Computer Aided Geometric Design 4 (1987) 257–268
North-Holland

A 4-point interpolatory subdivision scheme for curve design

Nira DYN, David LEVIN

School of Mathematical Sciences, Tel-Aviv University, Tel-Aviv 69978, Israel

John A. GREGORY

Department of Mathematics, Brunel University, Uxbridge, Middlesex UB8 3PH, UK

Received October 1986
Revised September 1987

Abstract. A 4-point interpolatory subdivision scheme with a tension parameter is analysed. It is shown that for a certain range of the tension parameter the resulting curve is C^1. The role of the tension parameter is demonstrated by a few examples. The application to surfaces and some further potential generalizations are discussed.

Keywords. Subdivision, interpolation, curves, surfaces, point insertion.

1. Introduction

Subdivision schemes are efficient tools for computer-aided curve and surface design. Most of the important methods for curve design can be approached by subdivision processes [Boehm, Farin, Kahmann '84]. The common subdivision schemes are based on chopping corners of a control polygon, hence they are not interpolatory, and this is a drawback in some applications.

Recently a simple interpolation scheme based on a 4-point recursive subdivision was suggested by the second author. Given control points $\{\bar{p}_i\}_{i=-2}^{n+2}$, $\bar{p}_i \in \mathbb{R}^d$, intermediate points are added by the scheme

$$\bar{p}_{i+1/2} = \left(\tfrac{1}{2} + w\right)\left(\bar{p}_i + \bar{p}_{i+1}\right) - w\left(\bar{p}_{i-1} + \bar{p}_{i+2}\right), \quad -1 \le i \le n. \tag{1.1}$$

The augmented set of points $\{\bar{p}_{i/2}\}_{i=-2}^{2n+2}$ is then regarded as new control points, and the same rule is used to insert points between consecutive points, and so on repeatedly.

At each stage four consecutive points in the last set define a new intermediate point. Let us denote the control points at the k-level set by $\{\bar{p}_i^k\}_{i=-2}^{2^k n+2}$. Then the subdivision scheme defines the control points at level $k+1$ by

$$\begin{cases} \bar{p}_{2i}^{k+1} = \bar{p}_i^k, & -1 \le i \le 2^k n + 1, \\ \bar{p}_{2i+1}^{k+1} = \left(\tfrac{1}{2} + w\right)\left(\bar{p}_i^k + \bar{p}_{i+1}^k\right) - w\left(\bar{p}_{i-1}^k + \bar{p}_{i+2}^k\right), & -1 \le i \le 2^k n \end{cases} \tag{1.2}$$

where $\bar{p}_i^0 = \bar{p}_i$, $-2 \le i \le n+2$.

By letting k tend to infinity, this process defines an infinite set of points in $\mathbb{R}^d$. The questions we are confronting are whether these points lie on a continuous curve in $\mathbb{R}^d$ and what are the smoothness properties of this curve, depending on the values of the parameter w.

A nice geometric interpretation of the scheme (1.2) is revealed by writing (1.1) in the form

$$\bar{p}_{i+1/2} = \tfrac{1}{2}\left(\bar{p}_i + \bar{p}_{i+1}\right) + 2w\left[\tfrac{1}{2}\left(\bar{p}_i + \bar{p}_{i+1}\right) - \tfrac{1}{2}\left(\bar{p}_{i-1} + \bar{p}_{i+2}\right)\right]. \tag{1.3}$$

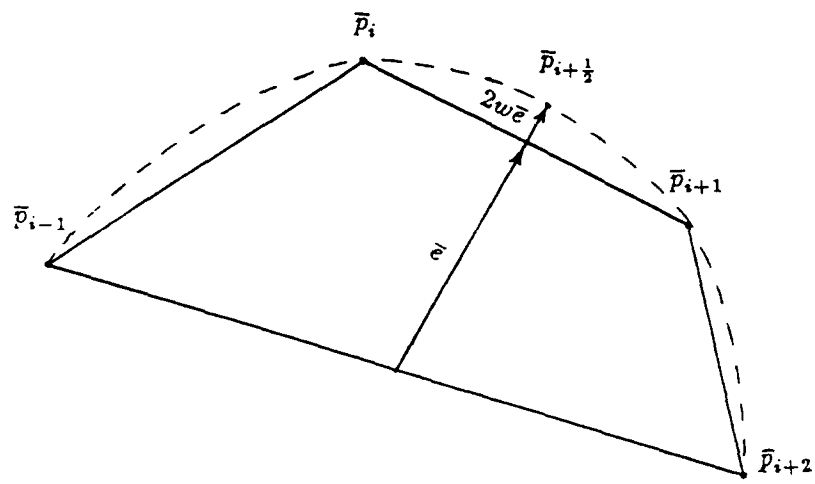

Fig. 1. Geometric interpretation of the scheme (1.2).

For $w = 0$ the point $\bar{p}_{i+1/2}$ is simply the mid-point of the linear segment $(\bar{p}_i, \bar{p}_{i+1})$. For $w \neq 0$ this mid-point is 'corrected' by a vector $2w\bar{e}$ where $\bar{e}$ is the vector from the mid-point of $(\bar{p}_{i-1}, \bar{p}_{i+2})$ to the mid-point of $(\bar{p}_i, \bar{p}_{i+1})$ (see Fig. 1). It is clear that if the initial control points are on a linear manifold $M \subset \mathbb{R}^d$, then all subsequent points also remain on M.

The above geometrical interpretation of the scheme clarifies the role of w as a tension parameter; decreasing the value of w to zero is equivalent to tightening the corresponding curve towards the control polygon – the piecewise linear curve between the control points.

It has been observed experimentally that the scheme (1.2) produces a continuous curve with continuous tangent vector for $0 < w < \frac{1}{4}$. In Sections 3, 4, we analyze the limit curve and prove that it is continuous for $|w| < \frac{1}{4}$ and has a continuous tangent vector for $0 < w < \frac{1}{8}$. Also, we show that for any $w > 0$ the components of the curve determined by a general set of control points do not have a second derivative in the parametrization $\bar{p}(t)$, $t \in [0, n]$, which attaches $\bar{p}_i^k$ to the parameter value $2^{-k}i$, $k \geq 0$, $0 \leq i \leq 2^k n$.

A scheme of the type (1,1) which can produce continuous curves with continuous tangent and curvature is of the form

$$\bar{p}_{i+1/2} = \left(\tfrac{9}{16} + 2\theta\right)(\bar{p}_i + \bar{p}_{i+1}) - \left(\tfrac{1}{16} + 3\theta\right)(\bar{p}_{i-1} + \bar{p}_{i+2}) + \theta(\bar{p}_{i-2} + \bar{p}_{i+3}). \tag{1.4}$$

For $\theta = 0$ this scheme corresponds to (1.1) with $w = \frac{1}{16}$. For any θ the scheme (1.4) satisfies the necessary condition derived in [Dyn, Levin '86] for a general $2r$-point scheme to produce curves with C^2 components. The values of θ which guarantee the continuity of the curvature of the curve are

$$0 < \theta < 0.02. \tag{1.5}$$

The proof of this result along the lines of the analysis in Section 3 can be found in [Weissmann '88].

In [Dyn, Levin '86] necessary and sufficient conditions for the $2r$-point scheme to produce C^r curves were stated. At this stage the two authors retreat from the sufficiency part of the claim due to a flaw in the proof.

In the next section the application of the scheme (1.2) to the design of curves is discussed and three examples are presented. The tensor-product extension of the scheme for the design of surfaces is briefly viewed.

The subdivision scheme can also be used to approximate a function $f(t)$, given its value on a set of equidistant points

$$f_i^0 = f(ih), \quad -2 \le i \le n+2, \quad nh = 1, \tag{1.6}$$

by taking $d = 2$ and $\bar{p}_i^0 = (ih, f_i^0)$, $-2 \le i \le n+2$, in (1.2). For this type of interpolation the question of approximation order is of interest. It is shown in Section 5 that for $f \in C^4[-\epsilon, 1+\epsilon]$ and $w = \frac{1}{16}$, the function defined by the scheme with the data (1.6) approximates $f(t)$ with an error of order $O(h^4)$ in [0, 1].

2. Applications and examples

The subdivision algorithm (1.2) is actually an insertion algorithm since all the points at stage k of the algorithm are carried over to stage $k+1$ and new points are inserted in between the old ones. Evidently the resulting limit curve interploates the initial control points. Another important property of the scheme is that, for any w, it produces a straight line in $\mathbb{R}^d$, whenever the control points lie on such a line. The local nature of the scheme (i.e., the curve segment $\bar{p}(t)$, $t \in (i, i+1)$ depends on $\bar{p}_{i-2}, \ldots, \bar{p}_{i+3}$), and the control of the tension by the parameter w are also important features for curve design. The algorithm discussed here is unique in combining the four ingredients: subdivision, locality, interpolation, shape control. As any other subdivision algorithm the above insertion algorithm is very fast and convenient for computing all the points required for the graphical display of the curve.

The evaluation of a point on the curve at a prescribed parameter value is also quite fast, due to the local nature of the scheme. For a parameter value $\hat{t} \in (j, j+1)$, let the diadic expansion of the fractional part of $\hat{t}$ be given by

$$\hat{t} - j = \sum_{m=1}^{N} \tau_m 2^{-m}, \quad \tau_m \in \{0, 1\}. \tag{2.1}$$

We start with the six initial points $\{\bar{p}_i\}_{i=j-2}^{j+3}$ which determine all the curve segment $\bar{p}(t)$, $t \in [j, j+1]$. Then at stage k we compute three new points and carry six points to the next stage. The six points that should be known for stage $k+1$ correspond to the parameter values $t^{(k)} + l \cdot 2^{-k}$, $-2 \le l \le 3$, where $t^{(k)} = j + \sum_{m=1}^{k} \tau_m 2^{-m}$. These points determine $\bar{p}(t)$ for $t \in (t^{(k)}, t^{(k)} + 2^{-k})$, and in particular $\bar{p}(\hat{t})$.

The tangent of the curve at the point $\bar{p}(\hat{t})$ for $\hat{t}$ given by (2.1) can be evaluated by the same process. Using Corollary 4.1, we conclude that

$$\bar{p}'(\hat{t}) = \frac{2^N}{1-4w} \left\{ \tfrac{1}{2}\left[\bar{p}(\hat{t} + 2^{-N}) - \bar{p}(\hat{t} - 2^{-N})\right] - w\left[\bar{p}(\hat{t} + 2^{-N+1}) - \bar{p}(\hat{t} - 2^{-N+1})\right] \right\}. \tag{2.2}$$

Thus the slope at a point which is inserted at stage N is computed from its four neighboring points in the Nth stage, two on each side.

To define a curve passing through $\bar{p}_0, \bar{p}_1, \ldots, \bar{p}_n$ by (1.2), one needs to supply additional points $\bar{p}_{-2}, \bar{p}_{-1}, \bar{p}_{n+1}, \bar{p}_{n+2}$ which affects the behaviour of the curve near its end points. These extra points can be used to control the slope of the curve at the end points. Using (2.2) with $N = 0$ we get the slopes of the curve at the control points

$$\bar{p}'(i) = \frac{1}{1-4w} \left\{ \tfrac{1}{2}(\bar{p}_{i+1} - \bar{p}_{i-1}) - w(\bar{p}_{i+2} - \bar{p}_{i-2}) \right\} \tag{2.3}$$

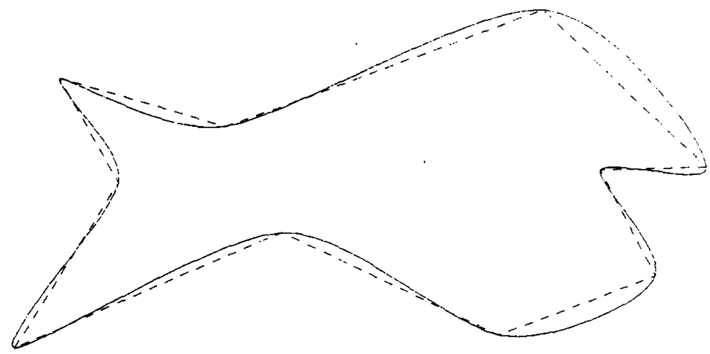

Fig. 2. Control polygon (broken line) and smooth curve (full line).

and in particular at the end points the dependence of $\bar{p}'(0)$ and $\bar{p}'(n)$ on the extra points becomes clear.

In case of a closed curve, additional end points are not needed, since by a cyclic continuation

$$\bar{p}_{-2}=\bar{p}_{n-1}, \quad \bar{p}_{-1}=\bar{p}_{n}, \quad \bar{p}_{n+1}=\bar{p}_{0}, \quad \bar{p}_{n+2}=\bar{p}_{1}. \tag{2.4}$$

The following two examples (Figs. 2 and 3) show closed curves which pass through a set of control points. The control polygons (corresponding to $w=0$) are drawn by a broken line, and the smooth curves obtained by (1.2) with $w=\frac{1}{16}$ by a full line.

Fig. 3 depicts also a dotted curve obtained with $w=\frac{1}{10}>\frac{1}{16}$, thus demonstrating the decrease in the tightness (tension) of the curve with the increase in w $(0 \le w < \frac{1}{8})$.

The values of w which are relevant for application to curve design are $0 \le w \le \frac{1}{10}$. Yet it is interesting to see the nature of a curve which is obtained with a parameter w outside the permissible range. In Fig. 4 the control polygon is as in Fig. 3. The broken line is now the curve with $w=\frac{1}{16}$ and the full line is the polygonal line connecting the points obtained in the fifth stage of the scheme (1.2) with $w=0.3>\frac{1}{4}$. The limiting curve corresponding to this polygonal line seems continuous, but it has many loops and sharp bends and probably is of Hausdorff dimension >1.

We remark that for $w>0$ the proposed scheme is not shape preserving, i.e., monotonicity or convexity of the original control polygon does not guarantee monotonicity or convexity of the corresponding curve. Also, it should be noted that if the distribution of the control points is highly non-uniform, the scheme may introduce kinks and loops in the curve.

The subdivision scheme (1.2) can be easily extended for the design of a surface which passes through a set of control points of the form:

$$\left\{ \bar{p}^0_{i,j};\, i=-2,\ldots,n+2,\ j=-2,\ldots,m+2 \right\} \tag{2.5}$$

First we apply (1.2) to the index i, inserting points between $\bar{p}^k_{i,j}$ and $\bar{p}^k_{i+1,j}$, $i=-1,\ldots,2^k n$,

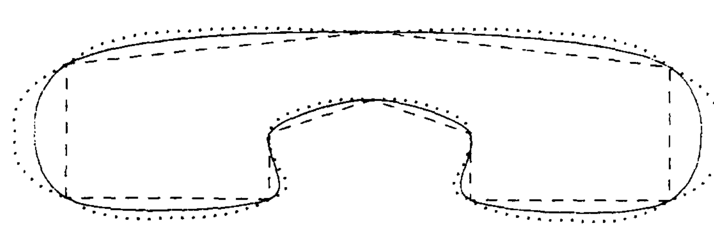

Fig. 3. Decreasing the tightness (tension) of a curve.

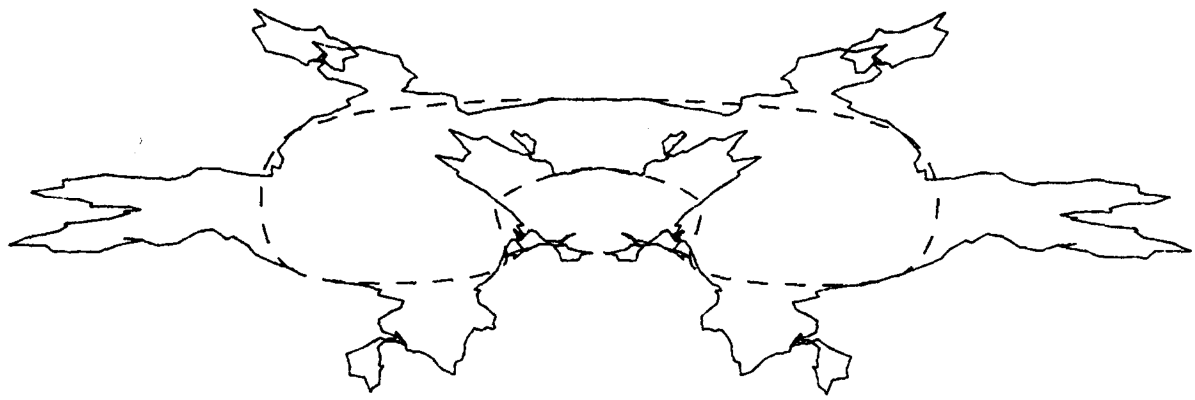

Fig. 4. Parameter w outside the permissible range.

$j = -2,\dots,2^k m+1$. Then (1.2) is applied to the index j. The overall step results in the $(k+1)$th set of points:

$$\begin{cases} \bar{p}^{k+1}_{2i,2j} = \bar{p}^k_{i,j}, & -1 \le i \le 2^k n+1, -1 \le j \le 2^k m+1, \\ \bar{p}^{k+1}_{2i+1,2j} = \left(\tfrac{1}{2}+w\right)\left(\bar{p}^k_{i,j}+\bar{p}^k_{i+1,j}\right) - w\left(\bar{p}^k_{i-1,j}+\bar{p}^k_{i+2,j}\right), & \\ & -1 \le i \le 2^k n, -1 \le j \le 2^k m+1, \\ \bar{p}^{k+1}_{i,2j+1} = \left(\tfrac{1}{2}+w\right)\left(\bar{p}^{k+1}_{i,2j}+\bar{p}^{k+1}_{i,2j+2}\right) - w\left(\bar{p}^{k+1}_{i,2j-2}+\bar{p}^{k+1}_{i,2j+4}\right), & \\ & -1 \le i \le 2^k n+1, -1 \le j \le 2^k m. \end{cases} \tag{2.6}$$

For $|w| < \frac{1}{4}$ the points computed by the above process can be extended into a continuous surface, which for $0 < w < \frac{1}{8}$ has continuous tangent plane. This surface may be viewed as a blending surface of the $m+2$ curves $\bar{p}_j(t)$, $t \in [0, n]$, obtained by the application of (1.2) to $\{\bar{p}_{ij};\ -2 \le i \le n+2\}$ for $j = -2,\dots,m+2$. The blending process uses (1.2) again on the sets of control points $\{\bar{p}_{-2}(t),\dots,\bar{p}_{m+2}(t)\}$, $t \in [0, n]$, to yield the surface $\bar{p}(t, s)$, $t \in [0, n]$, $s \in [0, m]$ which interpolates the curves $\{\bar{p}_j(t)\}_{j=0}^{n}$.

3. Convergence analysis – sufficient conditions

To investigate the continuity and smoothness properties of the limit curve defined by (1.2), we view it as a parametric curve. At stage k of the algorithm the points $\{\bar{p}^k_i\}_{i=0}^{2^k n}$ are interpreted as points on a parametric curve $\bar{p}(t)$, with a parameter range $0 \le t \le n$, corresponding to the parameter values $\{2^{-k} i\}_{i=0}^{2^k n}$. The smoothness properties of the limit curve $\bar{p}(t)$ are at least as strong as those of its components. Therefore, we consider the scheme (1.2) with $\{\bar{p}^k_i\}$ replaced by $\{f^k_i\}$, a sequence of real numbers representing one component of $\{\bar{p}^k_i\}$. We discuss the conditions under which the values $\{f^k_i\}_{i=0}^{2^k n}$, $k \ge 0$, can be extended to a continuous function on $[0, n]$ and the smoothness properties of this function, depending on the values of the parameter w.

Theorem 3.1. *Given $f^0_i = f_i$, $-2 \le i \le n+2$, let the values f^k_i at points $2^{-k} i$, $-2 \le i \le 2^k n+2$, $k \ge 0$, be defined recursively by*

$$\begin{cases} f^{k+1}_{2i} = f^k_i, & -1 \le i \le 2^k n+1, \\ f^{k+1}_{2i+1} = \left(\tfrac{1}{2}+w\right)\left(f^k_i+f^k_{i+1}\right) - w\left(f^k_{i-1}+f^k_{i+2}\right), & -1 \le i \le 2^k n. \end{cases} \tag{3.1}$$

Then, for $|w| < \frac{1}{4}$, there exists $f \in C[0, n]$ such that $f(2^{-k} i) = f^k_i$, $0 \le i \le 2^k n, k \ge 0$.

Proof. Let f^k be the piecewise linear interpolant to the values $\{f_i^k\}_{i=-2}^{2^k n+2}$. We wish to examine $\lim f^k$, as $k \to \infty$. The maximal difference between the functions f^{k+1} and f^k on any interval $[2^{-k}i, 2^{-k}(i+1)]$ is attained at the point $2^{-(k+1)}(2i+1)$, $-1 \le i \le 2^k n$, and its value is

$$f_{2i+1}^{k+1} - \tfrac{1}{2}\left(f_i^k + f_{i+1}^k\right) = w\left(f_i^k - f_{i+1}^k\right) + w\left(f_{i+1}^k - f_{i-2}^k\right). \tag{3.2}$$

Now, directly from (3.1), we have

$$f_{2i+1}^k - f_{2i}^k = \tfrac{1}{2}\left(f_{i+1}^{k-1} - f_i^{k-1}\right) + w\left(f_i^{k-1} - f_{i-1}^{k-1}\right) + w\left(f_{i+1}^{k-1} - f_{i+2}^{k-1}\right), \tag{3.3}$$

and

$$f_{2i+1}^k - f_{2i+2}^k = \tfrac{1}{2}\left(f_i^{k-1} - f_{i+1}^{k-1}\right) + w\left(f_i^{k-1} - f_{i-1}^{k-1}\right) + w\left(f_{i+1}^{k-1} - f_{i+2}^{k-1}\right). \tag{3.4}$$

Using (3.3) and (3.4) recursively gives

$$\max_{-2 \le i \le 2^k n+1} \left| f_{i+1}^k - f_i^k \right| \le \alpha^k \max_{-2 \le i \le n+1} \left| f_{i+1}^0 - f_i^0 \right| \equiv \alpha^k \cdot K, \tag{3.5}$$

where $\alpha = \frac{1}{2} + 2|w|$. Let $\|\cdot\|_\infty$ denote the maximum norm on $[0, n]$. Then by (3.2) and (3.5)

$$\begin{aligned} \| f^{k+1} - f^k \|_\infty &= \max_{0 \le i \le 2^k n-1} \left| w\left(f_i^k - f_{i-1}^k\right) + w\left(f_{i+1}^k - f_{i+2}^k\right) \right| \\ &\le 2|w| \max_{-1 \le i \le 2^k n+1} \left| f_{i+1}^k - f_i^k \right| \\ &\le 2K|w|\alpha^k. \end{aligned}$$

Thus if $|\alpha| < 1$, i.e., $|w| < \frac{1}{4}$, the sequence of continuous functions $\{f^k\}$ is a Cauchy sequence and

$$\lim_{k \to \infty} f^k = f \in C[0, n].$$

This completes the proof since obviously $f^m(2^{-k}i) = f_i^k$ for $0 \le i \le 2^k n$ and any $m \ge k$. □

Theorem 3.2. *Let $f \in C[0, n]$ be the limit function of the process* (3.1) *with* $|w| < \frac{1}{4}$. *Then for* $0 < w < \frac{1}{8}$, $f \in C^1[0, n]$.

Remark 3.1. A formula for the values of the derivative f' at mesh points in any level is given in Corollary 4.1.

Proof. Consider the divided differences

$$d_i^k = 2^k\left(f_{i+1}^k - f_i^k\right), \quad -2 \le i \le 2^k n + 1. \tag{3.6}$$

and let d^k denote the piecewise linear interpolant to d_i^k at the points $2^{-k}i$, $-2 \le i \le 2^k n + 1$, $d^k \in C[-2^{1-k}, n + 2^{-k}]$. We aim to show that $\{d^k\}_{k=0}^{\infty}$ is a Cauchy sequence. From (3.3) and (3.4) we have

$$\begin{cases} d_{2i}^{k+1} = d_i^k + 2wd_{i-1}^k - 2wd_{i+1}^k, \\ d_{2i+1}^{k+1} = d_i^k - 2wd_{i-1}^k + 2wd_{i+1}^k. \end{cases} \tag{3.7}$$

Hence

$$d_{2i+1}^{k+1} - d_{2i}^{k+1} = 4w\left(d_{i+1}^k - d_{i-1}^k\right) = 4w\left(d_i^k - d_{i-1}^k\right) + 4w\left(d_{i+1}^k - d_i^k\right), \tag{3.8}$$

and

$$\begin{aligned} d_{2i+2}^{k+1} - d_{2i+1}^{k+1} &= d_{i+1}^k - d_i^k + 2w\left(d_i^k + d_{i-1}^k\right) - 2w\left(d_{i+2}^k + d_{i+1}^k\right) \\ &= (1-4w)\left(d_{i+1}^k - d_i^k\right) - 2w\left(d_{i+2}^k - d_{i+1}^k\right) - 2w\left(d_i^k - d_{i-1}^k\right). \end{aligned} \tag{3.9}$$

Combining (3.8) and (3.9) we find that for $0 \le w \le \frac{1}{8}$,

$$\max_j \left| d_{j+1}^{k+1} - d_j^{k+1} \right| \le \max_j \left| d_{j+1}^k - d_j^k \right|. \tag{3.10}$$

This estimate is not strong enough to prove convergence. We thus consider double step inequalities. From (3.8) and (3.10) we get

$$\left| d_{2i+1}^{k+1} - d_{2i}^{k+1} \right| \le 8w \max_j \left| d_{j+1}^{k-1} - d_j^{k-1} \right|. \tag{3.11}$$

To estimate $\left| d_{2i+2}^{k+1} - d_{2i+1}^{k+1} \right|$ we make use of (3.8) and (3.9). Thus

$$\begin{aligned} d_{4i+2}^{k+1} - d_{4i+1}^{k+1} = {} & 4w^2\left(d_{i-1}^{k-1} - d_{i-2}^{k-1}\right) + 2w(1-2w)\left(d_i^{k-1} - d_{i-1}^{k-1}\right) \\ & + 2w(1-2w)\left(d_{i+1}^{k-1} - d_i^{k-1}\right) + 4w^2\left(d_{i+2}^{k-1} - d_{i+1}^{k-1}\right), \end{aligned} \tag{3.12}$$

and

$$\begin{aligned} d_{4i+4}^{k+1} - d_{4i+3}^{k+1} = {} & -2w\left(d_i^{k-1} - d_{i-1}^{k-1}\right) + (1-8w)\left(d_{i+1}^{k-1} - d_i^{k-1}\right) \\ & -2w\left(d_{i+2}^{k-1} - d_{i+1}^{k-1}\right). \end{aligned} \tag{3.13}$$

From (3.12) it follows that

$$\left| d_{4i+2}^{k+1} - d_{4i+1}^{k+1} \right| \le 4w \max_j \left| d_{j+1}^{k-1} - d_j^{k-1} \right|, \quad 0 \le w \le \tfrac{1}{2}, \tag{3.14}$$

and from (3.13)

$$\left| d_{4i+4}^{k+1} - d_{4i+3}^{k+1} \right| \le (1-4w) \max_j \left| d_{j+1}^{k-1} - d_j^{k-1} \right|, \quad 0 \le w \le \tfrac{1}{8}. \tag{3.15}$$

Altogether, by (3.11), (3.14), and (3.15) we obtain

$$\max_j \left| d_{j+1}^{k+1} - d_j^{k+1} \right| \le \beta \max_j \left| d_{j+1}^{k-1} - d_j^{k-1} \right|, \tag{3.16}$$

where $0 < \beta < 1$ provided $0 < w < \frac{1}{8}$.

To show that $\{d^k\}$ is a Cauchy sequence we first note that the maximal difference between d^k and d^{k+1} is attained at a point on the $(k+1)$th mesh. Using (3.7)

$$\begin{aligned} d^{k+1}(2^{-k}i) - d^k(2^{-k}i) &= d_{2i}^{k+1} - d_i^k \\ &= -2w\left(d_{i+1}^k - d_i^k\right) - 2w\left(d_i^k - d_{i-1}^k\right), \end{aligned}$$

and

$$\begin{aligned} & d^{k+1}(2^{-k}i + 2^{-(k+1)}) - d^k(2^{-k}i + 2^{-(k+1)}) \\ & \quad = d_{2i+1}^{k+1} - \tfrac{1}{2}\left(d_i^k + d_{i+1}^k\right) \\ & \quad = -\left(\tfrac{1}{2} - 2w\right)\left(d_{i+1}^k - d_i^k\right) + 2w\left(d_i^k - d_{i-1}^k\right). \end{aligned}$$

Thus

$$\| d^{k+1} - d^k \|_\infty \le \tfrac{1}{2} \max_j \left| d_{j+1}^k - d_j^k \right|, \quad 0 < w < \tfrac{1}{8}. \tag{3.17}$$

Combining (3.16) and (3.17) it is clear that $\{d^k\}$ is a Cauchy sequence, i.e.,

$$\lim_{k \to \infty} d^k = d \in C[0, n] \quad \text{for } 0 < w < \tfrac{1}{8}.$$

It remains to show that $d = f'$ where f is the limit function of the process. Consider the Bernstein polynomial for the data $\{f_i^k\}$ on $[0, n]$.

$$b_k(t) = \sum_{i=0}^{N} \binom{N}{i} \left(\frac{t}{n}\right)^i \left(1 - \frac{t}{n}\right)^{N-i} f_i^k, \quad N = 2^k n. \tag{3.18}$$

Then its derivative is the Bernstein polynomial for the data $\{d_i^k\}$:

$$b_k'(t) = \sum_{i=0}^{N-1} \binom{N-1}{i} \left(\frac{t}{n}\right)^i \left(1-\frac{t}{n}\right)^{N-1-i} d_i^k. \tag{3.19}$$

From the uniform convergence of the Bernstein polynomials we conclude that $b_k \to f$ and $b_k' \to d$ uniformly on $[0, n]$ as $k \to \infty$. Hence $f' = d \in C[0, n]$. □

To apply these results to the curves and surfaces determined by the subdivision schemes (1.2) and (2.6) respectively, let $\{B_i(t)\}_{i=-2}^{n+2}$ denote the finite support cardinal basis functions resulting from the application of the process (3.1) to the data $f_i = \delta_{ij}$, $j \in \mathbb{Z}$, $-2 \le i \le n+2$. Then the curve which is the limit of the subdivision process (1.2) can be represented as

$$\bar{p}(t) = \sum_{i=0}^{n} \bar{p}_i^0 B_i(t). \tag{3.20}$$

Also the tensor product surface discussed in Section 2 may be written as

$$\bar{p}(t, s) = \sum_{i=0}^{n} \sum_{j=0}^{m} \bar{p}_{ij}^0 B_i(t) B_j(s). \tag{3.21}$$

It is clear from these representations that the resulting curve and surface are at least as smooth as the basis functions. Hence for $0 < w < \frac{1}{8}$ the curve $\bar{p}(t)$ and the surface $\bar{p}(t, s)$ are continuous and have continuous tangent vectors and planes respectively.

4. Convergence analysis – necessary conditions

In this section we derive necessary conditions for the convergence of the process (3.1) to a $C[0, n]$ and to a $C^1[0, n]$ function, and show that for arbitrary sets of data the limit function does not have a second derivative at any point if $w \neq 0$.

Let us denote by $F^k = (F_{-2}^k, \ldots, F_2^k)^T \in \mathbb{R}^5$, the values attributed by the process (3.1) to the points

$$2^{-m} n_0 + 2^{-(k+m)} j, \quad j = -2, \ldots, 2, \quad k \ge 0 \tag{4.1}$$

for fixed $m \ge 0$ and $0 \le n_0 \le 2^m n$. Observe that $F_0^k = F_0^0$, $k \ge 0$.

Then it is easily seen that

$$F^{k+1} = AF^k \tag{4.2}$$

where

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ -w & \frac{1}{2}+w & \frac{1}{2}+w & -w & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & -w & \frac{1}{2}+w & \frac{1}{2}+w & -w \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}. \tag{4.3}$$

Since the process (3.1) reproduces linear functions, the vector $e = (1,1,1,1,1)^T \in \mathbb{R}^5$ and the vector $l = (-2, -1, 0, 1, 2)^T$ satisfy

$$Ae = e, \qquad Al = \tfrac{1}{2} l. \tag{4.4}$$

Let us denote by λ_3, λ_4, λ_5 the eigenvalues of A not specified by (4.4).

If the values generated by the subdivision process (3.1) define a continuous function f, then necessarily

$$\lim_{k\to\infty} \boldsymbol{F}^k = f(2^{-m}n_0)\boldsymbol{e} \tag{4.5}$$

where $f(2^{-m}n_0) = F_0^0$. Moreover, if f is differentiable, then necessarily

$$\lim_{k\to\infty} 2^{k+m}D\left(\boldsymbol{F}^k - F_0^k\boldsymbol{e}\right) = f'(2^{-m}n)\hat{\boldsymbol{e}} \tag{4.6}$$

where $\hat{\boldsymbol{e}} = (1, 1, 0, 1, 1)^{\mathrm{T}}$ and

$$D = \mathrm{diag}\left(-\tfrac{1}{2}, -1, 1, 1, \tfrac{1}{2}\right). \tag{4.7}$$

These observations lead to

Theorem 4.1. *The subdivision process* (3.1) *defines a continuous function f only if*

$$|\lambda_i| < 1, \quad i = 3, 4, 5. \tag{4.8}$$

Moreover, f is differentiable only if

$$|\lambda_i| < \tfrac{1}{2}, \quad i = 3, 4, 5. \tag{4.9}$$

Proof. Let $\boldsymbol{e}$, $\boldsymbol{l}$, $\boldsymbol{v}^3$, 4, $\boldsymbol{v}^5$ be the set of generalized eigenvectors of A. Since the third row of A is a left eigenvector for A corresponding to the eigenvalue 1,

$$v_0^j = 0, \quad j = 3, 4, 5,$$

and there exist $\alpha_2, \ldots, \alpha_5$ such that

$$\boldsymbol{F}^0 = F_0^0\boldsymbol{e} + \alpha_2\boldsymbol{l} + \sum_{i=3}^{5} \alpha_i\boldsymbol{v}^i. \tag{4.10}$$

Repeated application of A to (4.10), in view of (4.2) and (4.4), yields

$$\boldsymbol{F}^k = A^k\boldsymbol{F}^0 = \sum_{i=3}^{5} \alpha_i A^k\boldsymbol{v}^i + \alpha_2\left(\tfrac{1}{2}\right)^k\boldsymbol{l} + F_0^0\boldsymbol{e}. \tag{4.11}$$

For general data (not linear) $\sum_{i=3}^{5}|\alpha_i|^2 > 0$, and (4.5) holds for all $0 \le n_0 \le 2^m n$ and $m \ge 0$, only if $\lim_{k\to\infty} A^k\boldsymbol{v}^i = \boldsymbol{0}$, $i = 3, 4, 5$, i.e., only if (4.8) holds.

For the proof of the second part observe that

$$\begin{aligned} 2^{k+m}D\left(\boldsymbol{F}^k - F_0^0\boldsymbol{e}\right) &= 2^m\left[D\sum_{i=3}^{5} \alpha_i(2A)^k\boldsymbol{v}^i + \alpha_2 D\boldsymbol{l}\right] \\ &= 2^m\left[\alpha_2\hat{\boldsymbol{e}} + D\sum_{i=3}^{5} \alpha_i(2A)^k\boldsymbol{v}^i\right], \end{aligned} \tag{4.12}$$

and (4.6) can hold only if $\lim_{k\to\infty} (2A)^k\boldsymbol{v}^i = \boldsymbol{0}$, $i = 3, 4, 5$, or equivalently if (4.9) holds. In this case (4.9) and (4.12) imply that

$$f'(2^{-m}n_0) = 2^m\alpha_2. \quad \square \tag{4.13}$$

Remark 4.1. Formula (4.11), in view of (4.13), can be interpreted as a Taylor expansion of the first order, since it reads

$$f\left(2^{-m}n_0 + 2^{-(m+k)}j\right) = f(2^{-m}n_0) + f'(2^{-m}n_0)2^{-(m+k)}j + \sum_{i=3}^{5} \alpha_i\left(A^k\boldsymbol{v}^i\right)_j, \quad j = \pm 1, \pm 2.$$

The 'remainder' term, $\sum_{i=3}^{5} \alpha_i(A^k\boldsymbol{v}^i)_j$, measures the deviation from linearity.

Combining (4.11) and (4.13) we obtain an explicit formula for $f'(2^{-m}n_0)$.

Corollary 4.1. *If the function f defined by the process* (3.1) *is differentiable, then for fixed $m > 0$ and $n_0 \in \{0,\ldots,2^m n\}$*

$$f'(2^{-m}n_0) = \frac{2^{m+k}}{1-4w}\left[\tfrac{1}{2}\left(F_1^k - F_{-1}^k\right) - w\left(F_2^k - F_{-2}^k\right)\right], \quad k \geq 0, \tag{4.14}$$

where $F_j^k = f(2^{-m}n_0 + 2^{-(m+k)}j)$, $j = \pm 1, \pm 2$, $k \geq 0$.

Proof. Let $\boldsymbol{u}$ be the left eigenvector of A, corresponding to the eigenvalue $\frac{1}{2}$, and normalized so that $\boldsymbol{u} \cdot \boldsymbol{l} = 1$. Then by (4.11) and (4.13)

$$f'(2^{-m}n_0) = 2^m \alpha_2 = 2^{m+k}\boldsymbol{u} \cdot \boldsymbol{F}^k.$$

It is easy to check that

$$\boldsymbol{u} = \frac{1}{1-4w}\left(w, -\tfrac{1}{2}, 0, \tfrac{1}{2}, -w\right).$$

This completes the proof of (4.14). □

Next we obtain the ranges of the parameter w for which conditions (4.8) and (4.9) hold.

Theorem 4.2. *The values generated by the subdivision process* (3.1) *define a continuous function f only if*

$$|w| < \tfrac{1}{2}.$$

This function is differentiable everywhere only if

$$0 < w < \tfrac{1}{4}.$$

Proof. It is easy to see from the structure of A that $\frac{1}{2}$, λ_3, λ_4, λ_5 are the eigenvalues of the reduced matrix

$$A' = \begin{bmatrix} 0 & 1 & 0 & 0 \\ -w & \frac{1}{2}+w & -w & 0 \\ 0 & -w & \frac{1}{2}+w & -w \\ 0 & 0 & 1 & 0 \end{bmatrix}, \tag{4.15}$$

with corresponding eigenvectors $(v_{-2}^j, v_{-1}^j, v_1^j, v_2^j)^{\mathrm{T}}$, $j = 2,\ldots,5$. (Here $\boldsymbol{v}^2 = \boldsymbol{l}$.) By symmetry, the eigenvectors of A' satisfy $v_{-2}^j = \mu v_2^j$, and $v_{-1}^j = \mu v_1^j$ with $\mu = +1$ or -1. Direct computation then leads to

$$\lambda_2 = \tfrac{1}{2}, \quad \lambda_3 = 2w, \quad \lambda_4 = \tfrac{1}{4}\left(1 + \sqrt{1-16w}\right), \quad \lambda_5 = \tfrac{1}{4}\left(1 - \sqrt{1-16w}\right). \tag{4.16}$$

Hence (4.8) holds if $|w| < \frac{1}{2}$ and (4.9) holds if $0 < w < \frac{1}{4}$.

Remark 4.2. Theorem 4.2 shows the gap between the necessary conditions for obtaining a $C[0, n]$ or a $C^1[0, n]$ function by the scheme (3.1), and the sufficient conditions as derived in Section 3. It is clear that these sufficient conditions are not sharp, and can be improved by taking more steps in the recursion, before the triangle inequality is applied. It should be noted that the necessary and sufficient conditions for the convergence of uniform refinement schemes, given in [Micchelli, Prautzsch '87], are of too general nature to determine the precise range of w. For all practical purposes (see Section 2) the range $0 < w < \frac{1}{8}$ for a C^1 function is large enough.

We conclude this section by proving that the process (3.1) with $w \neq 0$ applied to arbitrary data, cannot produce functions with a second derivative.

Theorem 4.3. *The function defined by the process* (3.1) *with* $w \in (0, \frac{1}{4})$ *does not have in general a second derivative anywhere on* $[0, n]$.

Proof. A necessary condition for the function f, defined by (3.1), to be twice differentiable at $2^{-m}n_0$ is

$$\lim_{k\to\infty} 4^{k+m}D^2\left(\boldsymbol{F}^{k-1} - 2\boldsymbol{F}^k + F_0^0\boldsymbol{e}\right) = f''(2^{-m}n_0)\hat{\boldsymbol{e}}, \tag{4.17}$$

which in view of (4.11) and (4.4) becomes

$$4^{m+1}\lim_{k\to\infty} D^2 \sum_{j=3}^{5} \alpha_j (I - 2A)(4A)^{k-1}\boldsymbol{v}^j = f''(2^{-m}n_0)\hat{\boldsymbol{e}}. \tag{4.18}$$

If f is not linear, then $f''(2^{-m}\,n_0) \neq 0$ for some m and n_0. Therefore (4.18) holds only if $D^{-2}\hat{\boldsymbol{e}}$ is an eigenvector of A corresponding to the eigenvalue $\frac{1}{4}$. Now by (4.16), A has eigenvalue $\frac{1}{4}$ with $w \in (0, \frac{1}{4})$ only if $w = \frac{1}{8}$ or $w = \frac{1}{16}$. If $w = \frac{1}{8}$, then $\lambda_3 = \frac{1}{4}$ but $\boldsymbol{v}^3 \neq D^{-2}\hat{\boldsymbol{e}}$, while for $w = \frac{1}{16}$, $\lambda_4 = \frac{1}{4}$, $\boldsymbol{v}^4 = D^{-2}\hat{\boldsymbol{e}}$, and the process reproduces polynomials of degree ≤ 3 (A has the eigenvectors $D^{-j}\hat{\boldsymbol{e}}$ with corresponding eigenvalues $(\frac{1}{2})^j$, $j = 1, 2$). Returning to (4.18) with $w = \frac{1}{16}$, and observing that $0 < \lambda_3 < \frac{1}{4}$, $\lambda_5 = \lambda_4 = \frac{1}{4}$, we conclude that the limit does not exist whenever $\alpha_5 \neq 0$. □

Remark 4.3. Although $w = \frac{1}{16}$ does not yield a twice differentiable function, yet it improves the rate of approximation (see Section 5). For general $2r$-point schemes of the form

$$f_{2i}^k = f_i^k, \quad i \in \mathbb{Z},$$

$$f_{2i+1}^k = \sum_{j=-r+1}^{r} w_j f_{i+j}^k, \quad i \in \mathbb{Z},$$

the reproduction of polynomials of degree $\leq s$ is necessary for the scheme to define a limit function which is s-time differentiable [Dyn, Levin '86].

5. Rate of approximation

To analyse the approximation properties of the 4-point scheme, we consider data taken from a function g defined on $\hat{I} = [-\epsilon, 1 + \epsilon]$, at the equidistant points $\{ih \mid -2 \leq i \leq n+2\}$, $h = 1/n < \epsilon/2$. The subdivision process (3.1) produces values at the points $\{2^{-k}ih \mid -2 \leq i \leq 2^k n + 2\}$, $k > 0$, and we are interested in the order of h by which the limit function f^h, defined by these values, approximates the function g in the interval $[0, 1]$.

For this analysis we note that the 4-point scheme is designed so that it reproduces linear functions for any value of w. With the special value $w = \frac{1}{16}$, the scheme reproduces quadratic and cubic polynomials as well.

Using this property it is not difficult to prove the following approximation theorem.

Theorem 5.1. *Let* $f_i^h = g(ih)$, $-2 \leq i \leq n+2$, $h = 1/n < \epsilon/2$ *for* $g \in C^2(\hat{I})$, *and let* f^h *denote the limit function defined by the scheme* (3.1) *with* $|w| < \frac{1}{4}$. *Then for* $I = [0, 1]$

$$\max_{0\leq x\leq 1} |f^h(x) - g(x)| = \| f^h - g \|_{\infty,I} = O(h^2). \tag{5.1}$$

Moreover, if $g \in C^4(\hat{I})$ *and* $w = \frac{1}{16}$, *then*

$$\| f^h - g \|_{\infty,I} = O(h^4). \tag{5.2}$$

Proof. Consider the function B_i^h, $-2 \le i \le n+2$, defined by the process for the data $f_j^h = \delta_{ij}$, $j \in \mathbf{Z}$. Then clearly $B_i^h(x) \in C(\hat{I})$ and the support of B_i^h is the interval $[(i-3)h, (i+3)h]$. Since $B_i^h(x) = B_0^1(x/h - i)$

$$\| B_i^h \|_{\infty,\hat{I}} = \| B_0^1 \|_{\infty,[-3,3]} < \infty. \tag{5.3}$$

By the linearity of the scheme (3.1),

$$f^h(x) = \sum_{i=-2}^{n+2} f_i^h B_i^h(x), \quad x \in [0, 1], \tag{5.4}$$

and the restriction of f^h to $I_{i,i+1}^h = [ih, (i+1)h]$, denoted by $f_{[i]}^h$, is

$$f_{[i]}^h(x) = \sum_{j=i-2}^{i+3} f_j^h B_j^h(x), \quad i = 0, \dots, n-1. \tag{5.5}$$

For $|w| < \frac{1}{4}$, $g \in C^2(\hat{I})$, and $0 \le i \le n-1$ let p_i^h be a linear approximant to g on the interval $I_{i-2,i+3}^h = [(i-2)h, (i+3)h]$ satisfying

$$\| g - p_i^h \|_{\infty, I_{i-2,i+3}^h} \le C \| g'' \|_{\infty, I_{i-2,i+3}^h} h^2. \tag{5.6}$$

Then

$$\| g - f^h \|_{\infty, I_{i,i+1}^h} \le \| g - p_i^h \|_{\infty, I_{i,i+1}^h} + \| f^h - p_i^h \|_{\infty, I_{i,i+1}^h}, \tag{5.7}$$

and since the scheme (3.1) reproduces linear functions

$$\begin{aligned} \| f^h - p_i^h \|_{\infty, I_{i,i+1}^h} &= \Big\| \sum_{j=i-2}^{i+3} \left[g(jh) - p_i^h(jh) \right] B_j^h(x) \Big\|_{\infty, I_{i,i+1}^h} \\ &\le 6 \| g - p_i^h \|_{\infty, I_{i-2,i+3}^h} \| B_0^1 \|_{\infty,[-3,3]}. \end{aligned} \tag{5.8}$$

Combining (5.6), (5.7) and (5.8) we conclude that

$$\| f^h - g \|_{\infty, I_{i,i+1}^h} \le M \| g'' \|_{\infty, I_{i-2,i+3}^h} h^2, \tag{5.9}$$

with $M = (6 \| B_0^1 \|_{\infty,[-3,3]} + 1)C$.

To obtain (5.1) we consider (5.9) for all $i = 0, \dots, n-1$ and replace $\| g'' \|_{\infty, I_{i-2,i+3}^h}$ by the larger norm $\| g'' \|_{\infty,\hat{I}}$.

For the special case $w = \frac{1}{16}$, $g \in C^4(\hat{I})$ we repeat the above steps to obtain (5.2), only now we use a cubic approximation, q_i^h, to g on $I_{i-2,i+3}^h$, such that

$$\| g - q_i^h \|_{\infty, I_{i-2,i+3}^h} \le \hat{C} \| g^{(4)} \|_{\infty, I_{i-2,i+3}^h} h^4, \tag{5.10}$$

and recall that for $w = \frac{1}{16}$ the scheme (3.1) reproduces polynomials of degree ≤ 3. □

References

W. Boehm, G. Farin, J. Kahmann (1984), A survey of curve and surface methods in CAGD, Computer Aided Geometric Design 1, 1–60.

N. Dyn and D. Levin (1986), Smooth interpolation by bisection algorithms, in: Chui, Schumaker and Ward, eds., Approximation Theory 5, 335–337.

C.A. Micchelli and H. Prautzsch (1987), Uniform refinement of curves, to appear.

A. Weissman (1988), M.Sc. Thesis, Tel-Aviv University, to appear.