(n-1) + (z - t_{n-1})c(n)) \cdots),$$

for which reason Horner's method is also known as **Nested Multiplication**.

As to the second claim, note that $\Delta(z, t_{1:n-1}) r = \Delta(t_{1:n}) r$ since $\deg r < n$, hence $\widehat{c}(n) = \Delta(z, t_{1:n-1}) r$, while, directly from (18),

$$\Delta(\cdot, t_{1:j-1}) = \Delta(t_{1:j}) + (\cdot - t_j)\,\Delta(\cdot, t_{1:j}), \quad j \in \mathbb{N}, \tag{34}$$

hence, by (downward) induction,

$$\widehat{c}(j) = \Delta(z, t_{1:j-1}) r, \quad j = n{-}1, n{-}2, \ldots, 1.$$

$\square$

In effect, Horner's Method is another way of filling in a divided difference table, starting not at the left-most column but with a diagonal, and generating new entries, not from left to right, but from right to left:

$$
\begin{array}{cccccc}
\Delta(z)r & & & & & \\
 & \Delta(z,t_1)r & & & & \\
\Delta(t_1)r & & \Delta(z,t_{1:2})r & & & \\
 & \Delta(t_{1:2})r & & \cdot & & \\
\cdot & & \Delta(t_{1:3})r & & \Delta(z,t_{1:n-2})r & \\
 & \cdot & & \cdot & & \Delta(z,t_{1:n-1})r = \Delta(t_{1:n})r \\
\cdot & & \cdot & & \Delta(t_{1:n-1})r & \\
 & \cdot & & \cdot & & \Delta(t_{1:n})r \\
\cdot & & \cdot & & \cdot &
\end{array}
$$

Hence, Horner's method is useful for carrying out a *change of basis*, going from one Newton form to another. Specifically, $n-1$-fold iteration of this process, with $z = z_{n-1}, \ldots, z_1$, is an efficient way of computing the coefficients $(\Delta(z_{1:j})r : j = 1, \ldots, n)$, of the Newton form for $r \in \Pi_{<n}$ with respect to the centers $z_{1:n-1}$, from those for the Newton form with respect to centers $t_{1:n-1}$. Not all the steps need actually be carried out in case all the z_j are the same, i.e., when switching to the Taylor form (or local power form).

7 Divided differences of functions other than polynomials

Proposition 35. *On Π, the divided differences $\Delta(t_{1:j})$, $j = 1, \ldots, n$, provide a basis for the linear space of linear functionals spanned by*

$$\Delta(t_j)D^{\mu_j}, \quad \mu_j := \#\{i < j : t_i = t_j\}; \quad j = 1, \ldots, n. \tag{36}$$

Proof: By Proposition 7 and its proof,

$$\cap_{j=1}^n \ker \Delta(t_{1:j}) = w_{n,t}\Pi = \cap_{j=1}^n \ker \Delta(t_j)D^{\mu_j}.$$

Another proof is provided by Horner's method, which, in effect, expresses $(\Delta(t_{1:j}) : j = 1{:}n)$ as linear functions of $(\Delta(t_j)D^{\mu_j} : j = 1{:}n)$, thus showing the first sequence to be contained in the span of the second. Since the first is linearly independent (as it has $(w_{j-1,t} : j = 1{:}n)$ as a dual sequence) while both contain the same number of terms, it follows that both are bases of the same linear space. $\square$

This proposition provides a ready extension of $\Delta(t_{1:n})$ to functions more general than polynomials, namely to any function for which the derivatives mentioned in (36) make sense. It is exactly those functions for which the Hermite conditions (33) make sense, hence for which the Hermite interpolant r of (32) is defined. This leads us to G. Kowalewski's definition.

Definition 37 ([G. Kowalewski 1932]). *For any smooth enough function f defined, at least, at $t_1, \ldots, t_n$, $\Delta(t_{1:n})f$ is the leading coefficient, i.e., the coefficient of $()^{n-1}$, in the power form for the Hermite interpolant to f at $t_{1:n}$.*

In consequence, $\Delta(t_{1:n})f = \Delta(t_{1:n})p$ *for any polynomial p that matches f at $t_{1:n}$.*

Example 8. Assume that none of the t_j is zero. Then,

$$\Delta(t_{1:n})()^{-1} = (-1)^{n-1}/(t_1 \cdots t_n). \tag{38}$$

This certainly holds for $n = 1$ while, for $n > 1$, by (29), $0 = \Delta(t_{1:n})(()^{-1}()^1) = \Delta(t_{1:n})()^{-1}t_n + \Delta(t_{1:n-1})()^{-1}$, hence $\Delta(t_{1:n})()^{-1} = -\Delta(t_{1:n-1})()^{-1}/t_n$, and induction finishes the proof. This implies the handy formula

$$\Delta(t_{1:n})(z - \cdot)^{-1} = 1/w_{n,t}(z), \quad z \neq t_1, \ldots, t_n. \tag{39}$$

Therefore, with $\#\xi := \#\{j : \xi = t_j, 1 \leq j \leq n\}$ the multiplicity with which ξ occurs in the sequence $t_{1:n}$, and

$$1/w_{n,t}(z) =: \sum_{\xi \in t} \sum_{0 \leq \mu < \#\xi} \frac{\mu! A_{\xi\mu}}{(z-\xi)^{\mu+1}}$$

the partial fraction expansion of $1/w_{n,t}$, we obtain Chakalov's expansion

$$\Delta(t_0, \ldots, t_k)f = \sum_{\xi \in t} \sum_{0 \leq \mu < \#\xi} A_{\xi\mu} D^\mu f(\xi) \tag{40}$$

(from [Chakalov 1938]) for $f := 1/(z - \cdot)$ for arbitrary z since $D^\mu 1/(z - \cdot) = \mu!/(z - \cdot)^{\mu+1}$, hence also for any smooth enough f, by the density of $\{1/(z - \cdot) : z \in \mathbb{F}\}$.

This is an illustration of the peculiar effectiveness of the formula (39), for the divided difference of $1/(z - \cdot)$, for deriving and verifying divided difference identities. □

Example 9. When the t_j are pairwise distinct, (40) must reduce to

$$\Delta(t_{1:n})f = \sum_{j=1}^{n} f(t_j)/Dw_{n,t}(t_j), \tag{41}$$

since this is readily seen to be the leading coefficient of the polynomial of degree $< n$ that matches a given f at the n pairwise distinct sites $t_1, \ldots, t_n$ when we write that polynomial in Lagrange form,

$$\sum_{j=1}^{n} f(t_j) \prod_{i \in 1:n \backslash j} \frac{\cdot - t_i}{t_j - t_i}.$$

It follows (see the proof of [Erdős et al. 1940: Lemma I]) that, *for* $-1 \le t_1 < \cdots < t_n \le 1$,

$$\| \Delta(t_{1:n}) : C([-1\,..\,1]) \to \mathbb{F} \| = \sum_{j=1}^{n} 1/|Dw_{n,t}(t_j)| \ge 2^{n-2}, \tag{42}$$

with equality iff $w_{n,t} = (()^2 - 1)U_{n-2}$, where U_{n-2} is the second-kind Chebyshev polynomial.

Indeed, for any such

$$\tau := (t_1, \ldots, t_n),$$

the restriction λ of $\Delta(\tau)$ to $\Pi_{<n}$ is the unique linear functional on $\Pi_{<n}$ that vanishes on $\Pi_{<n-1}$ and takes the value 1 at $()^{n-1}$, hence takes its norm on the error of the best (uniform) approximation to $()^{n-1}$ from $\Pi_{<n-1}$, i.e., on the Chebyshev polynomial of degree $n-1$. Each such $\Delta(\tau)$ is an extension of this λ, hence has norm $\ge \|\lambda\| = 1/\operatorname{dist}(()^{n-1}, \Pi_{<n-1}) = 2^{n-2}$, with equality iff $\Delta(\tau)$ takes on its norm on that Chebyshev polynomial, i.e., iff τ is the sequence of extreme sites of that Chebyshev polynomial. □

8 The divided difference as approximate normalized derivative

Assume that f is differentiable on an interval that contains the nondecreasing finite sequence

$$\tau = (\tau_0 \le \cdots \le \tau_k),$$

and assume further that $\Delta(\tau)f$ is defined, hence so is the Hermite interpolant

$$P_\tau f$$

of f at τ.

Then $f - P_\tau f$ vanishes at $\tau_{0:k}$, therefore $D(f - P_\tau f)$ vanishes at some $\sigma = (\sigma_0, \ldots, \sigma_{k-1})$ that **interlaces** τ, meaning that

$$\tau_i \le \sigma_i \le \tau_{i+1}, \quad \text{all } i.$$

This is evident when $\tau_i = \tau_j$ for some $i < j$, and is Rolle's Theorem when $\tau_i < \tau_{i+1}$. Consequently, $DP_\tau f$ is a polynomial of degree $< k$ that matches Df at $\sigma_0, \ldots, \sigma_{k-1}$, hence must be its Hermite interpolant at σ. This proves the following.

Proposition 43 ([Hopf 1926]). *If f is differentiable on an interval that contains the nondecreasing $(k+1)$-sequence τ and smooth enough at τ so that its Hermite interpolant, $P_\tau f$, at τ exists, then there is a nondecreasing k-sequence σ interlacing τ and so that*

$$P_\sigma(Df) = DP_\tau f.$$

In particular, then

$$k\,\boldsymbol{\Delta}(\tau)f = \boldsymbol{\Delta}(\sigma)Df.$$

From this, induction provides the

Corollary ([Schwarz 1881-2]). *Under the same assumptions, but with f k times differentiable on that interval, there exists ξ in that interval for which*

$$k!\,\boldsymbol{\Delta}(\tau_0,\ldots,\tau_k) = D^k f(\xi). \tag{44}$$

The special case $k=1$, i.e.,

$$\boldsymbol{\Delta}(a,b)f = Df(\xi), \quad \text{for some } \xi \in (a\,..\,b),$$

is so obvious a consequence or restatement of L'Hôpital's Rule, it must have been around at least that long.

Chakalov [Tchakaloff 1934] has made a detailed study of the possible values that ξ might take in (44) as f varies over a given class of functions.

[A. Kowalewski 1917: p. 91] reports that already Taylor, in [Taylor 1715], derived his eponymous expansion (13) as the limit of Newton's formula, albeit for equally spaced sites only.

9 Representations

Determinant ratio. Let

$$\tau := (\tau_0,\ldots,\tau_k).$$

Kowalewski's definition of $\boldsymbol{\Delta}(\tau)f$ as the leading coefficient, in the power form, of the Hermite interpolant to f at τ gives, for the case of simple sites and via Cramer's Rule, the formula

$$\boldsymbol{\Delta}(\tau)f \;=\; \det Q_\tau[()^0,\ldots,()^{k-1},f]/\det Q_\tau[()^0,\ldots,()^k] \tag{45}$$

in which

$$Q_\tau[g_0,\ldots,g_k] := (g_j(\tau_i) : i,j=0,\ldots,k).$$

In some papers and books, the identity (45) serves as the definition of $\boldsymbol{\Delta}(\tau)f$ despite the fact that it needs awkward modification in the case of repeated sites.

Peano kernel (B-spline). Assume that $\tau := (\tau_0, \ldots, \tau_k)$ lies in the interval $[a\,..\,b]$ and that f has k derivatives on that interval. Then, on that interval, we have Taylor's identity

$$f(x) = \sum_{j<k} (x-a)^j D^j f(a)/j! + \int_a^b (x-y)_+^{k-1} D^k f(y)\,\mathrm{d}y/(k-1)!. \tag{46}$$

If now $\tau_0 < \tau_k$, then, from Proposition 35, $\Delta(\tau)$ is a weighted sum of values of derivatives of order $< k$, hence commutes with the integral in Taylor's formula (46) while, in any case, it annihilates any polynomial of degree $< k$. Therefore

$$\Delta(\tau) f = \int_a^b M(\cdot|\tau) D^k f/k!, \tag{47}$$

with

$$M(x|\tau) := k\,\Delta(\tau)(\cdot - x)_+^{k-1} \tag{48}$$

the Curry-Schoenberg **B-spline** (see [Curry & Schoenberg 1966]) with knots τ and normalized to have integral 1. While Schoenberg and Curry named and studied the B-spline only in the 1940's, it appears in this role as the Peano kernel for the divided difference already earlier, e.g., in [Popoviciu 1933] and [Tchakaloff 1934] (see [de Boor et al. 2003]) or [Favard 1940].

Contour integral. An entirely different approach to divided differences and Hermite interpolation begins with Frobenius' paper [Frobenius 1871], so different that it had no influence on the literature on interpolation (except for a footnote-like mention in [Chakalov 1938]). To be sure, Frobenius himself seems to have thought of it more as an exercise in expansions, never mentioning the word 'interpolation'. Nevertheless, Frobenius describes in full detail the salient facts of polynomial interpolation in the complex case, with the aid of the Cauchy integral.

In [Frobenius 1871], Frobenius investigates **Newton series**, i.e., *infinite* expansions

$$\sum_{j=1}^{\infty} c_j w_{j-1,t}$$

in the Newton polynomials $w_{j,t}$ defined in (2). He begins with the identity

$$(y-x) \sum_{j=1}^{n} \frac{w_{j-1,t}(x)}{w_{j,t}(y)} \;=\; 1 - \frac{w_{n,t}(x)}{w_{n,t}(y)}, \tag{49}$$

a ready consequence of the observations

$$\begin{aligned} x w_{j-1,t}(x) &= w_{j,t}(x) \;+\; t_j w_{j-1,t}(x), \\ \frac{y}{w_{j,t}(y)} &= \frac{1}{w_{j-1,t}(y)} \;+\; \frac{t_j}{w_{j-1,t}(y)} \end{aligned}$$

since these imply that

$$y\sum_{j=1}^{n}\frac{w_{j-1,t}(x)}{w_{j,t}(y)} = \sum_j \frac{w_{j-1,t}(x)}{w_{j-1,t}(y)} + \sum_j \frac{t_j w_{j-1,t}(x)}{w_{j,t}(y)},$$

$$x\sum_{j=1}^{n}\frac{w_{j-1,t}(x)}{w_{j,t}(y)} = \sum_j \frac{w_{j,t}(x)}{w_{j,t}(y)} + \sum_j \frac{t_j w_{j-1,t}(x)}{w_{j,t}(y)}.$$

Then (in §4), he uses (49), in the form

$$\sum_{j=1}^{n}\frac{w_{j-1,t}(z)}{w_{j,t}(\zeta)} + \frac{w_{n,t}(z)}{(\zeta-z)w_{n,t}(\zeta)} = 1/(\zeta - z),$$

in Cauchy's formula

$$f(z) = \frac{1}{2\pi\mathrm{i}}\oint \frac{f(\zeta)\,\mathrm{d}\zeta}{\zeta - z}$$

to conclude that

$$f(z) = \sum_{j=1}^{n} w_{j-1,t}c_j \; + \; w_{n,t}\frac{1}{2\pi\mathrm{i}}\oint \frac{f(\zeta)\,\mathrm{d}\zeta}{(\zeta-z)w_{n,t}(\zeta)}, \tag{50}$$

with

$$c_j \; := \; \frac{1}{2\pi\mathrm{i}}\oint \frac{f(\zeta)\,\mathrm{d}\zeta}{w_{j,t}(\zeta)}, \quad j = 1,\ldots,n.$$

For this, he assumes that z is in some disk of radius ρ, in which f is entire, and ζ runs on the boundary of a disk of radius $\rho' < \rho$ that contains z, with none of the relevant t_j in the annulus formed by the two disks.

Directly from the definition of the divided difference, we therefore conclude that, under these assumptions on f and t,

$$\Delta(t_{1:j})f \; = \; \frac{1}{2\pi\mathrm{i}}\oint \frac{f(\zeta)\,\mathrm{d}\zeta}{w_{j,t}(\zeta)}, \quad j = 0,1,2,\ldots. \tag{51}$$

Strikingly, Frobenius never mentions that (50) provides a general polynomial interpolant and its error. Could he have been unaware of it? To be sure, he could not have called it 'Hermite interpolation' since Hermite's paper [Hermite 1878] appeared well after Frobenius'. There is no indication that Hermite was aware of Frobenius' paper.

Genocchi-Hermite. Starting with (19) and the observation that

$$\Delta(x,y)f = \int_0^1 Df((1-s)x + sy)\,\mathrm{d}s,$$

induction on n gives the (univariate) Genocchi-Hermite formula

$$\Delta(\tau_0,\ldots,\tau_n)f = \int_{[\tau_0,\ldots,\tau_n]} D^n f, \tag{52}$$

with

$$\int_{[\tau_0,\ldots,\tau_n]} f \; := \\ \int_0^1 \int_0^{s_1} \cdots \int_0^{s_{n-1}} f((1-s_1)\tau_0 + \cdots + (s_{n-1}-s_n)\tau_{n-1} + s_n\tau_n)\,\mathrm{d}s_n \cdots \mathrm{d}s_1.$$

[Nörlund 1924: p.16] mistakenly attributes (52) to [Hermite 1859], possibly because that paper carries the suggestive title "Sur l'interpolation".

At the end of the paper [Hermite 1878], on polynomial interpolation to data at the n pairwise distinct sites $t_1,\ldots,t_n$ in the complex plane, Hermite does give a formula involving the righthand-side of the above, namely the formula

$$f(x) - Pf(x) = (x-t_1)\cdots(x-t_n)\int_{[t_n,\ldots,t_1,x]} D^n f$$

for the error in the Lagrange interpolant Pf to f at $t_{1:n}$. Thus, it requires the observation that

$$f(x) - Pf(x) = (x-t_1)\cdots(x-t_n)\,\Delta(t_n,\ldots,t_1,x)f$$

to deduce the Genocchi-Hermite formula from [Hermite 1878]. (He also gives the rather more complicated formula

$$f(x) - Pf(x) = \\ (x-a_1)^\alpha \cdots (x-a_n)^\lambda \int_{[a_n,\ldots,a_1,x]} [\![s_n - s_{n-1}]\!]^{\alpha-1} \cdots [\![1-s_1]\!]^{\lambda-1} D^{\alpha+\cdots+\lambda} f$$

for the error in case of repeated interpolation. Here, $[\![z]\!]^j := z^j/j!$.)

In contrast, [Genocchi 1869] is explicitly concerned with a *representation formula for the divided difference.* However, the 'divided difference' he represents is the following:

$$\Delta(x, x+h_1)\,\Delta(\cdot,\cdot+h_2)\cdots\Delta(\cdot,\cdot+h_n) = (\Delta_{h_1}/h_1)\cdots(\Delta_{h_n}/h_n)$$

and for it he gets the representation

$$\int_0^1 \cdots \int_0^1 D^n f(x + h_1t_1 + \cdots + h_nt_n)\,\mathrm{d}t_1 \cdots \mathrm{d}t_n.$$

[Nörlund 1924: p.16] cites [Genocchi 1878a], [Genocchi 1878b] as places where formulations equivalent to the Genocchi-Hermite formula can be found. So far, I've been only able to find [Genocchi 1878b]. It is a letter to Hermite, in which Genocchi brings, among other things, the above representation formula to Hermite's attention, refers to a paper of his in [Archives de Grunert, t. XLIX, 3e cahier] as containing a corresponding error formula for Newton interpolation. He states that he, in continuing work, had obtained such a representation also for Ampère's fonctions interpolatoires (aka divided differences), and finishes with the formula

$$\int_0^1 \cdots \int_0^1 s_1^{n-1} s_2^{n-2} \cdots s_{n-1} \\ D^n f(x_0 + s_1(x_1 - x_0) + \cdots + s_1 s_2 \cdots s_n(x_n - x_{n-1}))\, \mathrm{d}s_1 \cdots \mathrm{d}s_n$$

for $\Delta(x_0, \ldots, x_n)f$, and says that it is equivalent to the formula

$$\Delta(x_0, \ldots, x_n)f = \int \cdots \int D^n f(s_0 x_0 + s_1 x_1 + \cdots s_n x_n)\, \mathrm{d}s_1 \cdots \mathrm{d}s_n$$

in which the conditions $s_0 + \cdots + s_n = 1$, $s_i \geq 0$, all i, are imposed.

[Steffensen 1927: p.17f] proves the Genocchi-Hermite formula but calls it Jensen's formula, because of [Jensen 1894].

10 Divided difference expansions of the divided difference

By applying $\Delta(s_{1:m})$ to both sides of the identity

$$\Delta(\cdot) = \sum_{j=1}^{n} w_{j-1,t}\, \Delta(t_{1:j}) + w_{n,t}\, \Delta(t_{1:n}, \cdot)$$

obtained from (16), one obtains the expansion

$$\Delta(s_{1:m}) = \sum_{j=m}^{n} \Delta(s_{1:m}) w_{j-1,t}\, \Delta(t_{1:j}) + E(s,t),$$

where, by the Leibniz formula (29),

$$\begin{aligned} E(s,t) &:= \Delta(s_{1:m})(w_{n,t}\, \Delta(t_{1:n}, \cdot)) \\ &= \sum_{i=1}^{m} \Delta(s_{i:m}) w_{n,t}\, \Delta(s_{1:m}, t_{1:n}). \end{aligned}$$

But, following [Floater 2003] and with $p := n - m$, one gets the better formula

$$E(s,t) := \sum_{i=1}^{m} (s_i - t_{i+p})(\Delta(s_{1:i}) w_{i+p,t})\, \Delta(t_{1:i+p}, s_{i:m})$$

in which all the divided differences on the right side are of the same order, n. The proof (see [de Boor 2003a]), by induction on n, uses the easy consequence of (29) that

$$(s_i - y)\,\Delta(s_{i:m})f = \Delta(s_{i:m})((\cdot - y)f) - \Delta(s_{i+1:m})f.$$

The induction is anchored at $n = m$ for which the formula

$$\Delta(s_{1:m}) - \Delta(t_{1:m}) = \sum_{i=1}^{m}(s_i - t_i)\,\Delta(s_{1:i}, t_{i:m})$$

can already be found in [Hopf 1926].

11 Notation and nomenclature

It is quite common in earlier literature to use the notation

$$[y_1, \ldots, y_j]$$

for the divided difference of order $j-1$ of data $((t_i, y_i) : i = 1{:}j)$. This reflects the fact that divided differences were thought of as convenient expressions in terms of the given data rather than as linear functionals on some vector space of functions.

The presently most common notation for $\Delta(t_{1:j})p = \Delta(t_1, \ldots, t_j)p$ is

$$p[t_1, \ldots, t_j]$$

(or, perhaps, $p(t_1, \ldots, t_j)$) which enlarges upon the fact that $\Delta(z)p = p(z)$, but this becomes awkward when the divided difference is to be treated as a linear functional. In that regard, the notation

$$[t_1, \ldots, t_j]p$$

is better, but suffers from the fact that the resulting notation

$$[t_1, \ldots, t_j]$$

for the linear functional itself conflicts with standard notations, such as the matrix (or, more generally, the column map) with columns $t_1, \ldots, t_j$, or, in the special case $j = 2$, i.e.,

$$[t_1, t_2],$$

the closed interval with endpoints t_1 and t_2 or else the scalar product of the vectors t_1 and t_2. The notation

$$[t_1, \ldots, t_j; p]$$

does not suffer from this defect, as it leads to the notation $[t_1, \ldots, t_j; \cdot]$ for the linear functional itself, though it requires the reader not to mistakenly read that semicolon as yet another comma.

The notation, $\Delta\!\!\!|$, used in this essay was proposed by W. Kahan some time ago (see, e.g., [Kahan 1974]), and does not suffer from any of the defects mentioned and has the advantage of being literal (given that Δ is standard notation for a difference). Here is a TEX macro for it:

```
\def\divdif{\mathord\kern.43em\vrule width.6pt height5.6pt
              depth-.28pt \kern-.43em\Delta}
```

Although divided differences are rightly associated with Newton (because of [Newton 1687: Book iii, Lemma v, Case ii], [Newton 1711]; for a detailed discussion, including facsimiles of the originals and their translations, see [Fraser 1918], [Fraser 1919], [Fraser 1927]), the term 'divided difference' was, according to [Whittaker et al. 1937: p.20], first used in [de Morgan 1842: p.550], even though, by then, Ampère [Ampère 1826] had called it **fonction interpolaire**, and this is the term used in the French literature of the 1800s. To be sure, in [Newton 1711], Newton repeatedly uses the term "differentia sic divisa'.

Acknowledgments. On 20jun05, the change $i(1)$ `-->` $i(0)$ was made in the formula in Corollary 30. Thank you, Michael Floater!

References

Ampère, André-Marie [1826] Essai sur un noveau mode d'exposition des principes du calcul différentiel, du calcul aux différences et de l'interpolation des suites, considérées comme dérivant d'un source commune, *Annals de mathematicques pures et appliquées de Gergonne (Paris, Bachelier, in* 4^o*)* **16**, 329–349.

de Boor, C. [1972] On calculating with B-splines, *J. Approx. Theory* **6**, 50–62.

de Boor, C. [2003a] A divided difference expansion of a divided difference, *J. Approx. Theory* **122**, 10–12.

de Boor, C. [2003b] A Leibniz formula for multivariate divided differences, *SIAM J. Numer. Anal.* **41(3)**, 856–868.

de Boor, C. and A. Pinkus [2003] The B-spline recurrence relations of Chakalov and of Popoviciu, *J. Approx. Theory* **124**, 115–123.

Bulirsch, R. and H. Rutishauser [1968] Interpolation und genäherte Quadratur, in *Mathematische Hilfsmittel des Ingenieurs, Teil III*, S. Sauer & I. Szabó, eds, Grundlehren vol. 141, Springer-Verlag, Heidelberg, 232–319.

Cauchy, Augustin [1840] Sur les fonctions interpolaires, *C. R. Acad. Sci. Paris* **11**, 775–789.

Tchakaloff, L. [1934] Sur la structure des ensembles linéaires définis par une certaine propriété minimale, *Acta Math* **63**, 77–97.

Chakalov, L. [1938] On a certain presentation of the Newton divided differences in interpolation theory and its applications (in Bulgarian), *Annuaire Univ. Sofia, Fiz. Mat. Fakultet* **34**, 353–394.

Curry, H. B. and I. J. Schoenberg [1966] On Pólya frequency functions IV: the fundamental spline functions and their limits, *J. Analyse Math.* **17**, 71–107.

Erdős, P. and P. Turán [1940] On interpolation, III. Interpolatory theory of polynomials, *Appl. Math.(2)* **41**, 510–553.

Favard, J. [1940] Sur l'interpolation, *J. Math. Pures Appl.* **19**, 281–306.

Floater, M. [2003] Error formulas for divided difference expansions and numerical differentiation, *J. Approx. Theory* **122**, 1–9.

Fraser, Duncan C. [1918] Newton's interpolation formulas, *J. Instit. Actuaries* **LI**, 77–106.

Fraser, Duncan C. [1919] Newton's interpolation formulas. Further Notes, *J. Instit. Actuaries* **LI**, 211–232.

Fraser, Duncan C. [1927] Newton's interpolation formulas. An unpublished manuscript of Sir Isaac Newton, *J. Instit. Actuaries* **LVIII**, 53–95.

Frobenius, G. [1871] Über die Entwicklung analytischer Functionen in Reihen, die nach gegebenen Functionen fortschreiten, *J. reine angew. Math.* **73**, 1–30.

Genocchi, A. [1869] Relation entre la différence et la dérivée d'un même ordre quelconque, *Archiv Math. Phys. (I)* **49**, 342–345.

Genocchi, A. [1878a] Intorno alle funzioni interpolari, *Atti della Reale Accademia delle Scienze di Torino* **13**, 716–729.

Genocchi, A. [1878b] Sur la formule sommatoire de Maclaurin et les fonctions interpolaires, *C. R. Acad. Sci. Paris* **86**, 466–469.

Hermite, Ch. [1859] Sur l'interpolation, *C. R. Acad. Sci. Paris* **48**, 62–67.

Hermite, Ch. [1878] Formule d'interpolation de Lagrange, *J. reine angew. Math.* **84**, 70–79.

Hopf, Eberhard [1926] "Über die Zusammenhänge zwischen gewissen höheren Differenzenquotienten reeller Funktionen einer reellen Variablen und deren Differenzierbarkeitseigenschaften", dissertation, Universität Berlin (30 pp).

Jensen, J. L. W. V. [1894] Sure une expression simple du reste dans la formule d'interpolation de Newton, *Bull. Acad. Roy. Danemark* **xx**, 246–xxx.

Kahan, W. [1974] Divided differences of algebraic functions, Class notes for Math 228A, UC Berkeley, Fall.

Kowalewski, Arnold [1917] "Newton, Cotes, Gauss, Jacobi: Vier grundlegende Abhandlungen über Interpolation und genäherte Quadratur", Teubner, Leipzig.

Kowalewski, G. [1932] "Interpolation und genäherte Quadratur", B. G. Teubner, Berlin.

de Morgan, A. [1842] "The Differential and Integral Calculus", Baldwin and Cradock, London, UK.

Newton, I. [1676] Epistola posterior (Second letter to Leibniz via Oldenburg), 24 oct.

Newton, I. [1687] "Philosophiæ naturalis principia mathematica", Joseph Streater, London. (see
`ftp://ftp.math.technion.ac.il/hat/fpapers/newton1.pdf` and
`ftp://ftp.math.technion.ac.il/hat/fpapers/newton2.pdf`)

Newton, I. [1711] "Methodus differentialis", Jones, London.

Nörlund, N. E. [1924] "Vorlesungen über Differenzenrechnung", Grundlehren Vol. XIII, Springer, Berlin.

Opitz, G. [1964] Steigungsmatrizen, *Z. Angew. Math. Mech.* **44**, T52–T54.

Popoviciu, Tibère [1933] "Sur quelques propriétés des fonctions d'une ou de deux variables réelles", dissertation, presented to the Faculté des Sciences de Paris, published by Institutul de Arte Grafice "Ardealul" (Cluj, Romania).

Popoviciu, Tiberiu [1940] Introduction à la théorie des différences divisées, *Bull. Mathem., Societea Romana de Stiinte, Bukharest* **42**, 65–78.

Schwarz, H. A. [1881-2] Démonstration élémentaire d'une propriété fondamentale des fonctions interpolaires, *Atti Accad. Sci. Torino* **17**, 740–742.

Steffensen, J. F. [1927] "Interpolation", Williams and Wilkins, Baltimore.

Steffensen, J. F. [1939] Note on divided differences, *Danske Vid. Selsk. Math.-Fys. Medd.* **17(3)**, 1–12.

Taylor, Brook [1715] "Methodus incrementorum directa et inversa", London, England.

Whittaker, E. T. and G. Robinson [1937] "The Calculus of Observations, A Treatise on Numerical Mathematics, 2nd ed.", Blackie, Glasgow, UK.

C. de Boor
P.O. Box 1076
Eastsound WA 98245 USA
`deboor@cs.wisc.edu`
`http://www.cs.wisc.edu/~deboor/`