# Multivariate approximation*

Carl de Boor
Technical Summary Reprt #2950
August 1986
UNIVERSITY OF WISCONSIN-MADISON
Mathematics Research Center
610 Walnut Street
Madison  WI 53705
USA

## Abstract

The lecture addresses topics in multivariate approximation which have caught the author's interest in the last ten years. These include: the approximation by functions with fewer variables, correct points for polynomial interpolation, the B(ernstein,-ézier, -arycentric)-form for polynomials and its use in understanding smooth piecewise polynomial (pp) functions, approximation order from spaces of pp functions, multivariate B-splines, and surface generation by subdivision.

**AMS (MOS) Subject Classification:** 41-02, 41A05, 41A10, 41A15, 41A63, 41A65

**Key Words:** multivariate, interpolation, B-form, subdivision, splines, B-splines

Work Unit Number 3 – Numerical Analysis and Scientific Computing

* This paper has appeared in *State of the Art in Numerical Analysis*, A. Iserles and M. Powell (eds.), Institute Mathematics Applications (Essex); 1987; 87–109;

## SIGNIFICANCE AND EXPLANATION

This lecture was given at the IMA/SIAM conference on "State of the Art in Numerical Analysis" held April 14–18, 1986, in Birmingham, England. The lecture reviews developments in Multivariate Approximation in the last ten years. The selection of topics is quite subjective; it reflects entirely the author's research experience during that time.*

* The responsibility for the wording and views expressed in this descriptive summary lies with MRC, and not with the author of this report.

MULTIVARIATE APPROXIMATION

Carl de Boor

## 1. The set-up

The talk concerns the approximation of

$$f : G \subseteq \mathbb{R}^d \longrightarrow \mathbb{R},$$

i.e., of some real-valued function $f$ defined on some domain $G$ in $d$-dimensional space.

While my first publication dealt with such multivariate (actually, bivariate) approximation, I have concerned myself seriously with multivariate approximation only in the last ten years. This talk reflects some of the experiences I have had during that time.

The approximating functions are typically polynomials or piecewise polynomials, and just how one describes them will have an effect on one's work with them. Papers on multivariate approximation often sink under the burden of cumbersome notation. As a preventive measure against such a sad fate, I shall follow the "default" convention whereby symbols are left out if they can reasonably be guessed from the context. For example, if $x = \big(x(1), \ldots, x(d)\big)$ has just been declared to be a point in $\mathbb{R}^d$, I will feel free to write

$$\sum_i x(i) \quad \text{instead of} \quad \sum_{i=1}^{d} x(i).$$

For polynomials, **multi-index** notation is standard. With $x = (x(1), \ldots, x(d))$ the generic point in $\mathbb{R}^d$, one uses the abbreviation

$$x^\alpha := \prod_i x(i)^{\alpha(i)}, \quad x \in \mathbb{R}^d\ ,\ \alpha \in \mathbb{Z}_+^d.$$

The function $x \mapsto x^\alpha$ is a monomial **of degree** $\alpha$, or, of **(total) degree**

$$|\alpha| := \sum_i \alpha(i)$$

if only the exponent sum matters. More generally, a **polynomial of degree** $\leq \alpha$ is, by definition, any function of the form

$$x \mapsto \sum_{\beta \leq \alpha} x^\beta c(\beta),$$

Sponsored by the United States Army under Contract No. DAAG29-80-C-0041.

with real coefficients $c(\beta)$. The collection of all such polynomials is denoted by

$$\Pi_{\leq\alpha} = \Pi_{\leq\alpha}(\mathbb{R}^d),$$

the collection of all polynomials of total degree at most $k$ by

$$\Pi_{\leq k} = \Pi_{\leq k}(\mathbb{R}^d),$$

and the collection of all polynomials, of whatever degree, by

$$\Pi = \Pi(\mathbb{R}^d).$$

Many expressions simplify if one makes use of the **normalized power** function, i.e., the function

$$[\![]\!]^\alpha : x \mapsto x^\alpha/\alpha! \ := \ \prod_i x(i)^{\alpha(i)}/\alpha(i)!.$$

For example, with $\alpha, \xi, \upsilon, \ldots, \zeta \in \mathbb{Z}_+^d$, the **Multinomial Theorem** takes the simple form

$$[\![x + y + \cdots + z]\!]^\alpha = \sum_{\xi+\upsilon+\cdots+\zeta=\alpha} [\![x]\!]^\xi [\![y]\!]^\upsilon \cdots [\![z]\!]^\zeta. \tag{1.1}$$

At times, it pays to give up on the power form altogether. In some contexts, it is very convenient to describe polynomials in terms of the particular **homogenous** polynomials

$$\langle Y, \cdot\rangle : x \mapsto \prod_{y\in Y} \langle y, x\rangle,$$

with

$$\langle y, x\rangle := \sum_i y(i)x(i)$$

and $Y$ a **finite** subset of $\mathbb{R}^d$. I will make use of this form in Section 3. A particular instance is the B(ernstein-ézier)-form, which is the form of choice when dealing with piecewise polynomials on a triangulation (see Section 4).

## 2. Approximation by functions of fewer variables

The simplest approach to multivariate approximation uses **tensor products**, i.e., linear combinations of functions of the form

$$x \mapsto \prod_i g_i\big(x(i)\big),$$

each $g_i$ being a univariate function. This neatly avoids dealing with the realities of multivariate functions, but its effectiveness depends on having the information about $f$ correspondingly available in (cartesian) product form, e.g., as function values on a rectangular grid parallel to the coordinate axes. The recent book by Light and Cheney (1986) provides up-to-date material on this practically very important choice of approximating function and certain ready extensions.

The book also deals with the situation when the information about $f$ is not in product form. In that case, tensor product approximants still are attractive since they are composed of univariate functions. The general question of how to approximate a multivariate function by functions of fewer variables has received much attention. A reference with a Numerical Analysis slant is Golomb (1959).

The most remarkable result along this line is

**Kolmogorov's Theorem (Kolmogorov, 1957)**

$\exists \lambda \in ]0\,..\,1]^d$, $\exists \Phi \subseteq \mathrm{Lip}_\alpha[0\,..\,1] \cap$ strictly monotone, $\#\Phi = 2d+1$, such that
$\forall f \in C[0\,..\,1]^d \;\; \exists g \in C[0\,..\,d]$

$$f(x) = \sum_{\varphi\in\Phi} g\Big(\sum_i \lambda(i)\varphi\big(x(i)\big)\Big). \tag{2.1}$$

(Here and below,

$$\#A := \text{the number of elements in } A.)$$

The theorem claims the existence of a set $\Psi$ of $2d+1$ 'universal' maps $\psi : [0\,..\,1]^d \to [0\,..\,d]$ so that, for each continuous function $f$ on the unit cube $[0\,..\,1]^d$, a continuous function $g$ on the interval $[0\,..\,d]$ can be found for which

$$f = \sum_{\psi\in\Psi} g\circ\psi.$$

Moreover, each function $\psi\in\Psi$ is of the form

$$\psi(x) := \sum_{i=1}^{d} \lambda(i)\varphi(x(i)),$$

with each of the $2d+1$ functions $\varphi\in\Phi$ strictly increasing and in $\mathrm{Lip}_\alpha[0\,..\,1]$ for some positive $\alpha$, and $\lambda$ some $d$-vector with positive entries all bounded by 1.

The work of Kolmogorov and his pupil Arnol'd which culminated in this theorem was motivated by Hilbert's Thirteenth Problem which contained (implicitly) the conjecture that not all continuous functions of three variables could be written as superpositions of continuous functions of two variables. The version quoted here reflects further simplifications, chiefly by Lorentz (1962). For a proof and further discussion, see (Lorentz, 1966;pp.168ff).

Practical use of Kolmogorov's Theorem seems elusive since the 'universal' functions $\varphi\in\Phi$ have a fractal 'derivative' (see Section 8) and $g$ need not be smooth even if $f$ is smooth. But it remains a challenge to develop a practical Approximation Theory which can handle approximating functions of the form (2.1). In any case, it suggests a nontraditional form of approximation which is motivated by computational or algorithmic simplicity. Perhaps we have been too accepting of traditional approximation techniques in which we choose the approximating family according to linear degrees of freedom. Perhaps we should consider instead approximating families which are classified by the number of floating-point operations required for their evaluation. Approximation Theory as it now exists has little to offer in this direction, but Computer Science may have something to teach us.

A special case has had much exposure in the times before electronic computers, viz., the approximation by **nomographic functions**. These are functions of two variables of the specific form

$$\mathbb{R}^2 \longrightarrow \mathbb{R} : x \mapsto g\big(\varphi(x(1)) + \psi(x(2))\big).$$

For recent algorithmic work, see von Golitschek (1984).

## 3. Loss of Haar

The traditional approaches to approximation all start with polynomials, and so will I. Perhaps the greatest change when going to the multivariate set-up is the loss of the Haar property.

To recall, interpolation from a linear space $S$ of functions on some $G \subseteq \mathbb{R}^d$ at a point set $T \subset G$ can be viewed as the task of inverting the restriction map

$$S \longrightarrow \mathrm{I\!R}^T : p \mapsto p_{|T}.$$

We are to (re)construct some element $p \in S$ from its prescribed values $p_{|T}$ at the points in $T$. In these terms, $S$ has the Haar property if every $T \subseteq G$ with $\#T = \dim S$ is **correct** for $S$, i.e., is such that

$$S \longrightarrow \mathrm{I\!R}^T : p \mapsto p_{|T} \text{ is 1-1 and onto.}$$

In other words, we can interpolate, and uniquely so, from $S$ to any function values given on $T$.

If $d = 1$, then $\Pi_{\le k}$ has the Haar property (for any $G$ with more than $k$ points), but Mairhuber's switching yard argument (cf., e.g., the cover of Lorentz (1966)) shows that this property cannot hold for any $S$ of dimension $> 1$ on a multidimensional set $G$.

For the case of polynomials, it is possible to identify various point sets $T \subset \mathbb{R}^d$ which are correct for $\Pi_{\le k}$. A particularly nice example is provided by Chung and Yao (1977) who prove the following: Suppose that $V$ is a finite subset of $\mathbb{R}^d \backslash 0$, and that $V \cup 0$ is in **general position**, which means that $\Pi_1$ has the Haar property on $V \cup 0$. This implies that for every $W \subset V$ with $\#W = d$, $x_W \in \mathbb{R}^d$ is defined uniquely by the equations

$$1 + \langle w, x_W \rangle = 0, \qquad w \in W,$$

since these state that the linear polynomial $1 + \langle \cdot, x_W \rangle$ is to vanish on $W$ and take the value 1 at 0. Further, since $1 + \langle \cdot, x_W \rangle$ already vanishes on the $d$ points in $W$, it cannot vanish anywhere on $V \backslash W$, by the Haar property. It follows that the functions

$$\ell_W : x \mapsto \prod_{v \in V \backslash W} \frac{1 + \langle v, x \rangle}{1 + \langle v, x_W \rangle}$$

are well-defined, are made up of $\#V - d$ linear factors, and satisfy $\ell_W(x_{W'}) = \delta_{WW'}$ (since, for $W' \neq W$, at least one $v \in V \backslash W$ is in $W'$, making the corresponding linear factor zero at $x_{W'}$), hence in particular $x_{W'} \neq x_W$ for $W' \neq W$, while $\#T = \binom{\#V}{d} = \dim \Pi_{\#V-d}(\mathbb{R}^d)$. This proves

**Theorem (Chung and Yao, 1977)** $T := \{x_W\}$ is correct for $\Pi_{\#V-d}$.

The following result has a different flavor:

**Theorem (Hakopian, 1983)** If $T \subset \mathbb{Z}_+^d$ 'contains its shadow', i.e., $\beta \leq \alpha \in T$ implies $\beta \in T$, then $T$ is correct for $\operatorname{span}\left([\![\,]\!]^\alpha\right)_{\alpha \in T}$.

A totally different approach to the correctness problem has been taken by Kergin (1978). Kergin is interested in extending to a multivariate setting H. Whitney's (1957) characterization of functions on some subset $T$ of $\mathbb{R}$ which have extensions to a smooth function on all of $\mathbb{R}$. Since Whitney uses divided differences in an essential way, Kergin looks for a viable generalization of the divided differences. His approach retains the univariate choice of interpolating from $\Pi_{\leq k}$ at an arbitrary $(k+1)$-set $T$ in $\mathbb{R}^d$ and deals with the many more degrees of freedom available from $\Pi_{\leq k}$ by enforcing certain mean-value conditions. These conditions are that, for every sufficiently smooth $f$, the interpolant $Pf$ should be in $\Pi_{\leq k}$ and, for every $r \leq k$, for every homogeneous polynomial $q$ of degree $r$, and for every $(r+1)$-subset $W$ of $T$, there should exist a point in the convex hull of $W$ at which $q(D)f$ and $q(D)Pf$ agree. Here (to be more explicit) $q(D)$ is an $r$-th order homogeneous constant coefficient differential operator, i.e.,

$$q(D) = \sum_{|\alpha|=r} a(\alpha) D^\alpha$$

for certain coefficients $a(\alpha)$. Surprisingly, there exists exactly one **linear** map $P$ with these properties. This linear map can be characterized by the fact that, for every **plane wave**, i.e., every function $f$ of the form $f := g \circ \lambda : x \mapsto g(\langle x, \lambda \rangle)$, $P$ reduces to univariate interpolation at the projected point set $\langle T, \lambda\rangle = \{\langle t, \lambda\rangle : t \in T\} \subset \mathrm{I\!R}$, i.e., $P(g \circ \lambda) = \big(P_{\lambda T} g\big) \circ \lambda$, with $P_{\lambda T}$ univariate interpolation from $\Pi_{\le k}(\mathrm{I\!R})$ at $\lambda T = \langle T, \lambda\rangle$. In other words, for a plane wave $f$, $Pf(x)$ is the value at $\langle x, \lambda\rangle \in \mathrm{I\!R}$ of the univariate polynomial of degree at most $k$ which matches the value $f(t)$ at $\langle t, \lambda\rangle$, all $t \in T$, where Hermite interpolation is used in case of coincident points. Full understanding of this process (see Micchelli, 1980) led to an understanding of multivariate B-splines, of which more anon.

It seems more promising to give up on polynomials altogether and to choose the interpolating function space $S$ to depend on the point set $T$ at which data are given. The simplest general model has the form

$$\sum_{t \in T} \varphi(\cdot - t) c(t),$$

with $\varphi : \mathbb{R}^d \to \mathrm{I\!R}$ a function to be chosen 'suitably'. Duchon's **thin plate splines** (see, e.g., Meinguet, 1979) use

$$\varphi(x) := |x|^{2m-d} \begin{cases} ln|x|, & n \text{ even;} \\ 1, & n \text{ odd,} \end{cases}$$

motivated by a variational argument, while Hardy's **multiquadrics** correspond to the choice

$$\varphi(x) := \sqrt{1 + |x|^2}.$$

A good source of up-to-date information about such interpolation methods and, in particular, about the question of their correctness, is the recent survey article of Micchelli (1986).

## 4. The B-form

I now come to a discussion of **piecewise polynomial** functions, or **pp** functions for short. I have learned from the people in Computer-Aided Geometric Design that, in dealing with smooth pp functions on some triangulation, it is usually advantageous to write the polynomial pieces in **barycentric-Bernstein-Bézier form**, or **B-form** for short. This form relates polynomials to a given simplex. It is hard to appreciate the power and beauty of this form because, even with carefully chosen notation, it looks forbidding at first sight. Still, I want to point out its structure at least.

One starts with a $(d+1)$-subset $V$ of $\mathbb{R}^d$ in general position and considers the barycentric coordinates with respect to it, i.e., the Lagrange polynomials for linear interpolation at $V$. The typical Lagrange polynomial $\xi_v$ takes the value 1 at the vertex $v$ and vanishes on the facet spanned by $V \backslash v$. The B-form for $p \in \Pi_{\le k}$ employs all possible products of $k$ of these linear polynomials. Explicitly,

$$p =: \sum_{|\alpha|=k} B_\alpha \; c(\alpha) \tag{4.1}$$

with

$$B_\alpha(x) := |\alpha|! [\![\xi(x)]\!]^\alpha = |\alpha|! \prod_{v\in V} [\![\xi_v(x)]\!]^{\alpha(v)}. \tag{4.2}$$

Here

$$\xi(x) := \Big(\xi_v(x)\Big)_{v\in V} \tag{4.3}$$

is the $(d+1)$-vector containing the barycentric coordinates of $x$ with respect to $V$.

Note that the vector $\xi(x)$ and the multi-index $\alpha$ appearing here are conveniently and appropriately indexed by the elements of $V$ (rather than by the numbers $1, 2, \ldots, d+1$ or the numbers $0, 1, \ldots, d$, which would require an arbitrary indexing of the points in $V$).

The factor $|\alpha|!$ in the definition of the **Bernstein basis** element $B_\alpha$ is just right to make $\big(B_\alpha\big)_{|\alpha|=k}$ a **partition of unity**. Indeed,

$$\sum_{|\alpha|=k} B_\alpha(x) = k! \sum_{|\alpha|=k} \prod_{v\in V} [\![\xi_v(x)]\!]^{\alpha(v)} = k! [\![\sum_{v\in V} \xi_v(x)]\!]^k = 1,$$

using the Multinomial Theorem (see (1.1)) and the fact that $\sum_v \xi_v = 1$. The numerical analyst will delight in the alternative formulation of the form,

$$p(x) = \langle \xi(x), E\rangle^k c(0) \tag{4.4}$$

which makes use of the **shift operator** $E$ given by the rule

$$E^\beta c(\alpha) = c(\alpha+\beta).$$

More explicitly,

$$\langle z, E\rangle = \sum_{v\in V} z(v)E_v,$$

i.e.,

$$\langle z, E\rangle c(\alpha) = \sum_{v\in V} z(v)c(\alpha + e_v)$$

with $e_v$ the $v$-unitvector, i.e, $e_v(w) = \delta_{vw}, w \in V$. This form provides a most convenient starting point for the derivation of efficient algorithms for the evaluation and differentiation of the B-form. For details, see, e.g., Farin (1985) and de Boor (1986).

## 5. Smooth pp functions

The B-form is well suited to pp work since its typical term $B_\alpha$ vanishes $\alpha(v)$-fold on the facet spanned by $V\backslash v$. This means that the form readily provides information about the behavior of $p$ at all the bounding faces of the simplex with vertex set $V$. This is being increasingly exploited in studying the algebraic structure of the space

$$\Pi^\rho_{k,\Delta}$$

of pp functions of degree $\le k$ on a given triangulation $\Delta$ whose pieces join together smoothly to provide a function all of whose derivatives of order $\le \rho$ are continuous.

The problems being studied include: the dimension of such a space, a good basis for such a space, and the approximation power of such a space. For recent results, see Chui and his co-workers, and Schumaker. These results only deal with $d = 2$, and, even for this case, we know relatively little. For example, despite considerable efforts, we still do not know the dimension of the space of continuously differentiable piecewise cubic functions on an arbitrary triangulation in the plane. While we do know that this dimension depends on the quantitative details of the triangulation, we do not know exactly how.

As we understand these problems better and see some of their particular difficulties, we wonder whether $\Pi^\rho_{k,\Delta}$ is really the right space to study. It now seems that it might be more appropriate to seek out appropriate subspaces, e.g., the subspace spanned by certain compactly supported smooth piecewise polynomials as was done already in **Finite**

**Elements**. A particularly simple model is provided by approximation from a **scale** of pp functions.

## 6. Approximation from a scale

Associate with a given function space $S$ the **scale** $\big(S_h\big)$, with

$$S_h := \sigma_h S, \qquad \big(\sigma f\big)(x) := f(x/h),$$

and define the **approximation order** of $S$ to be

$$\max\,\{r : \forall \text{ smooth } f \quad \operatorname{dist}(f, S_h) = O(h^r)\}.$$

This order may well be 0, as it is for $S = \Pi_{\le k}$. But if $S$ contains functions whose support has diameter $\delta$, then $S_h$ contains functions with supports of diameter $h\delta$, and, for such $S$, one might hope to obtain closer approximations from $S_h$ as $h \to 0$. Work with specific examples has suggested the following conjectures in case $S \subseteq \Pi_{k,\Delta}$:

**Conjectures:** (i) The approximation order of $S$ equals the approximation order of

$$S_{\text{loc}} := \operatorname{span}\,\{\varphi \in S : \operatorname{supp}\varphi \text{ compact}\}.$$

(ii) $S$ has approximation order $\ge 1$ iff $S$ contains a local partition of unity.

(iii) The approximation order is always realized by a good quasi-interpolant.

Here, a map $Q$ into $S$ is a **good quasi-interpolant of order** $r$ in case it is a linear map which is stable in the sense that, for any $f$ and any $x \in G$,

$$|(Qf)(x)| \le \mathit{const}\sup\{|f(y)| : \|y - x\| \le R\}$$

with *const* and $R < \infty$ independent of $f$ or $x$, and which reproduces polynomials of degree $< r$. For example, if $\Phi$ is a local and nonnegative partition of unity in $S$, i.e., $\sup_{\varphi\in\Phi} \operatorname{diam}\operatorname{supp}\varphi < \infty$, $\varphi \ge 0$ for all $\varphi \in \Phi$, and $\sum_{\varphi\in\Phi}\varphi = 1$, then

$$f \mapsto \sum_{\varphi\in\Phi} \varphi\; f(\tau_\varphi)$$

is a good quasi-interpolant of order 1 (provided, e.g., that $\tau_\varphi \in \text{supp}\varphi$ for all $\varphi \in \Phi$).

This abstract model can be completely analysed in the very special case when $S$ is spanned by the integer translates of one function $\varphi$, i.e.,

$$S := S_\varphi := \text{ span}\Big(\varphi(\cdot - j)\Big)_{j\in\mathbb{Z}^d} = \{\sum_{j\in\mathbb{Z}^d} \varphi(\cdot - j)c(j) : c(j) \in \mathbb{R}\}.$$

For this case, the three conjectures are verified; in particular, Strang and Fix (1973) prove that $S$ **has approximation order** $r$ **iff** $\Pi_{<r} \subseteq S$.

Already for the slightly more general case when $S$ is the span of integer translates of *several* compactly supported functions, the situation becomes more complicated. A characterization of the approximation order is not yet known for this case, but the somewhat stronger (and practically more interesting) concept of **local** approximation order can be characterized very simply (de Boor and Jia, 1985): $S$ **has local approximation order** $r$ **iff there exists** $\psi \in S_{\text{loc}}$ **such that** $S_\psi$ **has approximation order** $r$.

For the general case, even simple questions such as whether a pp space with positive approximation order must contain a compactly supported element have so far remained unanswered.

## 7. Multivariate B-splines

The abstract theory of approximation from a scale has found new interest recently because of the advent of multivariate B-splines. These were introduced in 1976 in hopes that they would perform the same service in the study of multivariate smooth pp functions that the B-splines of Schoenberg and Curry provided so nicely for the theory of (univariate) splines.

In retrospect, well, in any case, the idea is simple enough. It involves a body $B \subset \mathbb{R}^n$ and the orthogonal projector $P : \mathbb{R}^n \to \mathbb{R}^d : x \mapsto (x(1), \ldots, x(d))$. The map $P$ is used to extend a function $\varphi$ on $\mathbb{R}^d$ to the function

$$\varphi \circ P : u \mapsto \varphi(Pu)$$

on all of $\mathrm{I\!R}^n$. The B-spline $M_B$ is defined as the distribution on $\mathbb{R}^d$ which represents integration over $B$ of the extended function. In formulae:

$$M_B : \varphi \mapsto \int_B \varphi \circ P \qquad \text{for all } \varphi \in C_0. \tag{7.1}$$

Here, $C_0 = C_0(\mathbb{R}^d)$ is the collection of all continuous functions on $\mathbb{R}^d$ with compact support. If $PB$ (the projection of the body) is $d$-dimensional, this can also be written

$$M_B(\varphi) = \int_{\mathbb{R}^d} M_B\varphi = \int_{PB} \mathrm{d}y\phi(y) \int_{B\cap P^{-1}y} 1 \tag{7.2}$$

showing that $M_B(y) = \mathrm{vol}_{n-d} B \cap P^{-1}y$. This latter formula was the original definition, motivated by a geometric characterization of the (univariate) B-spline due to Curry and Schoenberg (1966) and illustrated, for $n = 3, d = 1$, in Figure 7.1.

Figure 7.1 The quadratic B-spline as the "shadow" of a 3-simplex.

The value of the B-spline at a point $y$ equals the $(n-d)$-dimensional volume of the intersection of the simplex with the hyperplane $P^{-1}y$.

It is immediate that $M_B$ has **compact support**. Further, if $B$ is polyhedral with facets $\{B_i\}$, and if $z \in \mathrm{I\!R}^n$, then an application of Stokes' Formula shows that the directional derivative of $M_B$ along $Pz$ is

$$D_{Pz}M_B = -\sum_i \langle z, n_i\rangle M_{B_i}. \tag{8.1}$$

with $n_i$ the outward unit normal to the facet $B_i$ and $M_{B_i}$ the B-spline that is the "shadow" of the "body" $B_i$. Repeated applications of this differentiation formula show that all derivatives of $M_B$ of order $n-d+1$ must vanish identically away from the projections of the $(d-1)$-dimensional faces of $B$. Consequently,

$$M_B \in \Pi^{\rho}_{n-d,\Delta},$$

with $\Delta$ the partition whose partition interfaces are the projections of $(d-1)$-dimensional faces of $B$, and where $\rho$ is defined by the condition that $n-\rho-2$ equal the largest dimension of a face of $B$ projected entirely into one of the partition interfaces. Thus, in the generic case, we have $\rho = n-d-1$, which is as large as it can possibly be, given that the polynomial degree of $M_B$ is $n-d$.

This surprising smoothness is bought at a price. Since, for a *generic* partition $\Delta$, $\Pi^{k-1}_{k,\Delta}$ does not contain any locally supported functions, the partition for $M_B$ must be quite special. Figure 7.2 shows such a partition for a bivariate quadratic **simplex** spline, i.e., a B-spline that is the "shadow" of a simplex.

Figure 7.2 The partition for a bivariate quadratic simplex spline

Thus we cannot expect to obtain B-splines for every partition. At best, we can find B-splines whose partition refines a given one. For the case of simplex splines, such a collection of B-splines of degree $k$ can be constructed rich enough to provide a good quasi-interpolant of order $k+1$. There are even stable recurrence relations, found by Micchelli (1980), for their evaluation. But it seems that their use is computationally quite expensive (see, e.g., Grandine, 1986). It is therefore not likely that simplex splines will be used as a basis for a good subspace of a given smooth pp space of functions. Most likely, translates of a fixed B-spline will find practical employment.

There is a bit more hope for the **box** splines, i.e., the multivariate B-splines associated with the $n$-cube $B = [0\,..\,1]^n$. For their definition in terms of (7.1), one would allow $P$ to be, more generally, a linear map. Then, with $\xi_i := Pe_i$ the image of the $i$-th unit vector under $P$, the box spline can be characterized more explicitly by

$$\int_{\mathbb{R}^d} M(\cdot|\xi_1,\ldots,\xi_n)\varphi = \int_{[0..1]^n} \varphi\big(\sum_i \xi_i y(i)\big)\,\mathrm{d}y, \quad \varphi \in C_0.$$

By choosing the $\xi_i$ from $\mathbb{Z}^d$ appropriately, the resulting partition can be made to conform to a regular grid; see Figure 7.3 for the supports of two very well known box splines, the Courant element ($d = 2, n = 3$) and the Zwart-Powell element ($d = 2, n = 4$). Further, their evaluation can be accomplished by *subdivision*.

Figure 7.3 The supports of a linear and a quadratic bivariate box spline

De Boor (1982) gives an introduction to multivariate B-splines. Dahmen and Micchelli (1984) provide a survey of the literature available by 1983 to which they heavily contributed. Höllig (1986a) gives a more up-to-date introduction, and Höllig (1986b) summarizes what we know about box splines. In addition, I want to stress the beautiful, but more theoretical, developments to which Dahmen and Micchelli were led by their intensive study of box splines (see, e.g., Dahmen and Micchelli, 1985, 1986).

## 8. Subdivision

I hate to finish on a pessimistic note. I therefore bring up a totally different approach to the generation or approximation of surfaces which comes from Computer-Aided Design. I think that this technique has real promise for the generation of 'smooth' surfaces which fit to given points in 3-space of more or less arbitrary combinatorial structure. It is at present being used to evaluate linear combinations of box splines (see, e.g., Höllig, 1986a, for the relevant references). But since this idea has not yet been thoroughly studied, I will discuss it only in its original context of curve generation.

Figure 8.1 The steps of the simple subdivision algorithm.

Here is a very simple version of subdivision, which generalizes **Chaikin's algorithm** (Chaikin, 1974). Start off with points $a_j$ in $\mathbb{R}^d$, where $j$ runs over all of $\mathbb{Z}$, for simplicity. Think of these points as the vertices of a broken line. From the algorithm, one obtains a refined broken line in two steps. In the first step, one introduces the midpoints between neighboring vertices as new vertices, thus roughly doubling the number of vertices:

$$b_{2j} := a_j, \qquad b_{2j+1} := (a_j + a_{j+1})/2.$$

In the second step, one obtains each vertex of the refined broken line as an average of three neighboring vertices:

$$c_j := \beta b_{j-1} + \alpha b_j + \gamma b_{j+1}, \quad \text{with } \beta + \alpha + \gamma = 1.$$

In fact, for a curve of higher 'smoothness', one would repeat this averaging step one or more times, but I will stick with this simple model. Repetition quickly leads to a broken line which, for plotting purposes, is indistinguishable from the limiting curve.

Of course, it is not at all clear *a priori* that there is a limiting curve, though that is easily proved for reasonable choices of the weights, e..g, for $\alpha,\ \beta,\ \gamma \geq 0$. Nor is it clear just what the nature of that limiting curve might be. Chaikin's algorithm corresponds to the choice $\beta = 0, \alpha = \gamma = 1/2$. For this choice, the limiting curve is a parametric quadratic spline curve, viz. the curve

$$t \mapsto \sum_{\mathbb{Z}} M_3(t-j)a_j$$

with $M_3$ a quadratic cardinal B-spline (i.e., a B-spline having integer knots). For the symmetric choice $\beta = \gamma = (1-\alpha)/2$, the limiting curve is a parametric quadratic, resp. cubic, spline curve in case $\alpha = 0$, resp. $1/2$, but for any other choice, the limiting curve appears to be something unmentionable in standard terms, though this is not apparent from the curves themselves; see, e.g., Figure 8.2.

Figure 8.2 Curve iterates with $\alpha = 1/4, \beta = \gg$

Since the limiting curve depends linearly on the initial data $a_j$, it is sufficient for the complete analysis of the process to consider a broken line with just one actual break, e.g., the data

$$a_j := (j, j_+), \quad j \in \mathbb{Z} \tag{8.1}$$

in $\mathbb{R}^2$. The various curves generated will differ only on the segment between the two points $(-1,0)$ and $(1,1)$. Some of the iterates as well as the limiting curve segment are shown in Figure 8.3 for the particular choice $\alpha = 1/4$, and we see nothing unusual.

Figure 8.3 Iterates and limiting curve for one-break broken line

It is only when we look at derivatives that we realize that something is amiss. For our particular curve, the first component is just $t \mapsto t$, so it makes sense to consider the derivatives of the second component, $t \mapsto y(t)$ say, as an indication of the smoothness of the curve. Since we do not have a formula for $y(t)$, we cannot compute derivatives. But since the iteration is so simple, we can compute $y(t)$ for as many binary fractions $t$ as we care to. That done, we can then compute second divided differences, and these tell the story; see Figure 8.4: The second derivative of $y$ appears to be a **fractal**.

Traditional Approximation Theory views such curves with horror. There is even the seemingly practical objection that it would be difficult to machine curves with such a 'bad' second derivative. But I do not know of experiments which have established such a difficulty nor is it clear *a priori* that there should be any difficulty. On the other hand, the generation of such curves to plotting or machining accuracy is so swift, and their flexibility so great, that this technique is worth exploring in detail. It is this apparent local flexibility that makes subdivision techniques a promising tool for the generation of shape-controlled surfaces of arbitrary combinatorial structure.

Figure 8.4 **Wisconsin Winter**

Second divided differences $[t-h, t, t+h]y$ plotted against $t$, for $h = 2^{-8}$ (dots), and $h = 2^{-9}$ (stars)

For literature, see Catmull and Clark (1978), Doo (1978), and Doo And Sabin (1978). I am indebted to Professor Wanner for the surprising references to work by de Rham (1947, 1953, 1956, 1957, 1959) who, many years ago and, apparently, for pedagogical reasons, investigated a related subdivision algorithm for curves. Professor Wanner referred to it imaginatively as the '**the woodcarver's algorithm**'.

## References

C. de Boor (1982), Topics in Multivariate Approximation Theory, in *Topics in Numerical Analysis*, P. Turner ed., Springer Lecture Notes in Mathematics **965**, 39–78.

C. de Boor (1986), B-form basics, in *Geometric Modeling*, G. Farin ed., SIAM, Philadelphia PA, 131–148.

C. de Boor ed. (1986), *Approximation Theory*, Amer. Mathem. Soc., Providence RI.

C. de Boor and Jia R.-q. (1985), Controlled approximation and a characterization of the local approximation order, *Proc. Amer. Math. Society* **95**, 547–553.

E. E. Catmull and J. H. Clark (1978), Recursively Generated B-Spline Surfaces on Arbitrary Topological Meshes, *Computer Aided Design* **10**, 350–353.

G. M. Chaikin (1974), An Algorithm for High Speed Curve Generation, *Computer Graphics and Image Processing* **3**, 346–349.

K. C. Chung and T. H. Yao (1977), On lattices admitting unique Lagrange interpolations, *SIAM J. Numer. Anal.* **14**, 735–741.

H. B. Curry and I. J. Schoenberg (1966), Pólya frequency functions.IV, The fundamental spline functions and their limits, *J. d'Anal. Math.* **17**, 71-107.

W. Dahmen and C. A. Micchelli (1984), Recent Progress in Multivariate Splines, in *Approximation Theory IV*, C. K. Chui, L L. Schumaker and J. Ward., eds., Academic Press, New York, 27–121.

W. Dahmen and C. A. Micchelli (1985), On the solution of certain systems of partial difference equations and linear independence of translates of box splines, *Trans. Amer. Math. Society* **292**, 305–320.

W. Dahmen and C. A. Micchellii (1986), On the number of solutions to linear diophantine equations and multivariate splines (1986), RC 11725, IBM.

D. W. H. Doo (1978), A Subdivision Algorithm for Smoothing Down Irregularly Shaped Polyhedrons, *Proceedings: Interactive Techniques in Computer Aided Design*, Bologna, 157–165.

D. Doo and M. A. Sabin (1978), Behavior of Recursive Division Surfaces Near Extraordinary Points, *Computer Aided Design* **6**, 356–360.

G. Farin, Triangular Bernstein-Bézier patches, *CAGD Report 11-1-85*, Department of Mathematics, University of Utah, Salt Lake City UT.

M. von Golitschek (1984), Shortest path algorithms for the approximation by nomographic functions, in *Approximation Theory and Functional Analysis*, P. L. Butzer, R. L. Stens and B. Sz.-Nagy eds, Birkhäuser Verlag, Basel, ISNM Vol. 65, xxx–xxx.

M. Golomb (1959), Approximation by functions of fewer variables, in *On numerical approximation*, R. E. Langer ed., The University of Wisconsin Press, Madison WI, 275–327.

T. Grandine, The computational cost of simplex spline functions (1986), *MRC TSR #2926*, U. of Wisconsin, Madison WI.

H. Hakopian (1983), Integral remainder formula of the tensor product interpolation, *Bull. Polish Acad. Sci.* **31**, 267–272.

K. Höllig (1986a), Box splines, in *Approximation Theory V*, C. K. Chui, L. L. Schumaker and J. Ward eds., Academic Press, 71–95.

K. Höllig (1986b), Multivariate splines, in *Approximation Theory*, C. de Boor ed., *Proc. Symp. Appl. Math. 36*, Amer. Math. Soc. , Providence RI, 103–127.

P. Kergin (1978), *Interpolation of $C^k$ functions*, Ph.D. Thesis, University of Toronto, Canada; published as 'A natural interpolation of $C^k$ functions' in *J. Approximation Theory* **29** (1980), 278–293.

A. N. Kolmogorov (1957), On the representation of continuous functions of several variables by superpositions of continuous functions of one variable and addition, *Doklady* **114**, 679–681.

W. A. Light and W. E. Cheney (1986), *Approximation Theory in Tensor Product Spaces*, Lecture Notes in Mathematics 1169, Springer-Verlag.

G. G. Lorentz (1962), Metric entropy, widths, and superposition of functions, *American Math. Monthly* **69**, 469–485.

G. G. Lorentz (1966), *Approximation of Functions*, Holt, Rinehart and Winston, New York.

J. Meinguet (1979), Multivariate interpolation at arbitrary points made simple, *J. Appl. Math. Phys. (ZAMP)* **30**, 292–304.

C. A. Micchelli (1980), A constructive approach to Kergin interpolation in $\mathbb{R}^k$: multivariate B-splines and Lagrange interpolation (1980), *Rocky Mountains J. Math.* **10**, 485–497.

C. A. Micchelli (1986), Algebraic aspects of interpolation, in *Approximation Theory*, C. de Boor ed., *Proc. Symp. Appl. Math. 36*, Amer. Math. Soc. , Providence RI, 81–102.

G. de Rham (1947), Un peu de mathematique à propos d'une courbe plane, *Elemente der Mathematik* **II**, 73–76; 89–97. (*Collected Works*, 678–689)

G. de Rham (1953), Sur certaines équations fonctionnelles, *L'ouvrage publié à l'occasion de son centenaire par l'ecole polytechnique de l'université de Lausanne 1853-1953*, 95–97. (*Collected Works*, 690–695)

G. de Rham (1956), Sur une courbe plane, *J. Mathem. Pures Appl.* **335**, 24–42. (*Collected Works*, 696–713)

G. de Rham (1957), Sur quelques courbes définies par des équations fonctionnelles, *Rendiconti del seminario matematica dell'universitate del Politecnico di Torino* **16**, 101–113. (*Collected Works*, 714–727)

G. de Rham (1959), Sur les courbes limites de polygones obtenus par trisection, *l'Enseignement Math.* **5**, 29–43.(*Collected Works*, 728–743)

G. Strang and G. Fix (1973), A Fourier analysis of the finite element variational method, in *Constructive Aspects of Functional Analysis* G. Geymonat ed., C.I.M.E., 793–840.

H. Whitney (1957), On functions with bounded $n$th differences, *J. Math. Pures Appl.* **36**, 67–95.