A Multivariate Divided Difference

C. de Boor

Abstract. A multivariate kth divided difference is proposed and shown to lead to new error formulæ for polynomial interpolation on some standard multivariate point sets.

§1. Introduction

In trying to understand what might be a reasonable way to describe the error in interpolation by polynomials on $\mathbb{R}^d$, I came upon the following considerations:

The standard error formula

$$g(x) = g(a) + \int_0^1 D_{x-a} g(a + t(x-a))\, \mathrm{d}t$$

for the simplest possible case, that of interpolation by constant polynomials at a point, a, can be written in the following form

$$g(x) = g(a) + \sum_j \langle v_j, x-a \rangle \, [a, x; \varphi_j] g, \tag{1.1}$$

in which

$$\langle x, y \rangle := \sum_{j=1}^d x(j) y(j)$$

is the standard scalar product on $\mathbb{R}^d$, $(v_1, \ldots, v_d)$ is an arbitrary basis for $\mathbb{R}^d$, with $(\varphi_1, \ldots, \varphi_d)$ its dual basis, i.e.,

$$\langle v_i, \varphi_j \rangle = \delta_{ij},$$

Approximation Theory VIII, Vol. 1: Approximation and Interpolation
Charles K. Chui and Larry L. Schumaker (eds.), pp. 87–96.

ISBN 981-02-2971-2

and, for $x, y, \xi \in \mathbb{R}^d$, $[x, y; \xi]$ denotes the linear functional

$$[x, y; \xi] : g \mapsto \int_0^1 D_\xi g(x + t(y - x))\,\mathrm{d}t. \tag{1.2}$$

Now, for $j = 1, \ldots, d$, let $x_j \neq a$ be a point on the straight line common to the $d-1$ hyperplanes $h_i^{-1}\{0\}$, $i \neq j$, with

$$h_i : x \mapsto \langle v_i, x - a \rangle,$$

i.e., $x_j = \alpha_j \varphi_j + a$ for some $\alpha_j \neq 0$. Then

$$p_1 := g(a) + \sum_{j=1}^{d} h_j\,[a, x_j; \varphi_j]g$$

is a linear polynomial which, by (1.1), agrees with g at a *and* at $x_1, \ldots, x_d$, hence is the unique linear polynomial with that property. What about its error?

Evidently,

$$g = p_1 + \sum_{j=1}^{d} h_j\,\big([a,\ \cdot\,; \varphi_j] - [a, x_j; \varphi_j]\big)g,$$

and this involves the error in the interpolation at x_j to the function $[a,\ \cdot\,; \varphi_j]g$, hence *invites application of (1.1)*, giving

$$[a, x; \varphi_j] - [a, x_j; \varphi_j] = \sum_{i=1}^{d} \langle v_{ji}, x - x_j \rangle [x_j, x; \varphi_{ji}]\,[a,\ \cdot\,; \varphi_j],$$

where, for different j, we are free to choose different bases $(v_{j1}, \ldots, v_{jd})$ if that suits our ultimate purpose.

It is now easy to imagine repetition of this process, i.e., the repeated expansion of some or even all the error terms, having in hand at all times a formula for g of the form

$$g = p + e,$$

with p a polynomial, and e an explicit expression for the error which vanishes at certain points, thus identifying p as a polynomial which interpolates g at those points, and with both p and e involving terms of the form

$$[x_1, \ldots, x_k,\ \cdot\,; \xi_1, \ldots, \xi_k]g \ := \ [x_1,\ \cdot\,; \xi_1][x_2,\ \cdot\,; \xi_2] \cdots [x_k,\ \cdot\,; \xi_k]g, \tag{1.3}$$

evaluated at some point when occurring in the formula for p.

It is the purpose of this note to suggest that (1.3), evaluated at some point, be called a *kth divided difference*, and to record its basic properties.

§2. Historical Comments

While tensor products of univariate divided differences have long been used, more general multivariate divided differences (in Approximation Theory; see further comments below) all seem to be traceable to Kergin interpolation [8], or, more precisely, to Micchelli's explanation and treatment [10] (see, also [12]) of Kergin interpolation. This treatment involves the functional

$$f \mapsto \int_{[x_0,\dots,x_k]} f := \int_0^1 \int_0^{s_1} \cdots \int_0^{s_{k-1}} f(x_0+s_1\nabla x_1+\cdots+s_k\nabla x_k)\mathrm{d}s_k \cdots \mathrm{d}s_1 \tag{2.1}$$

(with $\nabla x_j := x_j - x_{j-1}$) called the *divided difference functional on* $\mathbb{R}^d$ by Micchelli in [11], and familiar from the Hermite-Genocchi formula for the univariate divided difference which, in these terms, reads

$$[t_0,\dots,t_k]g = \int_{[t_0,\dots,t_k]} D^k g. \tag{2.2}$$

In particular, it is not hard to show directly (as is done later in this note) that, in these terms,

$$[x_1,\dots,x_{k+1};\xi_1,\dots,\xi_k]g = \int_{[x_1,\dots,x_{k+1}]} D_{\xi_1}\cdots D_{\xi_k} g. \tag{2.3}$$

The section entitled "Multivariate divided differences" in Dahmen [5] contains such integrals, albeit with the integrand $\left(\prod_{1\le i<j\le k+1} D_{x_i-x_j}\right) g$ instead of $\left(\prod_{i=1}^k D_{\xi_i}\right) g$. Subsequently, Hakopian [6] defined a multivariate divided difference to be something of the form of the right-hand side of (2.3), but restricted the directions ξ_j to be coordinate vectors. Hence, one message of this note is that it pays to work with the more general definition (1.3).

In fact, Kergin interpolation involves the even more general functionals

$$g \mapsto \int_{[x_0,\dots,x_k]} q_k(D)g,$$

with q_k a homogeneous polynomial (of degree k). While these are not, in general, kth divided differences in the sense of (1.3), they do make up the linear span of all such kth divided differences. A more thorough discussion of the literature on multivariate divided differences, together with a detailed discussion of a rather different notion of multivariate divided difference, can be found in [9].

Finally, Micchelli's analysis of the simplex spline $M(\,\cdot\,|x_0,\ldots,x_k)$ is based on his observation that $M(\,\cdot\,|x_0,\ldots,x_k)/k!$ represents the functional $\int_{[x_0,\ldots,x_k]}$. This has led others to derive facts about this functional from facts about its representer. However, for many purposes, it seems simpler to work directly with the functional; compare, e.g., [3] with [14].

I am grateful to Florian Potra for having reminded me that the literature concerning the numerical solution of functional equations has, for many years, used the notion of a divided difference of operators in Banach spaces. Specifically, J. Schröder [15] in 1956 proposed to call, for a given map $f : X \to Z$ with X and Z normed linear spaces, any linear map

$$f_{\{x,y\}} : X \to Z$$

with the property

$$f_{\{x,y\}}(x-y) = f(x) - f(y)$$

a *divided difference of f at x and y*. While this definition fits, in general, many linear maps, Byelostotskij [1] seems to have been the first to have used specifically the linear map

$$f_{\{x,y\}} := \int_0^1 f^{(1)}(x + t(y-x))\mathrm{d}t,$$

with $f^{(1)}$ the Fréchet derivative of f. With this definition, if f is a scalar-valued map on some domain of $\mathbb{R}^d$, hence $f_{\{x,y\}}$ is a linear functional, then

$$[x, y; \xi]f = f_{\{x,y\}}\xi.$$

The first to have defined and used higher-order divided differences of operators on normed linear spaces seems to have been Serge'ev [16] who, in 1961, introduced the second divided difference as a divided difference of a divided difference, in a study of Steffensen iteration.

The same idea was used by Potra and Ptak, in [13: Appendix B], to define a kth order divided difference $[x_1,\ldots,x_{k+1}; f]$ of $f : D \subset X \to Z$ as a divided difference $[x_k, x_{k+1}; [x_1,\ldots,x_{k-1},\,\cdot\,; f]]$ of a $(k-1)$st divided difference. Moreover, they point out (without mentioning Hermite-Genocchi and in a different but equivalent formulation of (2.1)) that $\int_{[x_1,\ldots,x_{k+1}]} f^{(k)}$ is such a kth divided difference, with $f^{(k)}$ the kth Fréchet derivative of f. The value of this k-linear map at the k-sequence $(\xi_1,\ldots,\xi_k)$ is exactly what I denoted here, for the special case $X = \mathbb{R}^d$, $Z = \mathbb{R}$, by $[x_1,\ldots,x_{k+1}; \xi_1,\ldots,\xi_k]f$.

§3. Basic Properties

The divided difference $[x, y; \xi]$ defined in (1.2) is linear in ξ and satisfies

$$[x, y; y]g - [x, y; x]g = [x, y; y - x]g = g(y) - g(x). \tag{3.1}$$

In particular, if $d = 1$, then $[x, y; 1]$ is the ordinary divided difference:

$$[x, y; 1]g = [x, y; y - x]g/(y - x) = (g(y) - g(x))/(y - x) = [x, y]g. \tag{3.2}$$

However, for $d > 1$ and general ξ, $[x, y; \xi]g$ is not a scaled difference of values of g, hence the use of the term 'divided difference' might be misleading. Still, as we hope to illustrate here and explore in more detail elsewhere, this 'divided difference' can be made to play the same role in multivariate polynomial interpolation that (3.2) plays so successfully in the classic, univariate context, hence deserves the name for that reason.

$[x, y; \xi]$ is a linear functional; also, it is symmetric in x and y. On smooth functions, it is continuous in x and y and reduces to $g \mapsto D_\xi g(x)$ when $x = y$.

The discussion in the Introduction motivated the definition (1.3) of a multivariate kth divided difference. We now prove its identification (2.3). For this, it is convenient to start off with the following alternative definition (which differs from (1.3) in the order in which the first-order divided differences appear on the right): For smooth functions g and for arbitrary sequences $(x_1, \ldots, x_k)$ and $(\xi_1, \ldots, \xi_k)$, we define

$$[x_1, \ldots, x_k, \cdot; \xi_1, \ldots, \xi_k]g := [x_k, \cdot; \xi_k] \cdots [x_1, \cdot; \xi_1]g. \tag{3.3}$$

For $k = 0$, this defines $[x_1;]$ to be evaluation at x_1, hence (2.3) holds for $k = 0$ with the reasonable definition $\int_{[x_0]} g := g(x_0)$ (which should have been part of (2.1)). Assuming (2.3) to hold for $k = n-1$, we conclude that

$$\begin{gathered}[x_1, \ldots, x_n, \cdot; \xi_1, \ldots, \xi_n]g = [x_n, \cdot; \xi_n][x_1, \ldots, x_{n-1}, \cdot; \xi_1, \ldots, \xi_{n-1}]g = \\ \int_0^1 D_{\xi_n} \left[\int_0^1 \int_0^{s_1} \cdots \int_0^{s_{n-2}} F(s_1, \ldots, s_{n-1}, \cdot) \mathrm{d}s_{n-1} \cdots \mathrm{d}s_1 \right] (x_n + t(\cdot - x_n)) \mathrm{d}t,\end{gathered} \tag{3.4}$$

with

$$F(s_1, \ldots, s_{n-1}, \cdot) := (D_{\xi_1} \cdots D_{\xi_{n-1}} g)(x_1 + s_1 \Delta x_1 + \cdots + s_{n-1}(\cdot - x_{n-1}))$$

(and $\Delta x_j := x_{j+1} - x_j$), hence

$$D_{\xi_n} F(s_1, \ldots, s_{n-1}, \cdot) = (D_\Xi g)(x_1 + s_1 \Delta x_1 + \cdots + s_{n-1}(\cdot - x_{n-1})) s_{n-1},$$

with

$$D_\Xi g := D_{\xi_1} \cdots D_{\xi_n} g.$$

Therefore, interchanging, in (3.4), D_{ξ_n} with the integrals, we find that $[x_1, \ldots, x_n, \cdot\,; \xi_1, \ldots, \xi_n]g$ equals

$$\int_0^1 \int_0^1 \int_0^{s_1} \cdots \int_0^{s_{n-2}} (D_\Xi g)(x_1 + s_1 \Delta x_1 + \cdots + s_{n-1}((x_n + t(\cdot - x_n)) - x_{n-1})) \\ s_{n-1} \mathrm{d}s_{n-1} \cdots \mathrm{d}s_1 \mathrm{d}t.$$

Now, interchange integration with respect to t with all the other integrations and introduce the new variable $s_n := s_{n-1}t$, hence $\mathrm{d}s_n = s_{n-1}\mathrm{d}t$, to obtain, finally,

$$[x_1, \ldots, x_n, \cdot\,; \xi_1, \ldots, \xi_n]g = \\ \int_0^1 \int_0^{s_1} \cdots \int_0^{s_{n-1}} (D_\Xi g)(x_1 + s_1 \Delta x_1 + \cdots + s_{n-1} \Delta x_{n-1} + s_n(\cdot - x_n)) \\ \mathrm{d}s_n \cdots \mathrm{d}s_1,$$

which, evaluated at x_{n+1}, gives (2.3) for $k = n$.

It follows from (2.3) that *the divided difference* $[x_1, \ldots, x_{k+1}; \xi_1 \ldots, \xi_k]$ *is symmetric and continuous in its points,* $x_1, \ldots, x_{k+1}$, *and is symmetric, continuous and multilinear in its directions,* $\xi_1, \ldots, \xi_k$ (as long as it is applied to smooth functions). In particular, the order reversal in the definition (3.3), while convenient for the proof of (2.3), is ultimately irrelevant. This permits use of the convenient shorthand

$$[X; \Xi] := [x_1, \ldots, x_{k+1}; \xi_1, \ldots, \xi_k]$$

in which $x_1, \ldots, x_{k+1}$ are the entries, written in any particular order, of the $k+1$-sequence X and $\xi_1, \ldots, \xi_k$ are the entries, in any convenient order, of the k-sequence Ξ.

For example, it follows that, for any (point) sequences X, X', and corresponding (direction) sequences Ξ, Ξ',

$$[X; \Xi][X', \cdot\,; \Xi'] = [X, X'; \Xi, \Xi']. \tag{3.5}$$

As another example, (3.1) has the generalization

$$[X, a, b; \Xi, a - b] = [X, a; \Xi] - [X, b; \Xi], \tag{3.6}$$

which, as Vladimir Yegorov has pointed out to me, does not require the proof I gave for it since it follows directly from (3.5) and the observation that $[X, a; \Xi] - [X, b; \Xi] = [a, b; a - b][X, \cdot\,; \Xi]$, by (3.1).

The technique of *lifting* proposed and used in [11] is a ready means for deriving multivariate divided difference identities from their univariate case, and (2.3), (3.5), (3.6) could have been so derived (cf. [11:Thm 3]).

§4. Some Illustrations

Here are three error formulas for multivariate polynomial interpolation. The first one is well known, the other two are new; one is led to them by the construction outlined in the Introduction.

Kergin interpolation. The identity

$$[x;] = [x_0;] + [x_0, x; x - x_0]$$

is the standard error formula (for interpolation at some point) with which I started this note. If it is applied to its last term as a function of the first occurrence of x there, i.e., to the function $z \mapsto [x_0, z; x - x_0]$, we obtain

$$[x;] = [x_0;] + [x_0, x_1; x - x_0] + [x_0, x_1, x; x - x_0, x - x_1].$$

Repetition provides Micchelli's (see [10], also Micchelli and Milman [12]) formula

$$[x;] = \sum_{j=0}^{k} [x_0, \ldots, x_j; x - x_0, \ldots, x - x_{j-1}] + [x_0, \ldots, x_k, x; x - x_0, \ldots, x - x_k]$$

for Kergin's interpolation [8], in which, because of the multilinearity of the divided difference in its directions, we recognize

$$x \mapsto [x_0, \ldots, x_j; x - x_0, \ldots, x - x_{j-1}]$$

as a polynomial of degree j which vanishes at $x_0, \ldots, x_{j-1}$.

Chung-Yao interpolation. By this I mean interpolation from Π_k (the space of d-variate polynomials of degree $\le k$) at the points

$$\Theta_{\mathbb{H}} := \{\theta_H : H \in \binom{\mathbb{H}}{d}\},$$

with $\mathbb{H}$ a set of $d + k$ hyperplanes in $\mathbb{R}^d$ in general position and θ_H the unique point common to the d hyperplanes in such an $H \in \binom{\mathbb{H}}{d}$. Chung and Yao [4] were the first to show that such interpolation is possible and uniquely so, by exhibiting the interpolant $P_{\mathbb{H}} g$ to g in Lagrange form. The following error formula (see [2]) is, in effect, the *Newton form* for this interpolant:

$$g(x) - P_{\mathbb{H}} g(x) = \sum_{K \in \binom{\mathbb{H}}{d-1}} p_K(x) \, [\Theta_K, x; n_K, \ldots, n_K] g. \tag{4.1}$$

Here,

$$p_K(x) := \prod_{h\in\mathbb{H}\backslash K} \frac{h(x)}{h_\uparrow(n_K)},$$

with h denoting a particular hyperplane as well as a particular linear polynomial whose zero set coincides with that hyperplane, and $h_\uparrow$ its *leading term*, i.e., its linear homogeneous part; with

$$\Theta_K := \Theta \cap \bigcap K$$

the $k+1$ points from Θ on the unique straight line

$$\cap K := \cap_{h\in K} h$$

common to the $d-1$ hyperplanes in K; and with

$$n_K$$

some nontrivial vector parallel to $\cap K$.

For $k = 1$, (4.1) reduces to an error formula due to S. Waldron [17]. This would be the formula reached if the procedure outlined in the Introduction were applied with the choice of the *directions*

$$(\varphi_{ji} : i = 1, \ldots, d) := (x_i - x_j : i \neq j),$$

where $x_0 := a$. The general formula (4.1) is proved from this by induction on k, as is natural for the procedure outlined in the Introduction.

It is intriguing to explore the behavior of (4.1) as $\mathbb{H}$ approaches nongeneric configurations.

Tensor product interpolation. Because of lack of space, only the simplest case, of interpolation at the points of a bivariate mesh

$$(x_0, \ldots, x_k) \times (y_0, \ldots, y_h),$$

is given here.

We write the interpolant to given g in the standard form $(P\otimes Q)g$, with P and Q the corresponding maps of univariate polynomial interpolation. All error formulæ in the literature involve $D_1^{k+1}D_2^{h+1}g$ in addition to $D_1^{k+1}g$ and $D_2^{h+1}g$. However, the process outlined in the Introduction (when, in contrast to the preceding example, one chooses always the coordinate vectors $\mathbf{i}_j$ as the directions in the expansions (1.1) used) leads ultimately to the following formula for the error at the point $(x, y) \in \mathbb{R}^2$:

$$\begin{aligned}(g - (P\otimes Q)g)(x,y) = p(x)\,(Q[(x_0,\cdot),\ldots,(x_k,\cdot),(x,y);\mathbf{i}_1,\ldots,\mathbf{i}_1]g)(y)\\ +\; q(y)\,(P[(\cdot,y_0),\ldots,(\cdot,y_h),(x,y);\mathbf{i}_2,\ldots,\mathbf{i}_2]g)(x),\end{aligned}$$

in which only the two pure derivatives, $p_\uparrow(D)g = D_1^{k+1}g$ and $q_\uparrow(D)g = D_2^{h+1}g$, appear. In this formula, the term multiplying, e.g., $p(x)$ is the value at y of the (polynomial) interpolant, at $y_0, \ldots, y_h$, to the univariate function

$$t \mapsto [(x_0, t), \ldots, (x_k, t), (x, y); \mathbf{i}_1, \ldots, \mathbf{i}_1]g.$$

It can be shown (see [2]) that such an error formula, in terms of an H-basis $(p, q, \ldots)$ for the ideal of the polynomials vanishing at the set Θ of interpolation points, is available in the more general situation when Θ consists of a lower set of grid points from a rectangular grid in $\mathbb{R}^d$ (as it is for Chung-Yao interpolation; see (4.1)). Two extreme special cases are (i) the whole grid, i.e., tensor product interpolation, in which case the ideal is generated by just d polynomials, each the unique monic polynomial of minimal degree in one variable which vanishes on the entire grid; and (ii) the simplex grid, corresponding to interpolation from Π_k, in which case one H-basis for the ideal is $(p_\alpha : |\alpha| = k + 1)$, with p_α the unique polynomial with leading term $()^\alpha$ which vanishes on Θ (but this case has surely been done before; see, e.g., [7:p.298] for the bivariate case).

Acknowledgments. Supported by NSF grant DMS-9224748 and by ARO grant 32830-MA.

References

1. Byelostotskij, A. Ja., Some methods for the solution of functional equations (Russian), *Uspekhi Matem. Nauk* **17(5)** (1962), 192–193.
2. Boor, Carl de, Newton form and error in multivariate polynomial interpolation, ms, dec, 1994.
3. Boor, Carl de, On the Sauer-Xu formula for the error in multivariate polynomial interpolation, *Math. Comp.* **65** (1996), 1231–1234.
4. Chung, K. C. and T. H. Yao, On lattices admitting unique Lagrange interpolations, *SIAM J. Numer. Anal.* **14** (1977), 735–743.
5. Dahmen, W., On multivariate B-splines, *SIAM J. Numer. Anal.* **17** (1980), 179–191.
6. Hakopian, Hakop A., Multivariate divided differences and multivariate interpolation of Lagrange and Hermite type, *J. Approx. Theory* **34** (1982), 286–305.
7. Isaacson, Eugene and Herbert B. Keller, *Analysis of Numerical Methods*, John Wiley & Sons, New York, 1966.
8. Kergin, P., A natural interpolation of C^k functions, *J. Approx. Theory* **29** (1980), 278–293.
9. Kunkle, Thomas, Multivariate differences, polynomials, and splines, *J. Approx. Theory* **84(3)** (1996), 290–314.

10. Micchelli, C. A., A constructive approach to Kergin interpolation in $\mathbb{R}^k$: multivariate B-splines and Lagrange interpolation, *Rocky Mountain J. Math.* **10** (1980), 485–497.
11. Micchelli, C. A., On a numerically efficient method for computing multivariate B-splines, in *Multivariate Approximation Theory*, W. Schempp and K. Zeller (eds), Birkhäuser (Basel), 1979, 211–248.
12. Micchelli, C. A. and P. Milman, A formula for Kergin interpolation in $\mathbb{R}^k$, *J. Approx. Theory* **29** (1980), 294–296.
13. Potra, F.-A. and V. Ptak, *Nondiscrete induction and iterative processes*, Research Notes in Mathematics 103, Pitman Publishing Inc (Marshfield MA), 1984.
14. Sauer, T. and Yuan Xu, On multivariate Lagrange interpolation, *Math. Comp.* **64** (1995), 1147–1170.
15. Schröder, Johann, Nichtlineare Majoranten beim Verfahren der schrittweisen Näherung, *Archiv der Math.* **7** (1956), 471–484.
16. Serge'ev, A. S., On the method of chords, *Sibirsk. Matem. Zhournal* **2** (1961), 282–289.
17. Waldron, S., The error in linear interpolation at the vertices of a simplex, *SIAM J. Numer. Anal.* **35(3)** (1998), 1191–1200.
18. Waldron, S., A multivariate form of Hardy's inequality and L_p-error bounds for multivariate Lagrange interpolation schemes, *SIAM J. Math. Anal.* **28(1)** (1997), 233–258.

C. de Boor
CS, U. Wisconsin, 1210 W. Dayton St., Madison WI 53706-1685 USA
deboor@cs.wisc.edu

25mar02: supplied missing g in (4.1) and updated the references