Subdivision methods for geometric design

Joe Warren
Department of Computer Science
Rice University

November 15, 1995

Contents

Chapter 1

Introduction

Representing curved, complicated shape is the fundamental problem of geometric design. Building data structures and algorithms for generating, representing and manipulating such shapes is a difficult problem.

One powerful method for representing shape is based on iterated transformations. Let F be a function that maps one geometric shape into another geometric shape. If G_0 is an initial shape, then F defines the infinite sequence of shapes,

$$G_{i+1} = F(G_i).$$

If F is "well-behaved", then there exists a limit shape G that is a fixed point of F,

$$G = F(G).$$

A good example of this technique are the fractal methods of Barnsley [Bar93]. The function F is a collection of affine transformations. If each of the affine transformations in F is *contractive* (reduces the size of the shape in each dimension), then F has a unique fixed point,

$$G = F(G).$$

G is the fractal associated with F.

The beauty of this method is that very complicated shapes can be create with a small collection of affine transformations. Consider the set of affine transformations F that map the triangle of figure 1.1 into three shaded copies of its self. The fixed point of this transformation, shown of the right, is the Serpenski triangle.

Subdivision is another example of an iterated transformation. The geometric domain is piecewise linear objects, usually polygons or polyhedra. The function F consists of two distinct phases

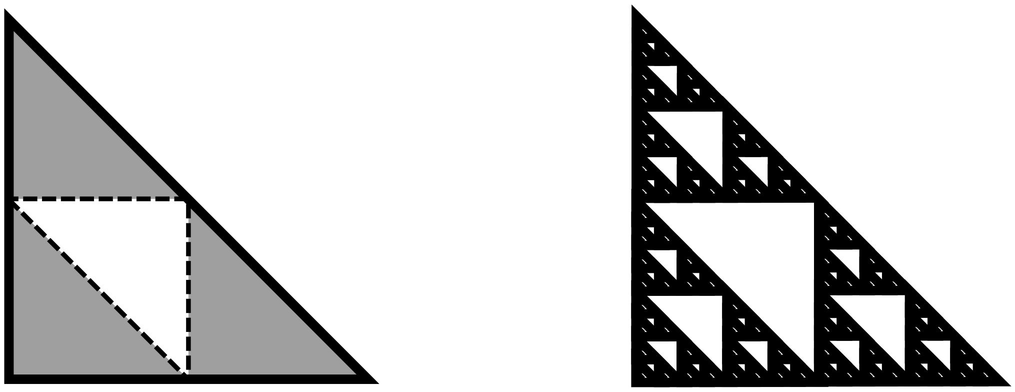

Figure 1.1 The Serpenski triangle

Splitting: Each edge or face is split into two edges or four faces, respectively.

Averaging: Each new vertex introduced by splitting is positioned at a fixed affine combination of its neighbor's positions.

Consider the following examples:

- In figure 1.2, a polygon is transformed into a new polygon with twice as many segments. For this particular transformation, the vertices of the new polygon are placed $\frac{1}{4}$ and $\frac{3}{4}$ of the way between the old vertices. Applying this process repeatedly yields a polygon with a great number of segments that closely approximate a smooth curve. What is this smooth curve? Reisenfeld [Rie75] shows that the curve is a uniform quadratic B-spline whose control points are the original polygon.

- Figure 1.3 depicts piecewise linear subdivision. Each triangular face is split into four subtriangles. New vertices are placed at the midpoints of old edges.

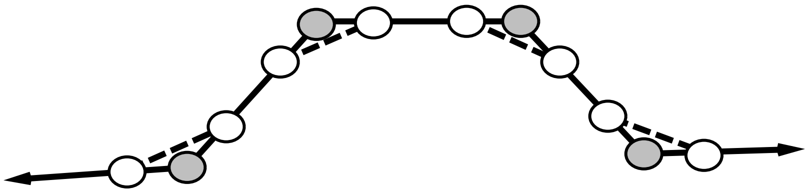

Figure 1.2 A subdivision method

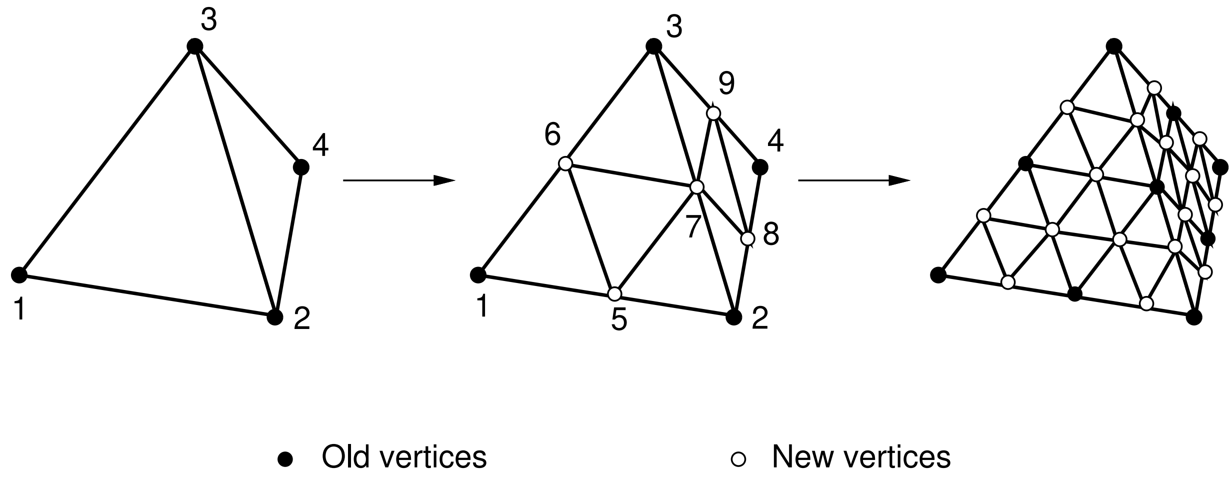

Figure 1.3 A linear method

- Figure 1.4 shows a subdivision scheme developed by Loop [Loo87]. Again, each triangle of a triangular mesh is split into four triangles. However, each new vertex is positioned using a fixed convex combination of the vertices of the original mesh. The final limit surface has a continuous tangent plane.

- Figure 1.5 shows an extension of Loop's method by Hoppe et al. [HDD+94] that incorporates sharp edges into the final limit surface. The vertices of the initial polyhedron of the left are tagged as belonging on a face, edge or vertex of the final limit surface. Based on this tag, different averaging masks are used to produce new polyhedra. In the example, different averaging masks are used on the white edges to produce sharp creases on the final limit surface on the right.

The benefits of subdivision are its simplicity and power.. Implementing a subdivision scheme is simple because only polyhedral modeling is needed. Each vertex of G_i is tagged, specifying whether the descendants of the vertex lies on a vertex, edge or face of the final limit surface. During subdivision, the appropriate averaging mask can be chosen based on this tag. The curved limit shape can be produced to any desired tolerance.

This approach also avoids the need for trimmed surface patches that arises in boundary representations. During subdivision, each curved face of an object is represented by a portion of a polyhedron. The topology of the polyhedron automatically ensures correct connectivity of the object.

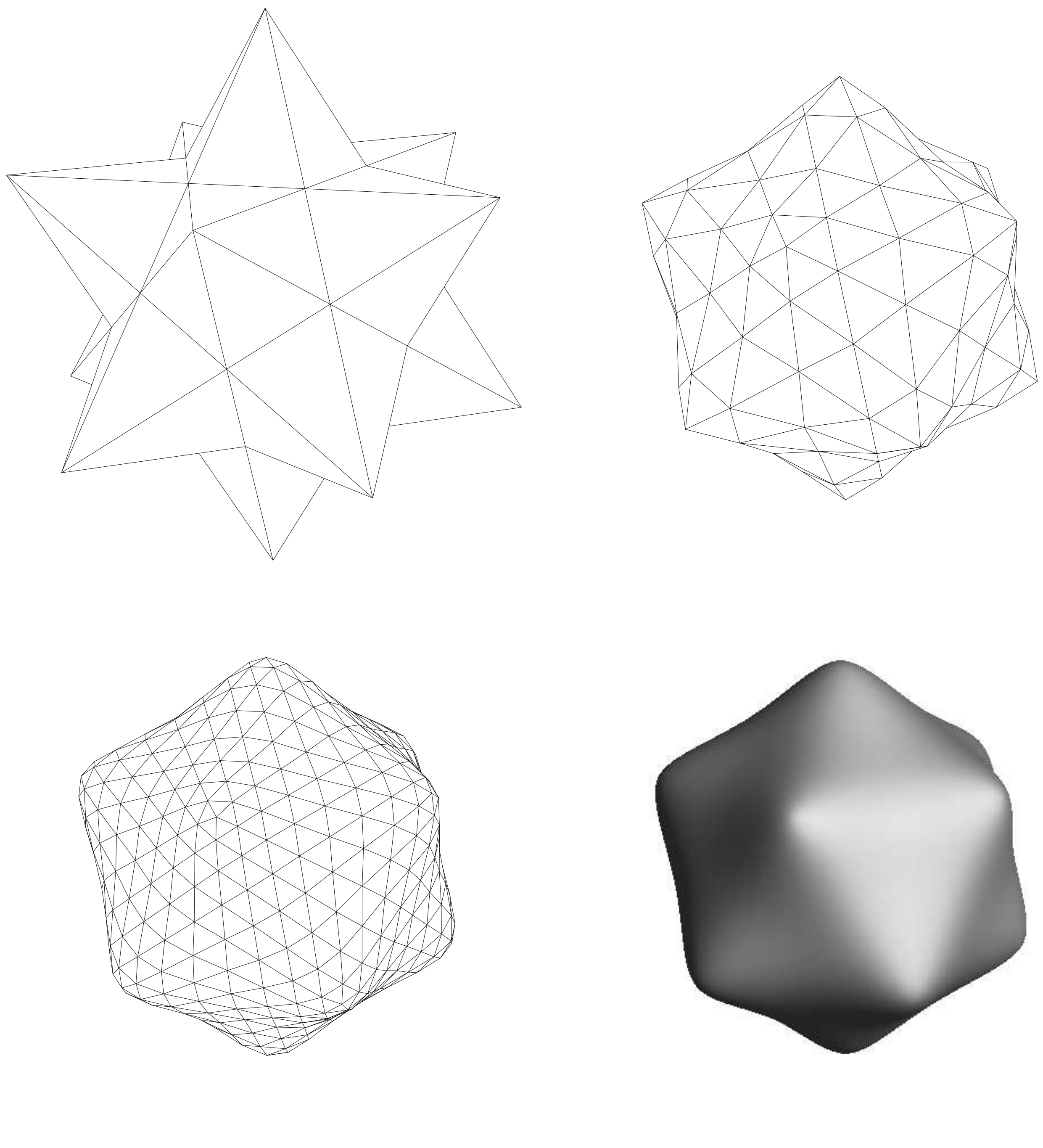

Figure 1.4 The smooth method of Loop

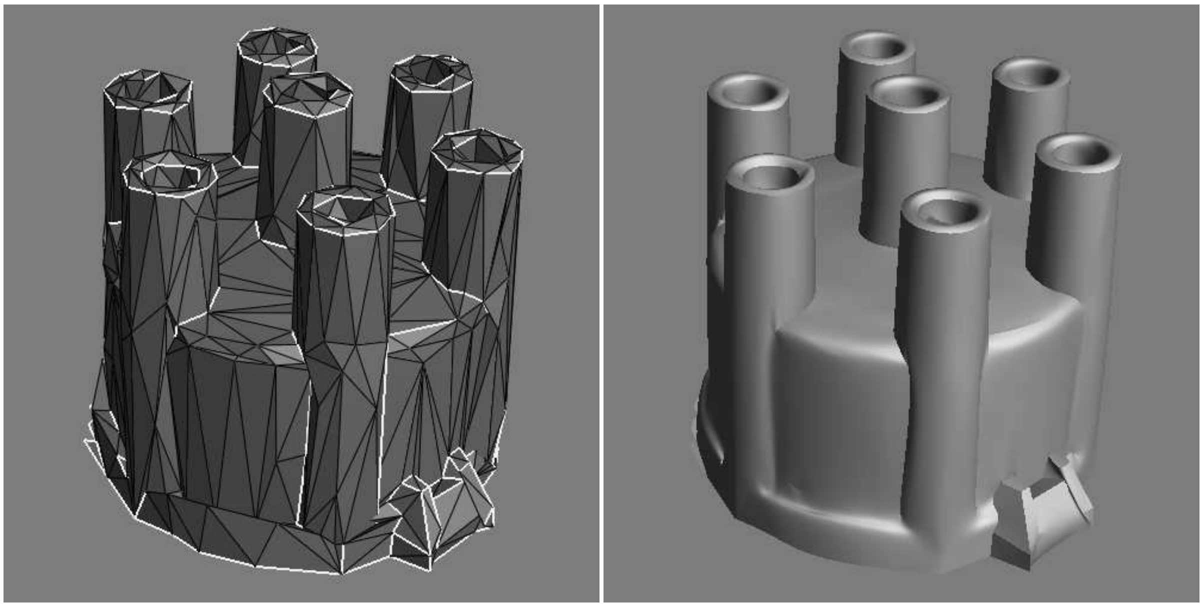

Figure 1.5 The smooth method of DeRose, Hoppe, et. al. (with creases)

Finally, subdivision automatically is powerful because it produces a hierarchy of polyhedra, G_0, G_1, ..., that approximate the final limit object G. Multi-resolution techniques, such as wavelets, for representing an object are easily defined using this hierarchy.

For subdivision methods, the fundamental question is this:

Given the averaging masks, what are the properties of the limit curve/surface?

This question includes determining whether a limit object exists and whether that object is smooth. Another problem is identifying those masks that allow the controlled introduction of edges (creases) into the limit surface.

Given interesting subdivision schemes, another important question is:

What can be computed from this representation?

This class of questions includes converting from other representations such as NURBS or CSG representation into a subdivision representation. Performing geometric operations such as intersection, lofting or fairing is another class of problems. Computing physical properties such as surface area or the solution to differential equations is another example.

This text address the first class of questions. In general, one would like a theory of subdivision that includes many of the techniques of B-splines such as knot insertion. This new theory should include methods for producing highly smooth surfaces from polyhedra of arbitrary topology.

Chapter 2

Subdivision methods for uniform B-splines

This section focuses on a very simple type of geometry, the graph of a function. We define uniform B-splines and describe a subdivision method for them.

2.1 Degree zero B-splines

The characteristic function $U(t)$ is

$$\begin{aligned} U(t) &= 1 \text{ if } 0 \leq t < 1, \\ &= 0 \text{ otherwise.} \end{aligned}$$

The functions $U_i(t) = U(t-i)$ are *translates* of $U(t)$. By construction, these functions are 1 between i and $i+1$. A degree zero B-spline is the sum of translates of $U(t)$.

$$f(t) = \sum_i p_i U_i(t).$$

This B-spline is *uniform* since the breaks between each pair of adjacent constant functions are evenly spaced.

The piecewise constant functions over the half-integers, quarter-integers, etc, are *dilates* of $U(t)$,

$$U_i^j(t) = U(2^j t - i).$$

Subdivision of $f(t)$ involves expressing $f(t)$ in terms of finer and finer dilates of $U(t)$.

$$f(t) = \sum_i p_i^j U_i^j(t).$$

If $U^j(t)$ denotes the row vector whose ith entry is $U_i^j(t)$, then in vector form,

$$f(t) = U^j(t) \cdot p^j. \tag{2.1}$$

This process is possible due to the fact that $U(t)$ can be expressed in terms of its dilates. The function $U(2t)$ is 1 for $0 \leq t < \frac{1}{2}$ and the function $U(2t-1)$ is 1 for $\frac{1}{2} \leq t < 1$. So,

$$U(t) = U(2t) + U(2t-1).$$

This relation is a *subdivision* formula for $U(t)$. More generally,

$$U_i^j(t) = U_{2i}^{j+1}(t) + U_{2i+1}^{j+1}(t).$$

In terms of matrices,

$$U^j(t) = U^{j+1}(t)S,$$

where S is the matrix whose entries $(2i, i)$ and $(2i+1, i)$ are 1 and zero otherwise. A finite portion of S (rows -4 to 3 and columns -4 to 3) is

$$\begin{pmatrix} 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \end{pmatrix}$$

S is the *subdivision* matrix associated with this process. If the initial set of coefficients p^0 are just p, then substituting into equation 2.1 yields the relation

$$p^{j+1} = Sp^j.$$

Application of the subdivision matrix S to p^j produces a new set of coefficients p^{j+1}.

2.2 Higher degree B-splines

Higher degree B-splines can be defined in a variety of ways. Perhaps the simplest definition is through convolution. The continuous convolution of two functions, $g(t)$ and $h(t)$, is

$$g(t) \otimes h(t) = \int_s g(s)h(t-s)ds.$$

We next consider several important properties of convolution.

Theorem 1 *If $f(t)$ is a C^k continuous function, then $U(t) \otimes f(t)$ is C^{k+1} continuous function.*

Proof: By definition,

$$f(t) \otimes U(t) = \int_s f(s)U(t-s)ds.$$

Now, $U(t-s)$ is one exactly when s is between $t-1$ and t. The convolution can be rewritten as

$$U(t) \otimes f(t) = \int_{t-1}^{t} f(s)ds.$$

Since integration raises the differentiability of a function, the theorem follows. □

Dilates and translates arise often during our analysis. The next theorem describes the effects of convolution on dilates and translates.

Theorem 2 *Let $m(t)$ be the convolution of $g(t)$ and $h(t)$.*

$$\begin{aligned} g(t-i) \otimes h(t-j) &= m(t-i-j), \\ g(2t) \otimes h(2t) &= \frac{1}{2}m(2t). \end{aligned}$$

Proof: Apply simple changes of variables to the defintiion of convolution. □

A B-spline basis function of degree n, $N(t)$, satisfies

$$N(t) = \bigotimes_{i=0}^{n} U(t).$$

If $n = 0$, then $N(t) = U(t)$. If $n = 1$, then $N(t) = U(t) \otimes U(t)$ and so on. Next, we list a few important properties of these functions.

- $N(t)$ is piecewise polynomial function of degree n.
- The support of $N(t)$ lies between 0 and $n+1$.
- $N(t)$ is a C^{n-1} function (theorem 1).
- The sum of the translates of $N(t)$ is the function 1.
- $N(t)$ is non-negative everywhere.

If we index the translates and dilates as done for the characteristic function, then

$$N_i^j(t) = N(2^j t - i).$$

A uniform B-spline of degree n is a function $f(t)$

$$f(t) = \sum_i p_i^0 N_i^0(t).$$

We next derive a subdivision formula for the basis function $N(t)$. By definition,

$$\begin{aligned} N(t) &= \bigotimes_{i=0}^{n} U(t), \\ &= \bigotimes_{i=0}^{n} (U(2t) + U(2t-1)). \end{aligned}$$

By the linearity of convolutions, this expression can be rewritten as the sum of various $n+1$-fold convolutions of $U(2t)$ and $U(2t-1)$. By theorem 2,

$$\frac{1}{2^n} N(2t) = \bigotimes_{i=0}^{n} U(2t).$$

Replacing several factors of $U(2t)$ by $U(2t-1)$ yields various translates of $N(2t)$. Therefore, there must exist constants s_k such that

$$N(t) = \sum_k s_k N(2t-k). \tag{2.2}$$

In the next section, we derive an exact expression for the s_k.

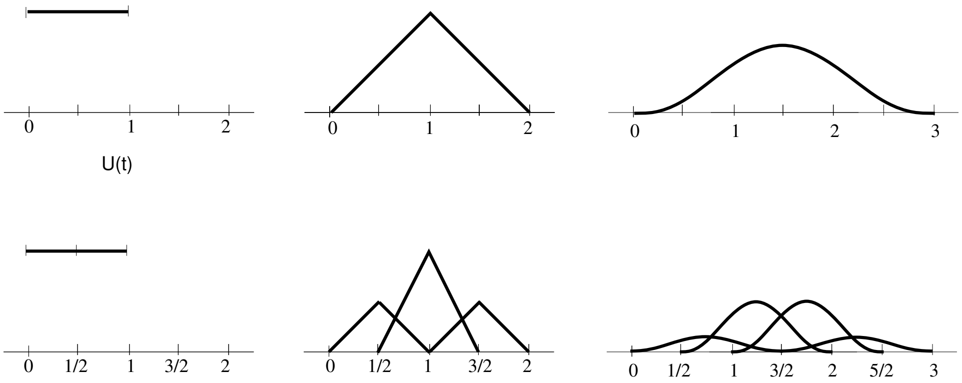

Figure 2.1 Subdivision of low degree B-spline basis functions

For dilates and translates of $N(t)$, the subdivision formula is

$$N_i^j(t) = \sum_k s_k N_{2i+k}^{j+1}(t).$$

In matrix notation, the basis functions are related by

$$N^j(t) = N^{j+1}(t)S,$$

where S is a matrix whose $S_{2i+k,i}$th entry is s_k and zero otherwise. The columns vectors of control points p^j and p^{j+1} are related by

$$p^{j+1} = Sp^j. \tag{2.3}$$

2.3 Subdivision as discrete convolution

In equation 2.3, the subdivision process is expressed as the repeated application of a fixed subdivision matrix to a set of coefficients. For B-splines, another view is possible in terms of discrete convolution.

The *discrete convolution* of two sequences a and b is a third sequence c such that

$$c_k = \sum_{i+j=k} a_i b_j.$$

Discrete convolution can be expressed as polynomial multiplication in the following manner. Associate with a sequence c a unique generating function $C(z)$ such that

$$C(z) = \sum_k c_k z^k.$$

(Lowercase letters denote sequences and upper case letters denote generating functions.) Multiplication of generating functions is equivalent to convolving their associated coefficient sequences. In terms of the definition of discrete convolution,

$$C(z) = A(z)B(z).$$

Discrete convolution can be used to derived the subdivision formula for the continuous convolution of two basis functions.

Theorem 3 *Let $h(t)$ denote the continuous convolution of two functions, $f(t)$ and $g(t)$, with subdivision formulas*

$$\begin{aligned} f(t) &= \sum_i a_i f(2t-i), \\ g(t) &= \sum_j b_j g(2t-j). \end{aligned}$$

Then, $h(t)$ has the subdivision formula

$$h(t) = \sum_k c_k h(2t-k),$$

where $C(z) = \frac{1}{2}A(z)B(z)$.

Proof: By the definition of continuous convolution

$$\begin{aligned} h(t) &= f(t) \otimes g(t), \\ &= (\sum_i a_i f(2t-i)) \otimes (\sum_j b_j g(2t-j)), \\ &= \sum_k (\sum_{i+j=k} a_i b_j f(2t-i) \otimes g(2t-j)). \end{aligned} \tag{2.4}$$

By theorem 2,

$$f(2t-i) \otimes g(2t-j) = \frac{1}{2} h(2t-i-j).$$

Substituting into equation 2.4,

$$\begin{aligned} h(t) &= \sum_k \sum_{i+j=k} \frac{1}{2} a_i b_j h(2t-k), \\ &= \sum_k c_k h(2t-k), \end{aligned}$$

where $C(z) = \frac{1}{2}A(z)B(z)$. □

Using this theorem, the subdivision formula for the B-spline basis function $N(t)$ of degree n can be derived. The subdivision formula for $U(t)$ is

$$U(t) = U(2t) + U(2t-1).$$

The generating function for this subdivision formula is $1+z$. By theorem 3, the generating function $S(z)$ associated with the subdivision formula of equation 2.2 is

$$S(z) = \frac{1}{2^n}(1+z)^{n+1}. \tag{2.5}$$

The exact coefficients s_k can be derived using the binomial theorem.

In the case of quadratic B-splines, a finite portion of S is

$$\begin{pmatrix} 0 & \frac{1}{4} & \frac{3}{4} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{3}{4} & \frac{1}{4} & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{4} & \frac{3}{4} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{3}{4} & \frac{1}{4} & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{4} & \frac{3}{4} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{3}{4} & \frac{1}{4} & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{4} & \frac{3}{4} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{3}{4} & \frac{1}{4} & 0 \end{pmatrix}$$

The rows of S specify the position of the new control points p^{j+1} in terms of the old control points p^j. In this case, the points are placed as specified in Chaikin's algorithm.

2.4 The Lane-Riesenfeld algorithm

The subdivision process of equation 2.3 can be viewed as multiplying the subdivision matrix S times a vector of coefficients p. By construction, $S_{2i+k,i} = s_k$, that is each column of S is shift of its neighboring column by two positions. Let $\hat{S}$ denote the matrix with $\hat{S}_{i+k,i} = s_k$. $\hat{S}$ is a matrix with columns similar to S except each column is shifted by only one entry. The product Sp can be rewritten as

$$Sp = \hat{S}\hat{p}$$

where $\hat{p}_{2i} = p_i$ and $\hat{p}_{2i+1} = 0$.

If $\hat{P}(z)$ is generating function for $\hat{p}$, then the generating function for $\hat{S}\hat{p} = Sp$ is

$$S(z)\hat{P}(z).$$

By construction, $\hat{P}(z)$ is exactly $P(z^2)$. Therefore, the generating function for the sequence Sp is

$$S(z)P(z^2).$$

If $P^j(z)$ denotes the generating function associated with p^j, then in terms of generating functions, equation 2.3 is

$$P^{j+1}(z) = S(z)P^j(z^2).$$

Substituting the definition of $S(z)$ in equation 2.5,

$$P^{j+1}(z) = (\frac{1+z}{2})^n((1+z)P^j(z^2)). \tag{2.6}$$

This formula has a simple geometric interpretation. Lane and Riesenfeld's algorithm [LR80] for subdividing a degree n uniform B-spline is roughly as follows:

- Replicate each coefficient once.
- Apply midpoint averaging to this new sequence n times.

Equation 2.6 is an algebraic expression of this algorithm. The coefficients of $(1+z)P^j(z^2)$ are coefficients of $P^j(z)$ replicated one. Each multiplication by $\frac{1}{2}(1+z)$ represented a midpoint averaging pass over the coefficient sequence.

Chaikin's algorithm is example of this method with $n = 2$. Given a coefficient sequence

$$..., p_0, p_1, p_2, p_3, ...,$$

replicating the coefficients yields the sequence

$$..., p_0, p_0, p_1, p_1, p_2, p_2, p_3, p_3, ...$$

Applying one round of midpoint averaging yields the sequence

$$..., p_0, \frac{p_0+p_1}{2}, p_1, \frac{p_1+p_2}{2}, p_2, \frac{p_2+p_3}{2}, p_3, ...$$

This sequence is equivalent to subdivision for a piecewise linear B-spline. A second round of averaging yields

$$...., \frac{3p_0+p_1}{4}, \frac{p_0+3p_1}{4}, \frac{3p_1+p_2}{4}, \frac{p_1+3p_2}{4}, \frac{3p_2+p_3}{4}, \frac{p_2+3p_3}{4}, ...$$

This sequence is the one produced by Chaikin's algorithm.

Chapter 3

Convergence analysis for uniform subdivision

In this section, we formalize the idea of convergence for a subdivision process. Let the matrix S satisfy $S_{2i+k,i} = s_k$ and be zero otherwise. Given an initial vector p^0, S defines a sequence of vectors p^j satisfying

$$p^{j+1} = Sp^j.$$

In this section, we define a sequence of functions associated with these vectors and examine their convergence properties.

3.1 Parameterization of subdivision methods

The key to interpreting p^j as a function is to assign each entry of p^j a parameter value and graph p^j over these parameter values. In the case of uniform B-splines, the appropriate parameter values arise naturally. The basis functions $N^j(t)$ are piecewise polynomials over dilates of the integers. The jth dilate, n^j, is the vector whose ith entry is $\frac{i}{2^j}$. This dilate n^j provides a suitable parameterization for p^j. After each step of subdivision, the number of coefficients is doubled while the parameter spacing is halved.

This parameterization defines a piecewise linear function $L[n^j, p^j](t)$ satisfying

$$L[n^j, p^j](n_i^j) = p_i^j$$

for all i (see figure 3.1). Under this interpretation, the subdivision process can be viewed as producing a sequence of piecewise linear functions $L[n^j, p^j](t)$ as $j \to \infty$. The rest of this section addresses two basic questions. Does this sequence of functions have a limit? Is this limit function continuous? The next section introduces tools for answering these question.

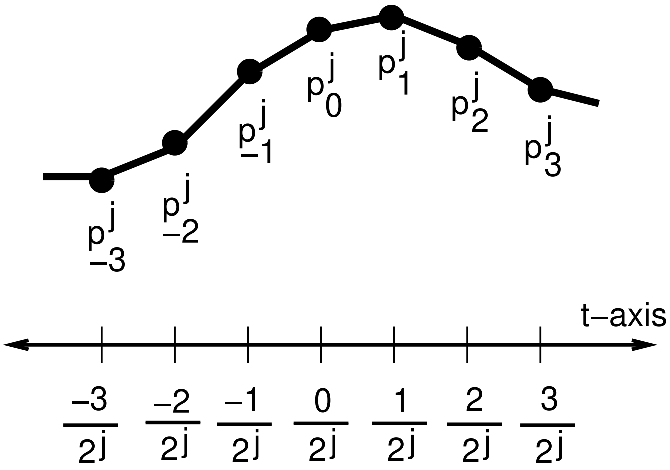

Figure 3.1 Parameterizing a coefficient vector

3.2 Convergence of sequences of functions

First, we recall the definition of a convergent sequence. An infinite sequence

$$f_0, f_1, f_2, \ldots$$

converges to a limit f if for all $\epsilon > 0$ there exist k such that for $j \geq k$,

$$|f_j - f| < \epsilon.$$

This limit is denoted

$$\lim_{j \to \infty} f_j = f$$

If $f_j(t)$ is function of t, then several types of convergence commonly arise. The simplest type of convergence, *pointwise* covergence, defines the limit

$$\lim_{j \to \infty} f_j(t) = f(t)$$

independently for each t.

The main drawback of pointwise convergence is that properties that are true for a sequence of functions $f_j(t)$ may not be true for their limit function $f(t)$. For example, consider the sequence of continuous functions $f_j(t) = t^j$. In the interval $0 \leq t \leq 1$, the limit function $f(t)$ is zero if $t < 1$ and one if $t = 1$. This function is discontinuous. Continuity is not necessarily preserved under pointwise convergence. Another drawback is that the derivatives of the functions $f_j(t)$ do not necessarily converge to the derivative of their pointwise limit. [Tay55, Ch. 18] gives several good examples of this behavior (and is the source for much of the material in this subsection).

The reason for this weakness is in the definition of pointwise convergence. Given an ϵ, each value of t has a distinct k associated with it. An alternative type of convergence, uniform convergence, requires that given an ϵ, a common k exists for all t. A sequence of functions $f_j(t)$ converges *uniformly* to a limit function $f(t)$ if for all $\epsilon > 0$ there exists k such that for all $j \geq k$

$$|f_j(t) - f(t)| < \epsilon$$

for all t Figure 3.2 illustrates this definition. For $j \geq k$, each function $f_j(t)$ must lie in the ribbon bounded above by $f(t) + \epsilon$ and below by $f(t) - \epsilon$.

Uniform convergence is sufficient to ensure that the limit of a sequence of continuous function is a continuous function.

Theorem 4 *Let the $f_j(t)$ be a sequence of continuous functions. If the $f_j(t)$ are uniformly convergent to a limit function $f(t)$, then $f(t)$ is continuous.*

Proof: We show that $f(t)$ is continuous at an arbitrary point t_0.

$$\begin{aligned} f(t) - f(t_0) &= (f(t) - f_j(t)) + (f_j(t) - f_j(t_0)) + (f_j(t_0) - f(t_0)), \\ |f(t) - f(t_0)| &\leq |f(t) - f_j(t)| + |f_j(t) - f_j(t_0)| + |f_j(t_0) - f(t_0)|. \end{aligned} \tag{3.1}$$

Given an $\epsilon > 0$, we must show that $|f(t) - f(t_0)| < \epsilon$ for t sufficiently close to t_0 (this is the definition of continuity). Due to uniform convergence, there exists k, independent of t, such that for all $j \geq k$,

$$|f_j(t) - f(t)| < \frac{\epsilon}{3}$$

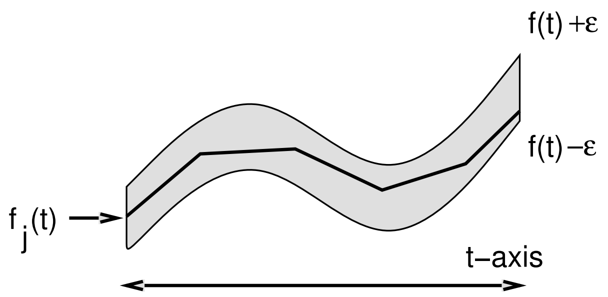

Figure 3.2 Uniform convergence

for all t. Applying this inequality to equation 3.1 yields that

$$|f(t) - f(t_0)| < |f_j(t) - f_j(t_0)| + \frac{2}{3}\epsilon.$$

Since $f_j(t)$ is continuous at t_0, $|f_j(t) - f_j(t_0)| < \frac{\epsilon}{3}$ for t sufficient close to t. So, $|f(t) - f(t_0)| < \epsilon$. This completes the proof. □

To aid in the subsequent analysis, we define the following norms. If $f(t)$ is a function, p is an infinite vector, and S is a bi-infinite matrix, then

$$\begin{aligned} ||f(t)|| &= \max_t |f(t)|, \\ ||p|| &= \max_i |p_i|, \\ ||S|| &= \max_i (\sum_k |S_{ik}|), \end{aligned}$$

where S_i is the ith row of S. A useful property of these norms are that

$$||Sp|| \le ||S|| \cdot ||p||.$$

3.3 Uniform convergence to a continuous function

Uniform convergence of a sequence of continuous functions forces a continuous limit function. We next derive sufficient conditions on S to ensure that the sequence of functions, $L[n^j, p^j](t)$, associated with the subdivision process uniformly converge. Our first condition is simple. We require that the row sums of S are equal to one, that is S times the vector of ones, $\mathbf{1}$, is $\mathbf{1}$. This restriction is a natural since it also insures that the subdivision scheme is affinely invariant. In chapter 5.2, we show that this condition is actually necessary for convergence to a continuous limit curve.

The key to this analysis is examining the behavior of the difference of adjacent coefficients, $p^j_{i+1} - p^j_i$. If Δ is the matrix whose main diagonal is -1 and whose adjacent upper diagonal is 1, then Δp denotes this difference. A finite portion of Δ is

$$\begin{pmatrix} -1 & 1 & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 & 0 \\ 0 & 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & -1 & 1 \\ 0 & 0 & 0 & 0 & -1 \end{pmatrix}. \tag{3.2}$$

If the entries of Δp^j converges to zero as $j \to \infty$, then intuitively the limit of the sequence of functions $L[n^j, p^j](t)$ should not have any discontinuities. The following theorem makes this precise.

Theorem 5 *If there exists $\beta > 0$ and $0 < \alpha < 1$ such that*

$$||\Delta p^j|| < \beta\alpha^j$$

for all $j > 0$, then as $j \to \infty$ the sequence $L[n^j, p^j](t)$ uniformly converges.

Proof:

$$\begin{aligned} L[n^{j+1}, p^{j+1}](t) - L[n^j, p^j](t) &= L[n^{j+1}, Sp^j](t) - L[n^{j+1}, S_1 p^j](t), \\ &= L[n^{j+1}, (S - S_1)p^j](t), \end{aligned}$$

where S_1 is the subdivision matrix for linear B-splines. ($S_1(z) = \frac{1}{2}z^{-1} + 1 + \frac{1}{2}z$.)

Since both $S\mathbf{1} = \mathbf{1}$ and $S_1\mathbf{1} = \mathbf{1}$, $(S - S_1)\mathbf{1} = 0$. Therefore, each row of $S - S_1$ can be written as combination of the rows of Δ. So there exists a matrix A such that $S - S_1 = A\Delta$. Therefore,

$$L[n^{j+1}, p^{j+1}](t) - L[n^j, p^j](t) = L[n^{j+1}, A\Delta p^j](t).$$

Since L defines piecewise linear functions,

$$\begin{aligned} ||L[n^{j+1}, A\Delta p^j](t)|| &\leq ||A\Delta p^j||, \\ &\leq ||A|| * ||\Delta p^j||. \end{aligned}$$

Substituting our hypothesis into this equation yields that

$$||L[n^{j+1}, p^{j+1}](t) - L[n^j, p^j](t)|| < ||A||\beta\alpha^j.$$

Consider the infinite sum

$$L[n^0, p^0](t) + \sum_j (L[n^{j+1}, p^{j+1}](t) - L[n^j, p^j](t)).$$

By the ratio test, this expression must converge to a limit value, call it $F(t)$, for each individual value of t. The difference between $F(t)$ and its approximations is

$$||F(t) - L[n^j, p^j](t)|| < ||A||\frac{\beta}{1-\alpha}\alpha^j$$

for all t. Therefore, the functions $L[n^j, p^j](t)$ uniformly converge to $F(t)$. $\square$

One point to note in the theorem is that β depends on the choice of the initial vector p^0.

We next derive a subdivision process for the difference vectors Δp^j. The subdivision matrix D for this process satisfies

$$\Delta p^{j+1} = D\Delta p^j. \tag{3.3}$$

Since $p^{j+1} = Sp^j$, this relation implies that

$$\Delta S p^j = D\Delta p^j. \tag{3.4}$$

If equations 3.3 and 3.4 are expressed in terms of generating functions, the relation between S and D becomes clear. Equation 3.3 is

$$(1-z)P^{j+1}(z) = D(z)(1-z^2)P^j(z^2)$$

where $D(z)$ is the generating function with coefficients $d_k = D_{2i+k,i}$. Since $P^{j+1}(z) = S(z)P^j(z^2)$, equation 3.4 is

$$(1-z)S(z)P^j(z^2) = D(z)(1-z^2)P^j(z^2).$$

Canceling $(1-z)P^j(z^2)$ on both sides yields that

$$S(z) = (1+z)D(z).$$

Since the row of S sum to one,

$$S(-1) = \sum_k s_k(-1)^k = \sum_k s_{2k} - \sum_k s_{2k+1} = 1 - 1 = 0.$$

Thus, $S(z)$ has a factor of $1+z$.

We can now give direct conditions on D to ensure that this subdivision process converges to zero.

Theorem 6 *Let $||p^0||$ be bounded. If there exists $k \geq 1$ such that $||D^k|| < 1$, then there exists $\beta > 0$ and $0 < \alpha < 1$ such that*

$$||\Delta p^j|| < \beta\alpha^j$$

for all j.

Proof: Recall that

$$\Delta p^j = \Delta S^j p^0 = D^j \Delta p^0.$$

Now, let $||D^k|| = \alpha < 1$. Then,

$$||\Delta p^j|| \leq ||D^j|| \cdot ||\Delta p^0|| < ||\Delta p^0|| \alpha^{\lfloor \frac{j}{k} \rfloor}.$$

This completes the proof. □ This condition is also necessary for uniform convergence of the difference process. [DGL91, Theorem 3.1] gives a proof of its necessity.

In the case of B-splines of degree $n > 0$, this theorem can be used to show that the function $L[n^j, p^j](t)$ converge to the B-spline function. The generating functions for the subdivision process are $S(z) = \frac{1}{2^n}(1+z)^{n+1}$. The generating function for the difference process is $D(z) = \frac{1}{2^n}(1+z)^n$. Since $||D|| = \frac{1}{2}$, this difference process converges to zero. Therefore, the original subdivision scheme converges to a continuous function. Since the B-spline basis functions are non-negative, locally supported and sum to one, the value of a B-spline at a parameter value is a convex combination of nearby coefficients. These coefficients are converging to single common value, the value of the B-spline.

3.4 Convergence to a smooth function

The test for whether a subdivision scheme produces a C^k continuous limit function is straightforward.

Theorem 7 *Let $S(z)$ define a subdivision scheme producing continuous limit curves. Then, $(\frac{1+z}{2})^k S(z)$ defines a scheme producing C^k continuous limit curves.*

Proof: Let $C(t)$ be the basis function satisfying

$$C(t) = \sum_k s_k C(2t - k).$$

Since $C(t)$ is continuous, convolving $C(t)$ by $U(t)$ k times creates a C^k continuous function (theorem 1). The subdivision mask for this function is exactly $(\frac{1+z}{2})^k S(z)$ by theorem 3. □

We conclude this chapter with an interesting example. The subdivision mask $S(z) = \frac{1}{2}z^{-1} + 1 + \frac{1}{2}z$ defines piecewise linear B-splines. (The factor of z^{-1} centers this scheme is the functional setting.) This subdivision scheme is interprelatory since $p_{2i}^{j+1} = p_i^j$. Each control polygon interpolates the vertices of previous control polygons. Consider the subdivision scheme of [DGL87] with generating function

$$S(z) = \frac{1}{16}(-z^{-3} + 4z^{-2} - z^{-1})(1+z)^4.$$

A portion of subdivision matrix S associated with this process is

$$\begin{pmatrix} -\frac{1}{16} & \frac{9}{16} & \frac{9}{16} & -\frac{1}{16} & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & -\frac{1}{16} & \frac{9}{16} & \frac{9}{16} & -\frac{1}{16} & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & -\frac{1}{16} & \frac{9}{16} & \frac{9}{16} & -\frac{1}{16} & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & -\frac{1}{16} & \frac{9}{16} & \frac{9}{16} & -\frac{1}{16} \end{pmatrix}.$$

Figure 3.3 shows the mask used to define the new shaded control point.

This method produces C^1 limit curves. To verify this fact, we divide $S(z)$ by $\frac{1}{2}(1+z)$ and test whether the generating function

$$\frac{1}{8}(-z^{-3} + 4z^{-2} - z^{-1})(1+z)^3$$

produces C^0 functions. This scheme produces C^0 functions if the difference scheme with generating functions

$$\frac{1}{8}(-z^{-3} + 4z^{-2} - z^{-1})(1+z)^2$$

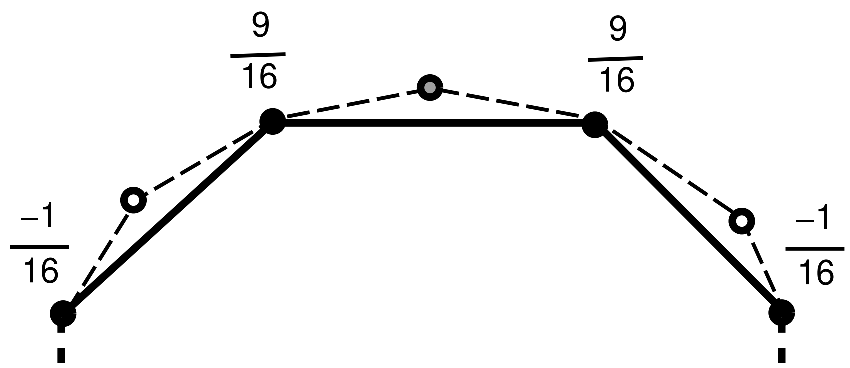

Figure 3.3 A four point interpolatory scheme

converges to zero. A finite portion of the matrix associated with this process is

$$\begin{pmatrix} -\frac{1}{8} & \frac{3}{4} & -\frac{1}{8} & 0 & 0 \\ 0 & \frac{1}{4} & \frac{1}{4} & 0 & 0 \\ 0 & -\frac{1}{8} & \frac{3}{4} & -\frac{1}{8} & 0 \\ 0 & 0 & \frac{1}{4} & \frac{1}{4} & 0 \\ 0 & 0 & -\frac{1}{8} & \frac{3}{4} & -\frac{1}{8} \end{pmatrix}.$$

The norm of this matrix is one. However, the norm of the square of this matrix is $\frac{3}{4}$. Therefore, the original scheme is C^1 continuous.

Chapter 4

Subdivision over irregular knot sequences

An extensive theory of univariate subdivision has been developed that relates the combinations used to the smoothness of the corresponding limit curve. [CDM91, Dyn92] give a comprehensive survey of this theory. However, this theory was developed with the assumption that the parameterization underlying the subdivision method is uniform, that is the ith control point is associated with a parameter value proportional to i. In the parametric case, this is not a substantial restriction. However, in the functional case, this is a non-trivial restriction. In the multi-variate case, this restriction is even worse. The uniform approach cannot deal with the irregular triangulations that often arise during the modeling of complicated shapes.

We next develop a theory of univariate subdivision for irregularly spaced knot sequences in the functional setting. The corresponding theory for the parametric case can then be derived easily. A theory for the irregular, functional case is also useful in applications such as finite element analysis that are intrinsically functional. Eventually, we hope that this theory will be a special case of a general multi-variate theory of subdivision over irregular triangulations.

Our approach to defining subdivision is similar to that of the regular case. A subdivision method is driven by a sequence of scalar values called *knots*. We require that there exist an $\epsilon > 0$ such that the initial knot sequence $n = n^0$ satisfies $n_{i+1} - n_i > \epsilon$ for all i. This restriction ensures that the knots appear in ascending order and fill the parameter line. Subsequent knot sequences n^j are defined by midpoint insertion.

$$n_{2i}^{j+1} \quad = \quad n_i^j, \tag{4.1}$$

$$n_{2i+1}^{j+1} = \frac{n_i^j + n_{i+1}^j}{2}. \tag{4.2}$$

4.1 Definition of irregular subdivision schemes

A subdivision scheme is a map from knot sequences n to a collection of subdivision rules $S[n]$. Given the sequence of knot vectors n^j, the subdivision rules $S[n^j]$ may be viewed as a matrix that maps the jth set of control points p^j into a new set of control point p^{j+1}.

$$p^{j+1} = S[n^j]p^j. \tag{4.3}$$

We restrict our attention to subdivision schemes that satisfy four important properties

Compact support: There exists nonnegative a and b such that, for all k, the kth column of $S[n]$ is zero except from row $2k - a$ to $2k + b$.

Affine invariance: For any scaled and translated knot sequence $\alpha n + \beta$ with $\alpha > 0$,

$$S[n] = S[\alpha n + \beta].$$

Index invariance: If $r(n)$ is the vectors whose ith entry is n_{i+1}, then

$$S[r(n)]_{i,j} = S[n]_{i+2,j+1} \quad \forall i, j.$$

Local definition: The kth column of $S[n]$ depends only on the knots n_i where $|k - i|$ is bounded independent of k.

Compact support ensures that the infinite sum in equation 4.3 is well-defined. The number $a + b + 1$ is the *column height* of S. The 2-1 slant of $S[n]$ doubles the number of control points during each step of subdivision. If the knot sequence n is regular (i.e $n_i = i$), then affine invariance ensures the resulting subdivision scheme is *stationary*, that is $S[n] = S[\frac{n}{2^j}] = S[n^j]$ for all j. For regular knot sequences n, the index shift $r(n)$ is also an affine transformation of n and $S[n] = S[r(n)]$. In this case, index invariance forces the subdivision scheme to be *uniform* ($S[n]_{i,j} = S[n]_{i+2,j+1}$). Local definition is critical in showing that midpoint subdivision of irregular knot sequences leads to stationary subdivision schemes.

As in the uniform case, the control points p^j are parameterized by the knot vector n^j. $L[n^j, p^j](t)$ denotes the piecewise linear function interpolating the points (n^j_i, p^j_i) for all i (see figure 4.1). $L[n^j, p^j](t)$ can be viewed as a function in t since the n^j_i are indexed in ascending value. The knot vectors n^j provide *parameter* values for the control points p^j.

The natural object to consider here is the limit of these functions $L[n^j, p^j](t)$ as j goes to infinity. Given an initial set of control points $p = p^0$ and an initial knot vector $n = n^0$, we define the limit function associated with the process of equation 4.3

$$F[n,p](t) = \lim_{j \to \infty} L[n^j, p^j](t).$$

Here, the limit is taken *point-wise*, that is individually for each distinct t. By construction, the limit operator F is linear in p. Specifically,

$$\begin{aligned} F[n, \alpha p](t) &= \alpha F[n,p](t), \\ F[n, p+q](t) &= F[n,p](t) + F[n,q](t). \end{aligned}$$

Scaling the knot vector n is also equivalent to scaling the parameter t,

$$F[\alpha n, p](t) = F[n,p](\frac{t}{\alpha}).$$

4.2 Basis functions

Let e^i be the vector whose ith entry is one with the remaining entries being zero. Given the knot sequence n^j, we associate the function $F[n^j, e^i](t)$ with each control point p^j_i. By the

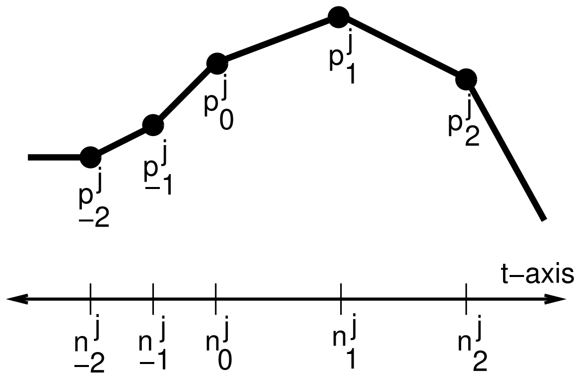

Figure 4.1 Parameterization

linearity of F, these functions are basis functions since

$$F[n^j, p^j](t) = \sum_i p_i^j F[n^j, e^i](t).$$

Just as in the case of B-splines, the knots of n^j determines the support of $F[n^j, e^i](t)$. The column height, $a+b+1$, of $S[n^j]$ determines the width of the support of $F[n^j, e^i](t)$.

Theorem 8 *$F[n^j, e^i](t) = 0$ for $t \notin [n^j_{i-a}, n^j_{i+b}]$.*

Proof: Without loss of generality, we show that $F[n^0, e^0]$ is zero for $t < n^0_{-a}$ and $t > n^0_b$. We keep track of the range of indices of non-zero coefficients during subdivision. After one round of subdivision, the non-zero coefficients range from p^1_{-a} to p^1_b. After k rounds of subdivision, the non-zero coefficients range from $p^k_{-a(2^k-1)}$ to $p^k_{b(2^k-1)}$. The limit of this range is

$$\begin{aligned} \lim_{k\to\infty} n^k_{-a(2^k-1)} = \lim_{k\to\infty} n^k_{-a2^k} &= n^0_{-a}, \\ \lim_{k\to\infty} n^k_{b(2^k-1)} = \lim_{k\to\infty} n^k_{b2^k} &= n^0_b. \end{aligned}$$

□

4.3 Example: Interpolating subdivision

As a running example, we focus on a generalization of the four point, interpolatory scheme of the previous chapter. Let r be an non-negative integer. Given a knot sequence n, $S[n]$ is defined a row at a time.

$$\begin{aligned} S[n]_{2i} &= e^i, \\ S[n]_{2i+1} &= m^i, \end{aligned}$$

where the $m^i_k = 0$ for $k < i - r$ and $k > i + r + 1$. The nonzero entries of m^i are defined by

$$\begin{pmatrix} 1 & 1 & \cdots & 1 \\ n_{i-r} & n_{i-r+1} & \cdots & n_{i+r+1} \\ \cdots & \cdots & \cdots & \cdots \\ (n_{i-r})^{2r+1} & (n_{i-r+1})^{2r+1} & \cdots & (n_{i+r+1})^{2r+1} \end{pmatrix} \begin{pmatrix} m^i_{i-r} \\ m^i_{i-r+1} \\ \cdots \\ m^i_{i+r+1} \end{pmatrix} = \begin{pmatrix} 1 \\ \frac{1}{2}(n_i + n_{i+1}) \\ \cdots \\ (\frac{1}{2}(n_i + n_{i+1}))^{2r+1} \end{pmatrix}$$

The masks m^i can be thought of in the following way. Consider the degree $2r+1$ polynomial $f(t)$ that interpolates the value f_k at knot n_k for $i - r \le k \le i + r + 1$. The nonzero entries of m^i are the combinations of the values f_k necessary to reproduce the value $f(\frac{n_i+n_{i+1}}{2})$ independent of the f_k chosen. The masks m^i are chosen to force polynomial precision of degree $2r+1$.

By the construction of $S[n]$, this subdivision scheme is interpolating. p_{2i}^{j+1} and p_i^j agree and share the same parameter value $n_{2i}^{j+1} = n_i^j$. For the case of regularly spaced n, this subdivision scheme has been heavily analyzed. If $r=0$, then the nonzero portion of $m^i = (\frac{1}{2}, \frac{1}{2})$. This subdivision scheme is simply piecewise linear interpolation. For $r=1$, the nonzero portion of the mask m^i is

$$(\frac{-1}{16}, \frac{9}{16}, \frac{9}{16}, \frac{-1}{16}).$$

This subdivision scheme is the four-point rule of [DGL87]. This method converges to C^1 continuous function.

For irregularly spaced n, this scheme has not been studied. Of course for $r=0$, the scheme produces piecewise linear interpolation. For $r=1$, this scheme produces an irregular four point rule that depends on the local knot spacing. Figure 4.2 show an example of the method. The purpose of this paper is to develop tools for analyzing schemes such as this one.

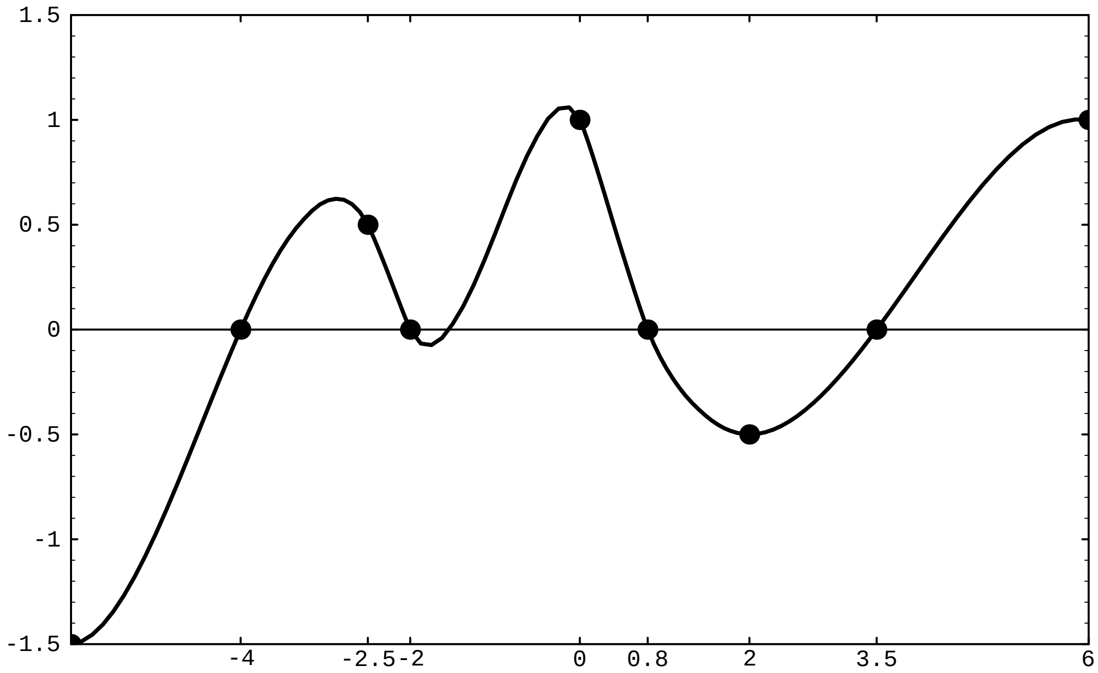

Figure 4.2 An example of the irregular four point method

4.4 Reduction to the stationary case

The key to analyzing the smoothness of subdivision schemes that are compact, affinely invariant, index invariant, and locally defined is reducing the scheme locally to an equivalent stationary scheme. Without loss of generality, we assume that $n_0^0 = 0$ and focus our analysis at the origin. Other case can be handled through translation and re-indexing. During subdivision, larger and larger clusters of knots on either side of $n_0^j = 0$ are regularly spaced. In particular, the knots from $n_{-2^j}^j$ to n_0^j and from n_0^j to $n_{2^j}^j$ are regularly spaced.

Theorem 9 *Let $n_0^0 = 0$. There exists an integer j and knot sequence $\hat{n}$ of the form*

$$\begin{aligned} \hat{n}_i &= (n_{-1}^j)i \quad (\forall \ i < 0), \\ \hat{n}_i &= (n_1^j)i \quad (\forall \ i \geq 0), \end{aligned} \tag{4.4}$$

such that

$$F[n^j, p^j](t) = F[\hat{n}, p^j](t)$$

for all $t \in (n_{-1}^j, n_1^j)$.

Proof: Track those control points produced after k subsequent subdivisions using n^j and $\hat{n}$ as initial knot sequences. By theorem 8, only those control points whose lower indices are in the range $-b + 1 - 2^k$ and $2^k + a - 1$ affect the limit functions in the parameter range (n_{-1}^j, n_1^j).

We next show that if those control point in this range agree after k subdivision, then the control points with indices in the range $-2^{k+1} - b + 1$ and $2^{k+1} + a - 1$ agree after $k + 1$ subdivisions. By construction, n_i^{j+k} and $\frac{\hat{n}_i}{2^k}$ agree for $-2^{j+k} \leq i \leq 2^{j+k}$. If j is chosen large enough, then the subdivision rules generated by n^{j+k} and $\frac{\hat{n}}{2^k}$ must agree for all control points whose indices are in the range $-2^{k+1} - b + 1$ and $2^{k+1} + a - 1$ (independent of k). This fact is a consequence of the local definition of the subdivision scheme.

So by induction, the two scheme agree on all control points assigned to the parameter range (n_{-1}^j, n_1^j) for all subdivisions. Therefore, the two associated limit functions agree. □

This theorem allows us to focus our analysis on the knot sequence in equation 4.4. Henceforth, we take the initial knot sequence $n = n^0$ to be of this form. Applying midpoint subdivision to this sequence is equivalent to division by two,

$$n^j = \frac{n}{2^j}.$$

Since our subdivision scheme is affinely invariant,

$$S[n^j] = S[\frac{n}{2^j}] = S[n].$$

Therefore, in a neighborhood of the origin, we may focus on analyzing the stationary subdivision scheme of the form

$$p^{j+1} = S[n]p^j.$$

Note that $S[n]$ depends only on the relative spacing on the knots on either side of the origin. To simplify, we refer to the matrix $S[n]$ as S.

Chapter 5

Univariate stationary subdivision

5.1 Spectral analysis

Our approach to analyzing the smoothness of $F[n,p](t)$ is to express this function locally as a linear combination of functions $F[n,x_i](t)$ where the x_i are eigenvectors of S. At first glance, determining these eigenvectors seems daunting because S is an infinite matrix. However, in practice, all of the interesting spectral properties of S are captured by a finite submatrix of S.

If the column height of S is $a+b+1$, then let the *bar* operator, $\bar{p}$, select the entries $p_{-b},...,p_a$ from the infinite vector p. These are exactly the entries of p that affect the limit function near the origin. The bar operator applied to the matrix S yields $\bar{S}$, the $a+b+1$ by $a+b+1$ matrix with entries S_{ij} where $-b \geq i,j, \geq a$.

Consider the regular, four point subdivision scheme of section 4.3. Since $r=1$, a and b are 3. So $\bar{S}$ is the $7x7$ matrix

$$\begin{pmatrix} -\frac{1}{16} & \frac{9}{16} & \frac{9}{16} & -\frac{1}{16} & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & -\frac{1}{16} & \frac{9}{16} & \frac{9}{16} & -\frac{1}{16} & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & -\frac{1}{16} & \frac{9}{16} & \frac{9}{16} & -\frac{1}{16} & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & -\frac{1}{16} & \frac{9}{16} & \frac{9}{16} & -\frac{1}{16} \end{pmatrix}. \tag{5.1}$$

The eigenvalues of this matrix are

$$1, \frac{1}{2}, \frac{1}{4}, \frac{1}{4}, \frac{1}{8}, \frac{-1}{16}, \frac{-1}{16}.$$

By construction, $\bar{S}$ contains exactly the nonzero portion of the main diagonal of S. Thus, these values are exactly the nonzero eigenvalues of S. The eigenvectors of $\bar{S}$ can be extended to form eigenvectors of S.

Theorem 10 *Let $\lambda \neq 0$ be an eigenvalue of $\bar{S}$ with associated eigenvector y. Then S has eigenvalue λ with a unique associated eigenvector x such that $\bar{x} = y$.*

The proof of this theorem is simple and left to the reader. The eigenvectors of $\bar{S}$ with eigenvalue zero have no effect on the final limit curve at the origin since after one round of subdivision the control points are mapped to zero. For the sake of simplicity, we assume that $\bar{S}$ has no zero eigenvalues.

If $\bar{S}$ does not have a full set of eigenvectors, then $\bar{S}$ is *defective.* A non-defective $\bar{S}$ has a full set of linearly independent eigenvectors $\bar{x}_0$, ..., $\bar{x}_{a+b}$. The extension of these eigenvectors, x_0, ..., x_{a+b}, can be used in the following manner.

Theorem 11 *Let $\bar{S}$ be a non-defective matrix with eigenvectors $\bar{x}_0$, ..., $\bar{x}_{a+b}$. If*

$$\bar{p} = \sum_{i=0}^{a+b} c_i \bar{x}_i,$$

then, for all $t \in (n_{-1}^0, n_1^0)$,

$$F[n,p](t) = \sum_{i=0}^{a+b} c_i F[n,x_i](t).$$

Proof: The vector $p - \sum_{i=0}^{a+b} c_i x_i$ is zero for entries $-b$ to a. By theorem 8,

$$F[n, p - \sum_{i=0}^{a+b} c_i x_i](t) = 0$$

for all $t \in (n_{-1}^0, n_1^0)$. The theorem follows by the linearity of the limit operator F. $\square$

For defective $\bar{S}$, *generalized* eigenvectors can be used in place of eigenvectors. Each of the generalized eigenvectors, $\bar{x}_0$, ..., $\bar{x}_{a+b}$, is either an eigenvector of $\bar{S}$ or satisfies

$$\bar{S}\bar{x}_i = \lambda_i \bar{x}_i + \bar{x}_{i-1}.$$

These generalized eigenvectors can be extend to a set of infinite vectors, x_0, ..., x_{a+b}, satisfying

$$S x_i = \lambda_i x_i + x_{i-1}. \tag{5.2}$$

The proof of this fact is exactly the same as the proof of theorem 10. Since the generalized eigenvectors are linearly independent, theorem 11 also holds for these vectors. For subsequent theorems, we assume that $\bar{S}$ is non-defective. Where appropriate, we state the variant of the theorem that holds for defective $\bar{S}$ using generalized eigenvectors.

5.1.1 A spectral recurrence

The main message of this chapter is that the smoothness properties of a stationary subdivision scheme are tied to the spectral properties of its subdivision matrix S. In particular, the limit function associated with an eigenvector of a stationary subdivision scheme satisfies a fundamental relation.

Theorem 12 *Let x be an eigenvector of S satisfying $Sx = \lambda x$. Then,*

$$\lambda F[n, x](t) = F[n, x](\frac{t}{2}). \tag{5.3}$$

Proof: The proof consist of simply recalling the definition of F.

$$\begin{aligned} \lambda F[n,x](t) &= F[n, \lambda x](t), \\ &= F[n, Sx](t), \\ &= \lim_{j\to\infty} L[n^j, S^j(Sx)](t), \\ &= \lim_{j\to\infty} L[2n^{j+1}, S^{j+1}x](t), \\ &= \lim_{j\to\infty} L[n^{j+1}, S^{j+1}x](\frac{t}{2}), \\ &= F[n,x](\frac{t}{2}). \end{aligned}$$

□

If x_i is a generalized eigenvector as in equation 5.2, then

$$\lambda_i F[n, x_i](t) + F[n, x_{i-1}](t) = F[n, x_i](\frac{t}{2}). \tag{5.4}$$

Again, the proof is exactly as above.

5.1.2 Properties of the recurrence

The recurrence of theorem 12 is a powerful tool for analyzing stationary subdivision schemes. The following lemma illustrates several properties of such recurrences.

Lemma 1 *Let $g(t)$ be a function satisfying*

$$\lambda g(t) = g(\frac{t}{2}). \tag{5.5}$$

1. *If $g(t)$ is bounded on the interval $(\frac{1}{2}, 1]$ and $|\lambda| < \frac{1}{2^k}$, then the limit as $t \to 0$ of $\frac{g(t)}{t^k}$ is zero.*

2. *If $\lambda = 1$ and $g(t)$ is continuous at the origin, then $g(t)$ is the constant function.*

3. *If $g(d) \neq 0$ for $d \neq 0$ and $|\lambda| > 1$ then the limit as $t \to 0$ of $g(t)$ diverges.*

Proof: For part one, let

$$v = \max_{\frac{1}{2} < t \leq 1} |g(t)|.$$

If $\frac{1}{2^{i+1}} < t \leq \frac{1}{2^i}$, then iterating equation 5.5 yields

$$\begin{aligned} |g(t)| &= |\lambda^i g(2^i t)|, \\ &\leq |\lambda|^i v. \end{aligned}$$

Dividing the lefthand side of this relation by t^k and the righthand side by $\frac{1}{2^{k(i+1)}}$ respectively yields

$$\begin{aligned} |\frac{g(t)}{t^k}| &\leq 2^{k(i+1)} |\lambda|^i v, \\ &\leq (2^k |\lambda|)^i 2^k v \end{aligned}$$

Since $|\lambda| < \frac{1}{2^k}$, the limit as $i \to \infty$ and therefore as $t \to 0$ of $\frac{g(t)}{t^k}$ must be zero.

For part two, we observe that if there exists $d \neq e$ such that

$$g(d) - g(e) = c \neq 0,$$

then by equation 5.5

$$g(\frac{d}{2^i}) - g(\frac{e}{2^i}) = c \neq 0. \tag{5.6}$$

As i goes to infinity, $\frac{d}{2^i}$ and $\frac{e}{2^i}$ approach zero. Equation 5.6 contradicts that fact that $g(t)$ is continuous at the origin. Therefore, $g(d) = g(e)$ for all d and e.

For part three, we note that by hypothesis there exists $d \neq 0$ such that $g(d) \neq 0$. Since $g(\frac{d}{2^i}) = \lambda^i g(d)$ and $|\lambda| > 1$, $g(t)$ must diverge as $i \to \infty$ and $t = \frac{d}{2^i} \to 0$. $\square$

5.2 Necessary conditions for C^k continuity

By theorem 11, we can restrict our smoothness analysis to those limit functions $F[n, x](t)$ where x is an eigenvector of S with eigenvalue λ. If $F[n, x](t)$ is a C^k continuous function, then x and λ must have special properties. Let $F^{(i)}[n,x](t)$ denote the ith derivative of $F[n,x](t)$.

Theorem 13 *Let $Sx = \lambda x$ with $|\lambda| \geq \frac{1}{2^k}$. If $F[n,x](t)$ is C^k continuous and not identically zero, then there exist $0 \leq i \leq k$ such that*

- $\lambda = \frac{1}{2^i}$,
- $F[n,x](t) = c_i t^i$ *for* $c_i \neq 0$.

Proof: We first show that $F^{(k)}[n,x](t)$ is a constant function. Take the kth derivative of equation 5.3.

$$(2^k\lambda)F^{(k)}[n,x](t) = F^{(k)}[n,x](\frac{t}{2}).$$

If $F^{(k)}[n,x](t)$ is not a constant function, then $F^{(k)}[n,x](t)$ must diverge as $t \to 0$ either by part two of lemma 1 ($2^k\lambda = 1$) or part three of lemma 1 ($|2^k\lambda| > 1$). However, this contradict the continuity of $F^{(k)}[n,x](t)$.

Since $F^{(k)}[n,x](t)$ is a constant function, $F[n,x](t)$ is a polynomial function of degree k. Since $F[n,x](t)$ is not identically zero, choose the minimal i such that $F^{(i)}[n,x](0) \neq 0$ and take the ith derivative of equation 5.3.

$$\lambda F^{(i)}[n,x](t) = \frac{1}{2^i}F^{(i)}[n,x](\frac{t}{2}).$$

For lefthand and righthand side of this equation to agree at $t = 0$, λ must be $\frac{1}{2^i}$. Since $F^{(i)}[n,x](t)$ is continuous, by part two of lemma 1, $F^{(i)}[n,x](t)$ must be the constant function. Since the lower order derivatives are zero at the origin, $F[n,x](t)$ is a multiple of t^i. □

If a generalized eigenvector (equation 5.2) produces a C^k limit curve, then its eigenvalue must have modulus less than $\frac{1}{2^k}$. Consider the continuous (C^0) case. If $\lambda_0 = 1$, then $F[n,x_0](t)$ is the constant function. If x_1 were a generalized eigenvector satisfying

$$Sx_1 = \lambda_1 x_1 + x_0,$$

then $\lambda_0 = \lambda_1 = 1$. In terms of limit functions,

$$\begin{aligned} \lambda_1 F[n,x_1](t) + F[n,x_0](t) &= F[n,x_1](\frac{t}{2}), \\ F[n,x_1](t) + 1 &= F[n,x_1](\frac{t}{2}). \end{aligned}$$

As $t \to 0$, $F[n, x_1](t)$ must diverge.

The four point rule is known to yield C^1 continuous functions. This theorem states that the only possible eigenvalues with modulus greater or equal to $\frac{1}{2}$ are 1 and $\frac{1}{2}$. Also the eigenvectors x_0 and x_1 must produce the limit functions 1 and t up to a constant multiple. Using matrix 5.1, one can verify that the kth entries of x_0 and x_1 are 1 and k respectively. The eigenvector $\frac{1}{4}$ has multiplicity two, but has only a single associated eigenvector. Therefore, this scheme is strictly a C^1 continuous.

5.3 Sufficient conditions for C^k continuity

For regularly spaced knot sequences, the subdivision schemes described here generate subdivision matrices whose columns are just shifts of each other. Powerful tools such as Laurent polynomials are available for analyzing such schemes. For knot sequence whose knots are regularly-spaced away from the origin, all but a finite number of columns of the subdivision matrix must agree with the regular case (due to the local definition of the subdivision rules). If the regular case produces C^k continuous curves, then this irregular case also produces a C^k curve except possibly at $t = 0$.

By theorem 11, we need only analyze the smoothness of the functions $F[n, x](t)$ where x is an eigenvaalue of S. By theorem 13, an eigenvector x with eigenvalue $|\lambda| \geq \frac{1}{2^k}$ must reproduce a polynomial if the scheme is to be C^k continuous. This reproduction is usually ensured during construction of S (e.g. interpolatory subdivision). Only if $|\lambda| < \frac{1}{2^k}$ does the smoothness of $F[n, x](t)$ need to be verified.

Theorem 14 *Let $Sx = \lambda x$ with $|\lambda| < \frac{1}{2^k}$. If $F[n, x](t)$ is C^k continuous everywhere except at $t = 0$, then $F[n, x](t)$ is C^k continuous everywhere.*

Proof: We show that $F^{(i)}[n, x](t)$ exists and is continuous at $t = 0$ for $0 \leq i \leq k$. The proof proceeds by induction on i. For the base case $i = 0$, we note that $F[n, x](0)$ must be zero by theorem 12 since $|\lambda| < 1$. The limit as t goes to zero of $F[n, x](t)$ is also zero by part one of lemma 1. Therefore, $F[n, x](t)$ is continuous at $t = 0$.

We proceed with the induction step. First, we show that $F^{(i)}[n, x](0)$ exists and is zero at $t = 0$. By definition,

$$F^{(i)}[n, x](0) \quad = \quad \lim_{t \to 0} \frac{F^{(i-1)}[n, x](t) - F^{(i-1)}[n, x](0)}{t},$$

$$= \lim_{t \to 0} \frac{F^{(i-1)}[n,x](t)}{t},$$

since $F^{(i-1)}[n,x](0)$ is zero by the inductive hypothesis. Taking the $i-1$st derivative of equation 5.3 yields

$$\lambda 2^{i-1} F^{(i-1)}[n,x](t) = F^{(i-1)}[n,x](\frac{t}{2}).$$

The $(i-1)$st derivatives exists by our inductive hypothesis. Now, if $|\lambda| < \frac{1}{2^k}$ and $i \leq k$, then $|\lambda| 2^{i-1} < \frac{1}{2}$. Therefore, by part one of lemma 1, the limit as t goes to zero of $\frac{F^{(i-1)}[n,x](t)}{t}$ must also be zero.

Given that the ith derivative of $F[n,x](t)$ exists at $t = 0$, we can take the ith derivative of equation 5.3. Applying part one of lemma 1 shows that $F^{(i)}[n,x](t)$ is continuous at $t = 0$. This step completes the induction and the proof. □

This theorem also holds for generalized eigenvectors (equation 5.2) whose eigenvalues have moduli less than $\frac{1}{2^k}$. The proof involves modifying part one of lemma 1 to use equation 5.4.

Together theorems 13 and 14 yield necessary conditions for a subdivision scheme to produce C^k continuous limit functions. If S necessarily produces C^k limit curves and there exists p such that the ith derivative of $F[n,p](0) \neq 0$ for all $0 \leq i \leq k$, then there must exist eigenvectors of S that reproduce each monomial t^i up to degree k. So, the subdivision scheme defined by S can reproduce any polynomial up to degree k. The spectrum of S includes 1, $\frac{1}{2}$, ..., $\frac{1}{2^k}$. If the basis functions associated with the scheme are linearly independent, then these eigenvalues must also be unique.

Under these assumptions, we can summarize the necessary conditions for a C^k continuous scheme. Indexing the eigenvalues of S in order of decreasing moduli, $|\lambda_i| \geq |\lambda_{i+1}|$,

1. $F[n,x_i](t) = c_i t^i$ for all $0 \leq i \leq k$,

2. $\lambda_i = \frac{1}{2^i}$ for all $0 \leq i \leq k$,

3. $|\lambda_i| < \frac{1}{2^k}$ for all $i > k$.

If a full span of derivatives at the origin does not exist, then the spectrum may be missing certain powers of two. If the scheme has linearly dependent basis functions, several powers of two may be repeated. This result is a generalization of a similar result for uniform subdivision in [CDM91].

Since the regular four point rule produces C^1 functions, we can now use theorem 14 to show that the irregular four point rule also produces C^1 functions. For $c > 0$, consider the knot sequence n

$$..., -3, -2, -1, 0, c, 2c, 3c, ...$$

where $n_i = i$ is $i \leq 0$ and $n_i = ci$ if $i \geq 0$. Then the matrix $\bar{S}$ is

$$\begin{pmatrix} -\frac{1}{16} & \frac{9}{16} & \frac{9}{16} & -\frac{1}{16} & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & \frac{-(1+2c)}{16(2+c)} & \frac{3(1+2c)}{8(1+c)} & \frac{3}{8}+\frac{3}{16c} & \frac{-3}{8c(2+3c+c^2)} & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & \frac{-3c^3}{8(1+3c+2c^2)} & \frac{3(2+c)}{16} & \frac{3(2+c)}{8(1+c)} & \frac{-(2+c)}{16(1+2c)} & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & -\frac{1}{16} & \frac{9}{16} & \frac{9}{16} & -\frac{1}{16} \end{pmatrix}.$$

The matrix, parameterized by c, has eigenvalues that are independent of c.

$$1, \frac{1}{2}, \frac{1}{4}, \frac{1}{4}, \frac{1}{8}, \frac{-1}{16}, \frac{-1}{16}.$$

The leading eigenvalues are 1 and $\frac{1}{2}$ while the remaining eigenvalues have moduli less that $\frac{1}{2}$. Since two leading eigenvectors reproduce 1 and t (by construction), this scheme produces C^1 limit curves.

For interpolatory schemes with higher order polynomial precision, this analysis becomes more difficult. For quintic precision ($r = 2$), Mathematica can be used to derive the characteristic polynomial of $\bar{S}$. This polynomial and its roots vary as a function of c. By representing the polynomial in various bases, one can show that for $c > 0$, the spectrum of $\bar{S}$ satisfies the C^2 conditions. Unfortunately, this type of analysis fails for $r > 2$. We conjecture that such interpolatory schemes with order $2r + 1$ polynomial precision are C^r continuous.

5.4 Derivative schemes

Theorem 14 relies on knowing the subdivision scheme is C^k continuous away from the origin. In this section and the next, we give direct conditions on S for the scheme to be C^k continuous everywhere. To start, we must first establish that S defines a continuous scheme. The analysis technique of section 3.3 can be used to verify the continuity of the scheme. Recall that we defined a difference scheme associated with S and show that the difference scheme uniformly converges to zero. Establishing the smoothness of the scheme is more difficult. Since the knot sequence is no longer regularly spaced, theorem 3 cannot be used. The following provides a method for analyzing the smoothness of an irregular subdivision scheme.

5.4.1 Linear parameterizations

We previously showed that any stationary subdivision scheme that is at least C^1 continuous and that can produce functions with non-zero first derivatives at the origin must have linear precision. In particular, the eigenvector x_1 associated with eigenvalue $\frac{1}{2}$ must produce a constant multiple of the function t. Let l denote the multiple of x_1 such that

$$F[n,l](t) = t.$$

To simplify the following analysis, we restrict our attention to a fairly specific class of subdivision schemes, those in which the function t has a locally unique representation l in terms of coefficients that form a monotonically increasing sequence. For example, this restriction rules out schemes that produce linearly dependent basis functions. (However, a similar, but more complex theory is possible if t has multiple representations.)

Our functional scheme can also be expressed as the parametric curve

$$(t, F[n,p](t)) = (F[n,l](t), F[n,p](t)).$$

Applying the subdivision matrix S repeatedly to the initial parametric control polygon $(l,p)^T$ produces the appropriate functional limit curve. This observation suggests that the spacing given by $S^j l$ might be a more appropriate *linear* parameterization for the piecewise linear function defined by p^j. After j subdivisions, this linear parameterization satisfies

$$S^j l = \frac{l}{2^j} = l^j.$$

Note that the knot sequence n^j is still used to define the subdivision matrices $S[n^j]$. Only the parameter values associated with the p^j are changed.

The following example highlights the main reason for using this parameterization. Consider cubic B-splines over the initial knot sequence n

$$\ldots, -9, -6, -3, 0, 6, 12, 18, \ldots$$

Applying midpoint subdivision to this knot sequence leads to a stationary subdivision scheme with matrix $\bar{S}$

$$\begin{pmatrix} \frac{1}{8} & \frac{25}{32} & \frac{3}{32} & 0 & 0 \\ 0 & \frac{5}{8} & \frac{3}{8} & 0 & 0 \\ 0 & \frac{5}{24} & \frac{29}{40} & \frac{1}{15} & 0 \\ 0 & 0 & \frac{3}{5} & \frac{2}{5} & 0 \\ 0 & 0 & \frac{3}{20} & \frac{29}{40} & \frac{1}{8} \end{pmatrix}.$$

The derivative of $F[n,p](t)$ is a C^2 function. The derivatives of $L[n^j, S^j p](t)$ (defined using the knot parameterization) do not converge to a C^1 function. The eigenvector l for this scheme is

$$..., -9, -6, -3, 1, 6, 12, 18, ...$$

The derivatives of $L[l^j, S^j p](t)$ (defined using the linear parameterization) do converge to a C^1 function. In fact this function is the derivative of $F[l,p](t)$. In summary, only the linear parameterization ensures that the limit of the derivatives of control polygons is the derivative of the final limit curve. Section 5.4.3 proves this result.

Parameterizing by either n or l produce the same results as long as the convergence is uniform.

Theorem 15 *Let $||p||$ be bounded. If $L[n^j, S^j p](t)$ uniformly converges to a continuous function as $j \to \infty$, then*

$$F[n,p](t) = F[l,p](t)$$

Proof: We first show that the difference between n_i and l_i is bounded for all i. If n is regularly spaced, then $S[n]$ must be uniform due to affine and index invariance. Since the subdivision rules are locally defined and n has a single non-uniformity, only a finite number of columns of $S[n]$ can deviate from the uniform case. For the uniform case, [CDM91] and [DGL94] show that the polynomial t can be reproduced by a vector that is a sample of a linear function over a regular knot sequence. Since this representation is unique (as assumed at the start of the section), l_i must agree with the uniform analysis for large $|i|$. Therefore, there must exist k such $n_i = l_i + c_0$ for all $i \leq -k$ and $n_i = l_i + c_1$ for all $i \geq k$. So, the difference $||n - l||$ must be bounded.

The ith entry of p is plotted at either n_i or l_i depending on the parameterization chosen. The maximum deviation between the two functions $L[n,p](t)$ and $L[l,p](t)$ at $t = l_i$ is bounded by the number of knots in n between n_i and l_i and the maximum change in $L[n,p](t)$

between a pair of knots. If d is the maximum number of knots between l_i and n_i over all i (d exists since $||n - l||$ is bounded), then this difference is bounded by $d||\Delta_u p||$. Repeating this argument after j subdivisions,

$$\begin{aligned} ||L[n^j, S^j p](t) - L[l^j, S^j p](t)|| &\leq d||\Delta_u S^j p||, \\ &\leq d||(D_u)^j \Delta_u p||, \\ &\leq d||(D_u)^j|| \cdot ||\Delta_u p||. \end{aligned}$$

Since $||p||$ is bounded, $||\Delta_u p||$ must be bounded. By the converse of theorem 6, $||(D_u)^j||$ converges to zero as $j \to \infty$. Therefore, the difference between these two function converges to zero and the two functions uniformly converge to the same limit. □

5.4.2 Non-uniform differencing operator

Our ultimate goal is to build a subdivision matrix D whose limit functions are the derivatives of limit functions produced by S. To this end, we use an approach similar to that of Gregory and Qu in [GQ92, GR94]. The key is to construct a differencing operator Δ that behaves like a discrete derivative. Δ maps a set of control points for S to a new set of control points for D. S, D, and Δ satisfy a commutative relation. Differencing a set of subdivided control points, Sp, should produce the same results as subdividing (using D) a set of differenced control points, Δp. In terms of Δ, S, and D,

$$D\Delta = (2\Delta)S. \tag{5.7}$$

The cause of the extra factor of 2 will become clear momentarily.

If Δ is to act as a derivative, Δ should annihilate the vector $\mathbf{1}$ since the derivative of the function one is zero. Applied before any subdivision, Δ should map l to the vector $\mathbf{1}$ since the derivative of t is one. Thus, we want

$$\begin{aligned} \Delta\mathbf{1} &= \mathbf{0}, \\ \Delta l &= \mathbf{1}. \end{aligned}$$

Let Δ_u denote the uniform differencing operator of equation 3.2. If we let Γ be the diagonal matrix whose (i,i)th entry is $1/(l_{i+1} - l_i)$, then Δ can be expressed as

$$\Delta = \Gamma\Delta_u.$$

Δ_u annihilates the vector $\mathbf{1}$ and Γ scales $\Delta_u l$ to yield $\mathbf{1}$.

Note that Δ behaves correctly only if appled before subdivision. After j rounds of subdivision, the function t is represented by the control points $\frac{l}{2^j}$. Therefore, the correct differencing operator in this case is $2^j\Delta$. This observation explains the factor of two in equation 5.7.

Given Δ, we can explicitly construct D. Let

$$\Delta^* = \Delta_u^* \Gamma^{-1}.$$

Δ^* is a left and right inverse of Δ. Now, the matrix D satisfies

$$D = 2\Delta S \Delta^*.$$

If S has column height $a+b+1$, then it is easy to show D has column height $a+b$.

5.4.3 Derivative schemes

D defines a stationary subdivision scheme with associated limit functions

$$\lim_{j\to\infty} L[l^j, D^j q](t).$$

It remains to show that these functions are derivatives of those produced by S.

Theorem 16 *Let the sequence $L[l^j, D^j q](t)$ uniformly converge to a continuous function for all bounded $||q||$. Then, for all bounded $||p||$,*

$$\lim_{j\to\infty} L[l^j, D^j(\Delta p)](t)$$

is the derivative of $F[l,p](t)$ with respect to t.

Proof: Recall that by definition $F[l,p](t) = g(t)$ is the limit of the functions

$$g_j(t) = L[l^j, S^j p](t).$$

The derivatives, $g_j'(t)$, of the $g_j(t)$ are piecewise constant functions over the knot sequence l^j with piecewise values $(2^j\Delta)S^j p$. By the construction of D,

$$(2^j\Delta)S^j = D^j\Delta.$$

So, the $g_j'(t)$ have piecewise values $D^j(\Delta p)$.

By hypothesis, the sequence

$$h_j(t) = L[l^j, D^j(\Delta p)](t)$$

uniformly converges to a continuous function, call it $h(t)$. By the converse of theorem 5, the difference between the function $g_j'(t)$ and $h_j(t)$ must uniformly converge to zero. Therefore, the sequence $g_j'(t)$ uniformly converges to $h(t)$. Figure 5.1 illustrates this process.

We next prove that $\int_0^t g_j'(s)ds$ is point-wise convergent to $\int_0^t h(s)ds$ for any t. Fix t. By the uniform convergence of the $g_j'(t)$ to $h(t)$, for all $\epsilon > 0$, there exist an n such that for all $j \geq n$

$$|g_j'(t) - h(t)| < \epsilon.$$

In terms of the integrals, for all $j \geq n$,

$$\begin{aligned} |\int_0^t g_j'(s)ds - \int_0^t h(s)ds| &\leq \int_0^t |g_j'(s) - h(s)|ds, \\ &\leq \epsilon t. \end{aligned}$$

Therefore, $g_j(t) - g_j(0)$ converges to $\int_0^t h(s)ds$ for any t. However, $g_j(t) - g_j(0)$ also converges to $g(t) - g(0)$. So,

$$g(t) = \int_0^t h(s)ds + g(0)$$

for any t. Thus, $h(t)$ is the derivative of $g(t)$. □

This theorem is a slight variant of theorem IV on page 599 of [Tay55]. Figure 5.2 shows the derivative curve associated with the irregular four point curve of figure 4.2. The control points for this curve were computed from the original control points by applying Δ.

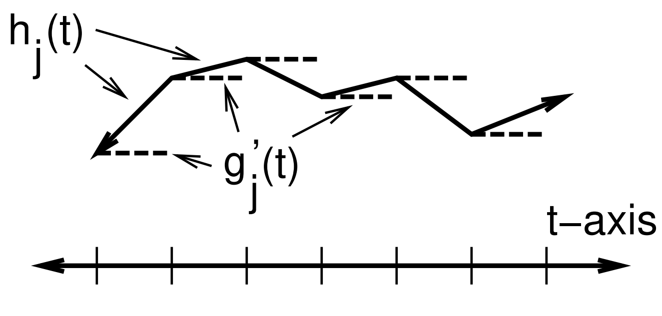

Figure 5.1 Convergence of $g_j'(t)$ and $h_j(t)$

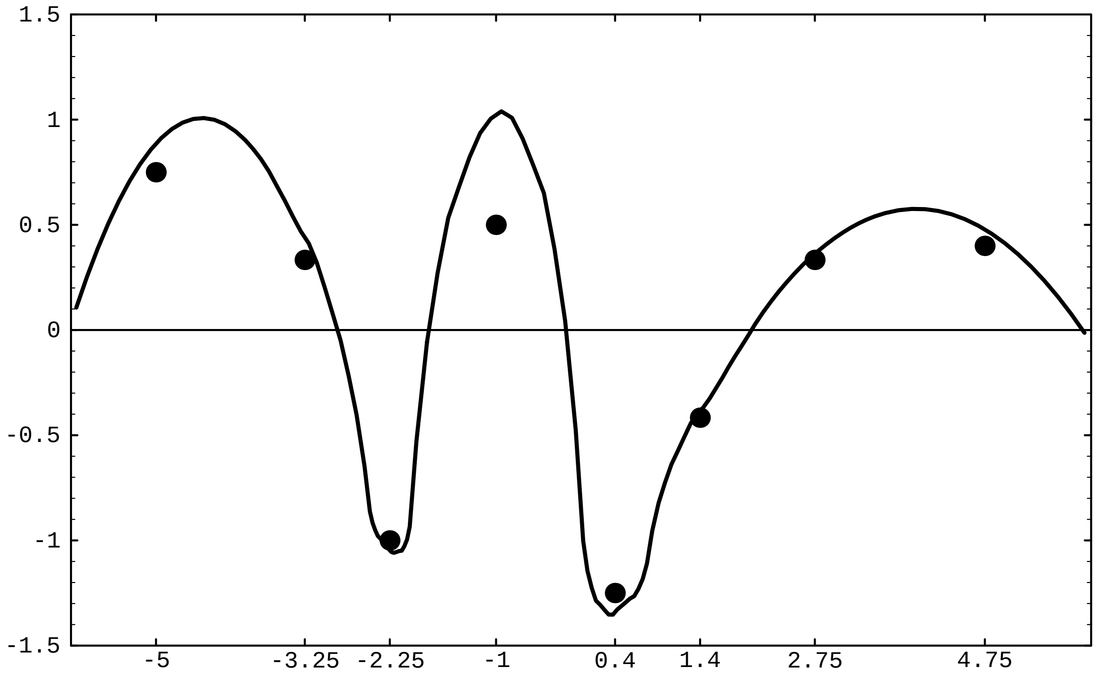

Figure 5.2 The derivative of an irregular scheme

The previous analysis can be repeated with D in place of S. Since D is a stationary subdivision scheme, theorem 5 can determine whether the new scheme is uniformly convergent to a continuous function. If D satisfies the necessary conditions for a C^1 scheme, then a new linear parameterization for its limit curves can be computed. If $F[n, x_2](t) = \frac{1}{2}t^2$ for S, then Δx_2 is the correct linear parameterization for D. A new non-uniform difference operator can be defined with this parameterization. Finally, theorem 16 can be applied to build a derivative scheme for D.

5.5 Parametric analysis

Given a control polyon $(p_1, p_2)^T$, we can apply the stationary subdivision matrix S to each component separately and build a sequence of polygons $(S^j p_1, S^j p_2)^T$. When does a limit polygon exist and when is it smooth?

The spectral properties of S can be used to give conditions for smoothness. Given subdivision matrix S, we can compute in descending order its eigenvalues $1, \gamma, ...$ and their associated eigenvectors $x_0, x_1, ...,$. If $0 < \gamma < 1$, then the eigenvector $l = x_1$ provides the natural parameterization for this subdivision scheme. Note that γ need not be $\frac{1}{2}$. $\frac{1}{2}$ arose in the preceeding functional analysis due to our use of binary subdivision. The only restriction placed on l is that

$$l_{i+1} - l_i \geq \epsilon \geq 0$$

for all i. This restriction guarantees that the parameterization is $1-1$ and fills the parameter line. (Here the ordering of the control points determines the ordering of the rows and columns of S and thus, the ordering of the entries of l.)

Given a parameterization for our control polygon, we now can apply the functional analysis of the previous section to each component of our parametric curve (with $\frac{1}{2}$ replaced by γ). If each component is a C^k function, then the parametric curve

$$(F[l,p_1](t), F[l,p_2](t))$$

is a C^k curve for those t such the derivative of $F[l,p_1](t)$ or $F[l,p_2](t)$is non-zero.

Chapter 6

Multi-variate subdivision over regular grids

We next turn our attention to the problem of generalizing subdivision from the univariate domain to the multi-variate domain. As in the univariate case, we first investigate subdivision methods over regular (uniform) grids. Given a set of coefficients p^0 attached to each vertex of an initial grid T^0, a multi-variate subdivision should produce a sequence of coefficient sets p^j attached to finer and finer grids T^j. As in the univariate case, we defer the formal presentation of these details to subsequent chapters. Instead, we concentrate on a simple generalization of uniform B-splines to the multi-variate setting.

6.1 B-splines as cross-sectional volumes

Univariate B-splines were defined through convolution. However, there are a variety of alternative definitions. One particularly nice definition involves taking cross sections of high dimensional boxes.

Let B be a point set in R^n with coordinates $x = (x_1, ..., x_n)$. We wish to construct a function $N(t)$ in d-dimensional parameter domain using B. Consider the restriction of B to the d-dimensional subspace of R^n such that the first d coordinates of x agree with $t = (t_1, ..., t_d)$. If we graph the d-dimensional volume of this restriction as a function of t, a function $N(t)$ results.

$$N(t) = volume_d(B \cap (x_{1..d} = t_{1...d})).$$

Partitioning B into several disjoint pieces and applying this construction yields several new functions whose sum is $N(t)$. If this partition is into pieces that are scaled translates of

B, then the cross sectional areas of these pieces are just translates and dilates of $N(t)$. The partition of B yields a subdivision formula for $N(t)$.

In figure 6.1, B is square ($n = 2$). The cross sectional area of B in its first coordinate is the piecewise linear hat function. Subdividing this square into 4 subsquares partitions this function into three distinct translates and dilates. Note that the middle pair of square project onto the same central basis function. This pattern is reflected in the subdivision formula

$$N(t) = \frac{1}{2}N(2t) + N(2t-1) + \frac{1}{2}N(2t-2)$$

is exactly the subdivision formula for a uniform quadratic B-spline.

In figure 6.2, B is a cube ($n = 3$). Graphing the cross sectional area in the first coordinate yields a different function. Subdividing B into eight subcubes and projecting yields four distinct basis functions. Three cubes project onto the second basis function. Three cubes project onto the third basis function. This relation yields the subdivision formula

$$N(t) = \frac{1}{4}N(2t) + \frac{3}{4}N(2t-1) + \frac{3}{4}N(2t-2) + \frac{1}{4}N(2t-3)).$$

This subdivision formula is exactly that of cubic B-splines.

These observation gives us a general prescription for creating subdividable basis functions in d dimensions. Given a set of directions D in R^d, define a box in $R^{|D|}$ whose edges project

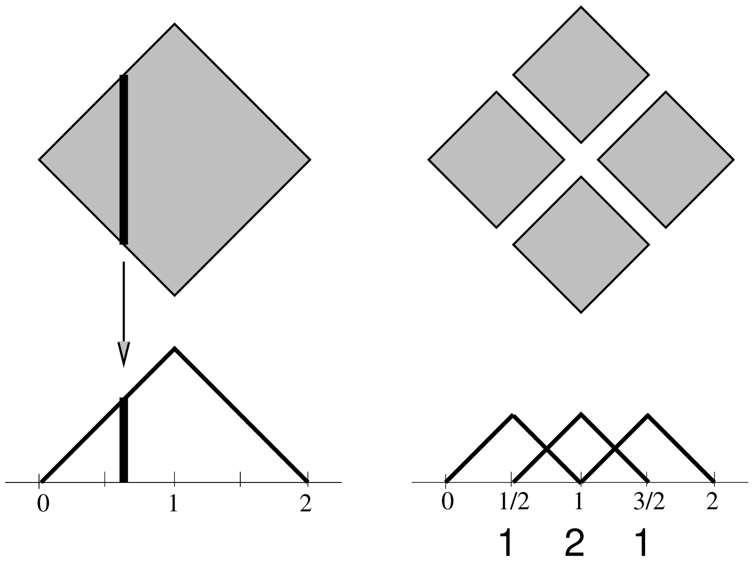

Figure 6.1 The cross sectional area of a square

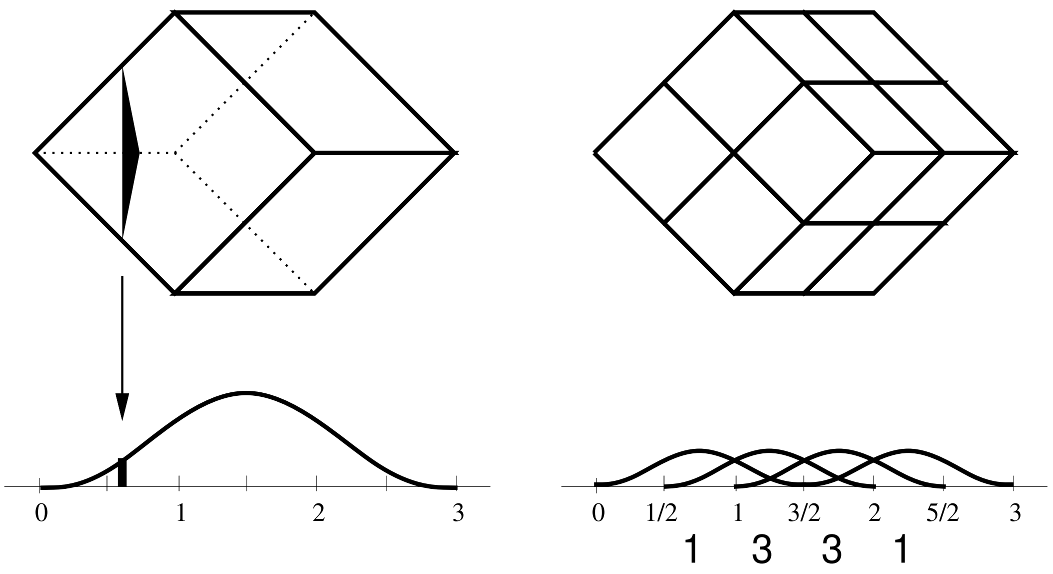

Figure 6.2 The cross sectional area of a cube

into R^d along the directions specified by D. The graph of the d-dimensional cross-sectional area is a function that has a subdivision formula. Linear combinations of translates of this basis function yields a type of spline known as a *box* spline.

6.2 Box splines

The basis functions for box splines are determined by a set of *direction* vectors D. Each entry δ of D is a d-tuple whose entries are integers. The simplest box spline has a set of direction vectors D consisting of each of the unit direction $\{e_1, ..., e_d\}$. The basis function for this set of direction vectors is

$$\begin{aligned} N_D(t) &= 1 \text{ if } 0 \le t_i < 1 \ \forall i, \\ &= 0 \quad \text{otherwise.} \end{aligned}$$

This functions is the d-variate generalization of the univariate step function $U(t)$.

Like the univariate step function, the multi-variate step function has a subdivision formula,

$$N_D(t) = \sum_{i_1=0}^{1} ... \sum_{i_d=0}^{1} N_D(2t - \mathbf{i})$$

where $\mathbf{i} = (i_1, ..., i_d)$. As in the univariate case, the coefficients of this subdivision formula defined a generating function. If $S(z)$ is the generating function associated with this subdivision formula, then $S(z)$ satisfies

$$S(z) = \prod_{j=1}^{d} (1 + z_j). \tag{6.1}$$

For larger sets of direction vectors, adding a new direction vector δ to an existing set of direction vectors D yields a new basis function that satisfies

$$N_{D \cup \delta}(t) = \int_0^1 N_D(t - u\delta)du. \tag{6.2}$$

If $N_D(t)$ has a subdivision formula, then $N_{D \cup \{\delta\}}(t)$ has a subdivision formula. If $z^\delta = \prod_{j=1}^{d} z_j^{\delta_j}$, then the next theorem relates the generating functions for these two formulas.

Theorem 17 *Let $f(t)$ have a subdivision formula with associated generating function $S(z)$. If*

$$g(t) = \int_0^1 f(t - u\delta)du,$$

then $g(t)$ has the subdivision formula with generating function

$$\frac{1}{2} S(z)(1 + z^\delta).$$

Proof: By hypothesis,

$$\begin{aligned}
g(t) &= \int_0^1 f(t - u\delta)du, \\
&= \int_0^1 (\sum_{\mathbf{i}} s_{\mathbf{i}} f(2(t - u\delta) - \mathbf{i}))du, \\
&= \frac{1}{2} \int_0^2 (\sum_{\mathbf{i}} s_{\mathbf{i}} f(2t - u\delta - \mathbf{i}))du, \\
&= \frac{1}{2} \int_0^1 (\sum_{\mathbf{i}} s_{\mathbf{i}} f(2t - u\delta - \mathbf{i})) + (\sum_{\mathbf{i}} s_{\mathbf{i}} f(2t - u\delta - \mathbf{i} - \delta))du, \\
&= \frac{1}{2} \int_0^1 (\sum_{\mathbf{i}} (s_{\mathbf{i}} + s_{\mathbf{i}-\delta}) f(2t - u\delta - \mathbf{i}))du, \\
&= \frac{1}{2} \sum_{\mathbf{i}} (\sum_{\mathbf{i}} (s_{\mathbf{i}} + s_{\mathbf{i}-\delta}) \int_0^1 f(2t - u\delta - \mathbf{i}))du, \\
&= \frac{1}{2} \sum_{\mathbf{i}} (s_{\mathbf{i}} + s_{\mathbf{i}-\delta}) g(2t - \mathbf{i}).
\end{aligned}$$

The generating function associated with this subdivision formula is exactly $\frac{1}{2}S(z)(1 + z^\delta)$. $\square$

Let $N_D(t)$ have the subdivision formula

$$N_D(t) = \sum_{\mathbf{i}} s_{\mathbf{i}} N_D(2t - \mathbf{i}).$$

In the base case where $D = \{e_1, ..., e_d\}$, the generating function $S_D(z)$ for this formula satisfies

$$S_D(z) = 2^d \prod_{j=1}^{d} \frac{1}{2}(1 + z^{e_j}).$$

Recursively applying theorem 17 to larger sets of direction vectors, we note that the generating function is

$$S_D(z) = 2^d \prod_{\delta \in D} \frac{1}{2}(1 + z^{\delta}).$$

The generating function $S_D(z)$ is indepedent of the order in which the direction vectors in D are processed. Therefore, the subdivision formula for $N_D(t)$ is independent of the order in which its defining direction vectors were processed. Since this subdivision formula uniquely determines the function $N_D(t)$, the definition of box splines using equation 6.2 is order independent.

This expression describes how box spline basis functions subdivide. A box spline is a linear combination of translates of these basis functions

$$f(t) = \sum_{\mathbf{i}} p_{\mathbf{i}} N_D(t - \mathbf{i}).$$

Subdividing the basis functions $N_D(t)$ allows the function $f(t)$ to be expressed in terms of denser and denser sets of coefficients. The subdivision process for the coefficients can expressed very compactly using generating functions. If $P^0(z)$ is the generating function

$$P^0(z) = \sum_{\mathbf{i}} p_{\mathbf{i}} z^{\mathbf{i}},$$

then the generating functions $P^j(z)$ for successive sets of coefficients can be generated as follows:

$$P^{j+1}(z) = S_D(z) P^j(z^2).$$

Here, z^2 denotes the vector $(z_1^2, ..., z_d^2)$.

Box spline subdivision can be interpreted in a manner similar to that of the Lane-Reisenfeld algorithm for univariate B-splines. The generating function $S_D(z)$ consists of two parts, $\prod_{j=1}^{d}(1 + z^{e_j})$ and subsequent factors introduced by convolution. The actions of these parts are as follows:

Replicate Multiplying the factor $\prod_{j=1}^{d}(1 + z^{e_j})$ and $P^j(z^2)$ yields a generating function where coefficients in $P^j(z)$ are replicated 2^d times.

Average Multiplying each of the remaining factors $\frac{1}{2}(1 + z^{\delta})$ times this expression averages the replicated coefficients in the direction δ.

Section 6.4 gives several examples of this subdivision algorithms in practice.

6.3 Properties of box splines

Before showing several examples of box-splines, we first describe some basis properties of box splines. We begin with a characterization of the degree and the support of box spline basis function.

Theorem 18 *The basis function $N_D(t)$ is a piecewise polynomial function of degree $|D| - d$ with support*

$$\{\sum_{\delta_i \in D} \alpha_i \delta_i | 0 \leq \alpha_i \leq 1\}. \tag{6.3}$$

Proof: If $|D| = d$, then the box spline has direction vector $D = \{e_1, ..., e_d\}$. By definition, the associated basis function is just a piecewise constant function supported over the unit box. Of course, this is also the region defined in equation 6.3

Given a set of direction vectors D, we next explore the effect of adding a direction vector δ to D. By equation 6.2, the basis function $N_{D \cup \{\delta\}}(t)$ is the integral of $N_D(s)$ as s varies from t to $t - \delta$. If $N_D(t)$ has degree $|D| - d$, then $N_{D \cup \{\delta\}}(t)$ has degree $|D| + 1 - d$. Moreover, if $N_D(t)$ has supported as in equation 6.3, then the new basis function is also supported over any point t in this region plus its translates of the form $t + \alpha\delta$ where $0 \leq \alpha \leq 1$. □

The actual extent of each polynomial piece of the basis function can characterized using a finer analysis. Extend each direction vector into an infinite line and take all integer translates of these lines. The corresponding box spline basis function is polynomial over the resulting partition of the plane. The next section shows the supports for various types of box splines and their partition into polynomial pieces.

In the univariate case, the smoothness of the B-spline increased with each convolution. Unfortunately, this result does not follow in the multi-variate case. Adding a new direction vector δ increases the smoothness of the new basis function in the direction δ. However, the smoothness is other directions is not necessarily increased. The next theorem characterizes the smoothness of box spline basis functions.

Theorem 19 *Let $\hat{D} \subseteq D$ be a maximal set of direction vectors that do not span R^d. Then, the basis function $N_D(t)$ is a C^k function where*

$$k = |D| - |\hat{D}| - 2.$$

For a proof, we refer the interested reader to [dBHR93]. If every set of d vector in D are linearly independent , then $|\hat{D}| = d-1$ and $N_D(t)$ has smoothness of order $|D| - d - 1$. Note that this smoothness is maximal since the basis function is a piecewise polynomial of degree $|D| - d$. Only if a subset of $d-1$ direction vectors are linearly dependent (e.g. a repeated direction vector) does $N_D(t)$ fail to have maximal smoothness.

6.4 Examples

Our first example consists of the three direction vectors

$$D = \{(1,0), (0,1), (1,1)\}.$$

By definition, two direction vectors yields a piecewise constant function. Three direction vectors yield a piecewise linear basis function. The generating function $S(z)$ for the subdividing this basis function is

$$S(z) = \frac{1}{2}(1+z_1)(1+z_2)(1+z_1 z_2).$$

Figure 6.3 shows the support of a single hat function. The bold lines separate the polynomial pieces of the function. The thin lines are the similar partition after subdivision. Each label is the coefficient of one term of the subdivision formula normalized to be an integer. The label is placed over the corresponding basis function.

This subdivision formula can be interpreted as an instance of a replicate and average. Consider an initial set of coefficients

$$\begin{array}{cccc} \dots & \dots & \dots & \dots \\ \dots & p & q & \dots \\ \dots & r & s & \dots \\ \dots & \dots & \dots & \dots \end{array}$$

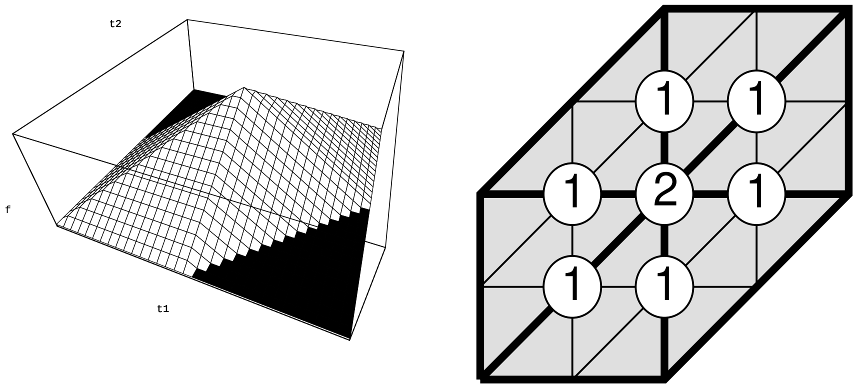

Figure 6.3 Piecewise linear hat functioun

After replicating each coefficient

$$\begin{array}{cccccc}
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
\cdots & p & p & q & q & \cdots \\
\cdots & p & p & q & q & \cdots \\
\cdots & r & r & s & s & \cdots \\
\cdots & r & r & s & s & \cdots \\
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots
\end{array}$$

Finally, averaging in the $(1,1)$ direction yields

$$\begin{array}{ccccc}
\cdots & \cdots & \cdots & \cdots & \cdots \\
\cdots & p & \frac{p+q}{2} & q & \cdots \\
\cdots & \frac{r+p}{2} & \frac{r+q}{2} & \frac{s+q}{2} & \cdots \\
\cdots & r & \frac{r+s}{2} & s & \cdots \\
\cdots & \cdots & \cdots & \cdots & \cdots
\end{array}$$

Our second example has four direction vectors

$$D = \{(1,0),(0,1),(1,1),(1,-1)\}.$$

This basis function is piecewise quadratic. Since any pair of vectors is linearly independent, the basis function is C^1 continuous. The generating function $S(z)$ for the subdividing this basis function is

$$S(z) = \frac{1}{4}(1+z_1)(1+z_2)(1+z_1z_2)(1+z_1z_2^{-1}).$$

Figure 6.4 shows support of a single basis function and its partition into polynomial pieces. The labels reflect the subdivision formula.

Our third example has six direction vectors

$$D = \{(1,0),(0,1),(1,1),(1,0),(0,1),(1,1)\}.$$

(Note that each direction vector is repeated twice.) This basis function is piecewise quartic. In this case $\hat{D}$ has a maximal size of two. Therefore, the basis function is C^2 continuous. The generating function $S(z)$ for the subdividing this basis function is

$$S(z) = \frac{1}{16}(1+z_1)^2(1+z_2)^2(1+z_1z_2)^2.$$

Figure 6.5 shows support of a single basis function and its partition into polynomial pieces. The labels reflect the subdivision formula.

One standard way to build bivariate basis functions that can be subdivided is take the tensor product of two univariate basis function with subdivision formulas. For example, if $N_1(t_1) = \sum_{i_1} s^1_{i_1} N_1(2t_1 - i_1)$ and $N_2(t_2) = \sum_{i_2} s^2_{i_2} N_2(2t_2 - i_2)$, then the tensor product of these two basis function had the subdivision formula

$$\begin{aligned} N(t_1,t_2) &= N_1(t_1)N_2(t_2), \\ &= \sum_{i_1,i_2} s^1_{i_1} s^2_{i_2} N_1(2t_1 - i_1) N_2(2t_2 - i_2), \\ &= \sum_{i_1,i_2} s^1_{i_1} s^2_{i_2} N(2t_1 - i_1, 2t_2 - i_2). \end{aligned}$$

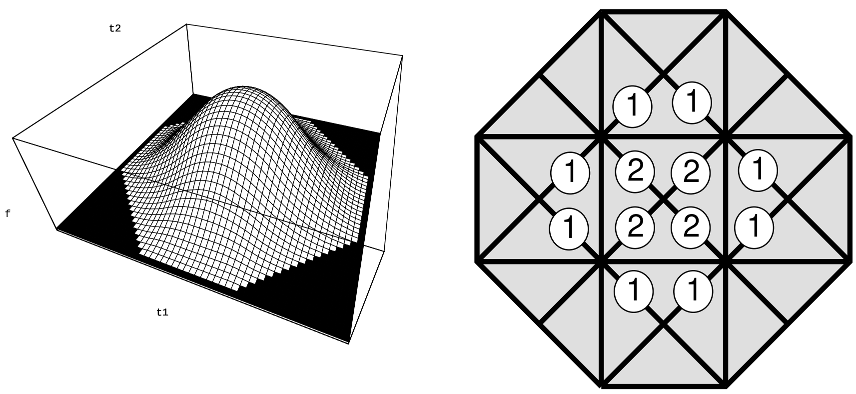

Figure 6.4 C^1 four direction quadratic box spline

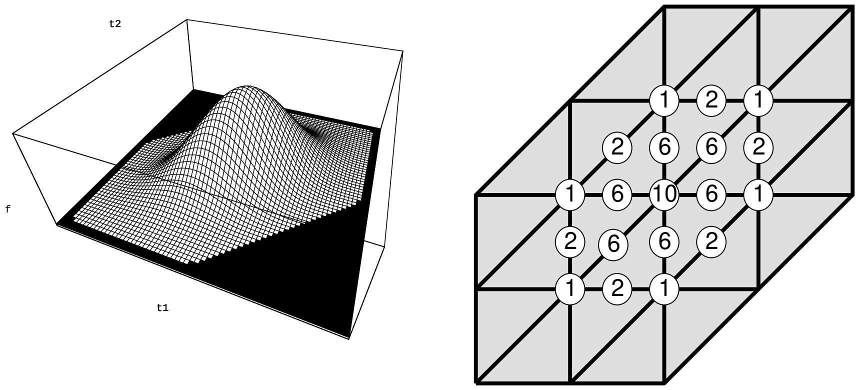

Figure 6.5 C^2 three direction quartic box spline

Tensor product B-splines can be expressed as box splines. For example, the set of six direction vectors

$$D = \{(1,0),(0,1),(1,0),(0,1),(1,0),(0,1)\}$$

yields a box spline basis function that are total degree quartic. The generating function associated with this box spline is

$$S(z) = \frac{1}{16}(1+z_1)^3(1+z_2)^3.$$

Factoring this generating function into $\frac{1}{4}(1+z_1)^3$ and $\frac{1}{4}(1+z_2)^3$ illustrates that the resulting function is actually the tensor product of two quadratic B-splines. Therefore, the basis functions are C^1 continuous. (Note this agrees with theorem 19 of the previous section.) Figure 6.6 shows the actual subdivision formula and the support of the resulting basis function.

Summary Box splines are a very natural extension of the uniform B-splines to the multi-variate setting. Of course, other types of multi-variate subdivision processes are possible besides box splines. In the case of box splines, the generating function $S(z)$ always split into linear factors. If $S(z)$ does not factor, what properties does the subdivision scheme possess? [Dyn92] gives a nice analysis of this case using matrix subdivision.

If one wishes to derive the box spline basis functions are the limit of the subdivision process, then the coefficients are assigned parameter values over a regular grid. Are subdivision methods over irregular grids possible? The next chapter attempts to answer this question.

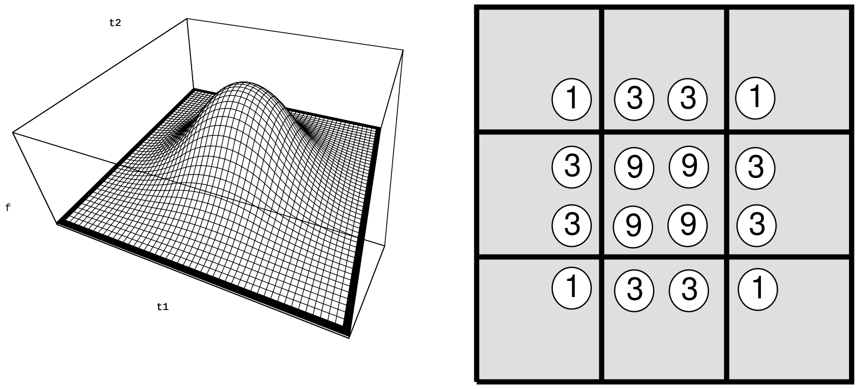

Figure 6.6 Bi-quadratic B-spline

Chapter 7

Subdivision over irregular triangulations

7.1 Bivariate subdivision schemes

A bivariate subdivision scheme can be viewed as being driven by a sequences of triangulations T^j. Given an triangulation T, let $D(T)$ denote the new triangulation in which each triangle of T is subdivided into four similar copies of itself. Figure 7.1 illustrates this process. Given an initial triangulation $T = T^0$, we define a sequence of triangulations T^j as follows:

$$T^{j+1} = D(T^j).$$

A triangular subdivision scheme is map from triangulations T to subdivision matrices $S[T]$. $S[T]$ maps a set of control points associated with the vertices of T to a new set of control points associated with the vertices of $D(T)$. Given an initial set of control points

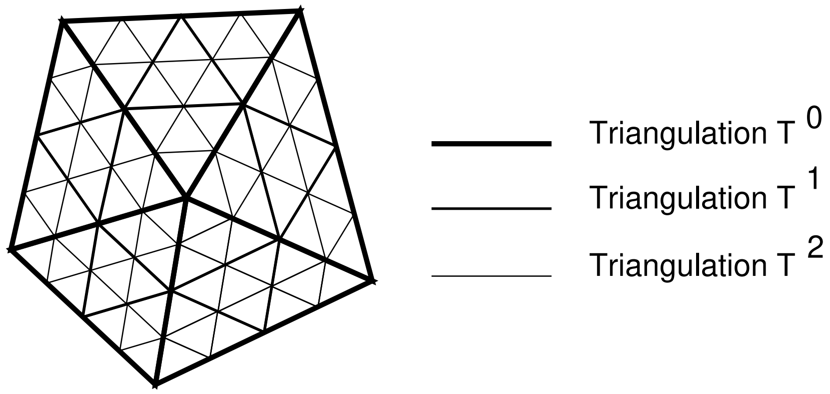

Figure 7.1 Regular subdivision of an irregular grid

$p = p^0$, we defined the sequence of new control points

$$p^{j+1} = S[T^j]p^j. \tag{7.1}$$

The m-disc of a vertex v in T is the set of all points in T that lie in a triangle at most $m-1$ vertex adjacent triangles away from v. Figure 7.2 shows the black vertex v (its own 0-disc). Each colored region is the difference between a disc and the next inner disc. We restrict our attention to subdivision schemes that produce matrices satisfying the following three properties.

Compact support There exists m such that for all k the support of the kth column of $S[T]$ lies in the m-disc in $D(T)$ of vertex n_k (in T). Compact support ensures that the infinite sum in equation 7.1 is bounded.

Affine invariance For any non-degenerate affine transformation A,

$$S[T] = S[A(T)].$$

If the triangulation T is *self-similar* under subdivision (i.e. $D(T) = A(T)$), then the resulting subdivision scheme is *stationary*, that is $S[T] = S[A(T)] = S[D(T)] = S[D^j(T)]$ for all j.

Local definition The kth column of $S[T]$ depends only on a bounded disc of T centered at n_k. Local definition is critical in showing that the resulting scheme is locally stationary.

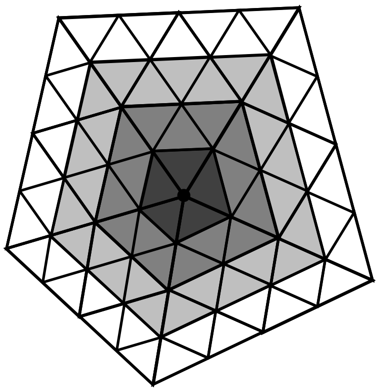

Figure 7.2 Various discs of a vertex in a triangulation

Although we will not use this property, we also assume that the subdivision scheme is *index invariant*, that is renumbering the vertices of T does not affect the subdivision process. Together with affine invariance, this property is sufficient to show the subdivision scheme is uniform on regular triangulations.

The vertices, n, of the triangulation T are the analog of *knots* in the univariate case. We associate with each vector p^j a piecewise linear function with vertices over the knots n^j of the triangulation T^j,

$$L[n^j, p^j](n^j_i) = p^j_i.$$

These linear functions are the *control polyhedra* defined by the subdivision process. Taking the limit of this sequence of functions yields the limit function associated with the initial data p

$$F[n,p](t) = \lim_{j\to\infty} L[n^j, p^j](t).$$

Here, the limit is taken *point-wise*, that is individually for each distinct t. By construction, the limit operator F is linear in p. Specifically,

$$\begin{aligned} F[n,\alpha p](t) &= \alpha F[n,p](t), \\ F[n,p+q](t) &= F[n,p](t) + F[n,q](t). \end{aligned}$$

Scaling the vector n is also equivalent to scaling the parameter t,

$$F[\alpha n, p](t) = F[n,p](\frac{t}{\alpha}).$$

7.1.1 Basis functions

Let e^i be the vector whose ith entry is one with the remaining entries being zero. Given the knot sequence n^j, the basis function associated with the control point p^j_i is $F[n^j, e^i](t)$. The edges of T^j bound the support of $F[n^j, e^i](t)$. The size of the support of the columns of $S[T^j]$ determines the size of the support of $F[n^j, e^i](t)$.

Theorem 20 *The support of $F[n^j, e^i](t)$ is the m-disc of n^j_i in T^j.*

Proof: Without loss of generality, we show that the support of $F[n^0, e^0]$ is the m-disc of n^0_0 in T^0. We keep track of the range of indices of non-zero coefficients during subdivision. After one round of subdivision, the non-zero coefficients lie in the m-disc of n^0_0 in T^1. After

k rounds of subdivision, the non-zero coefficients lie in the $m(2^k - 1)$-disc centered at n_0^0 in T^k. Note that the m-disc of n_0^0 in T^0 is exactly the $m2^k$-disc of n_0^0 in T^k. As $k \to \infty$, these two disc agree. □

7.1.2 Reduction to the stationary case

The key to analyzing the smoothness of a subdivision scheme is to reduce the scheme locally to an equivalent stationary scheme. For triangulations that are self-similar under subdivision, the subdivision process is stationary since $S[T] = S[A(T)] = S[D(T)]$. Of course, not every triangulation T is self-similar under subdivision.

However, compact support and local definition ensure the limit function $F[n, p](t)$ depends only on a finite portion of the triangulation near the dyadic point t. After a sufficient number of subdivisions, the mesh local to t is self-similar under subdivision. Specifically, there are three possible case:

- t lies in the interior of a triangle of T^0. After a sufficient number of subdivisions, the mesh around t is a regular three direction mesh.

- t lies on the interior of an edge of T^0. After a sufficient number of subdivisions, the mesh around t consists of the edge in T^0 separating two regular triangulations.

- t lies at a vertex of valence k of T^0. The mesh around t consist of k edges emanating from t that separate k distinct regular meshes.

The shaded regions in figure 7.3 illustrates these three possible cases. In each case, the resulting mesh is locally self-similar under subdivision.

In each of these three case, if t is translated to the origin, then $D(T)$ is a scaled version of itself, $\frac{1}{2}T$. This observation allows us to focus our analysis on triangulations that satisfy $D(T) = \frac{1}{2}T$. The resulting stationary schemes are centered at the origin. Note that the stationary subdivision matrix $S[T]$ still depends on the triangulation T. To simplify, we refer to the matrix $S[T]$ as S.

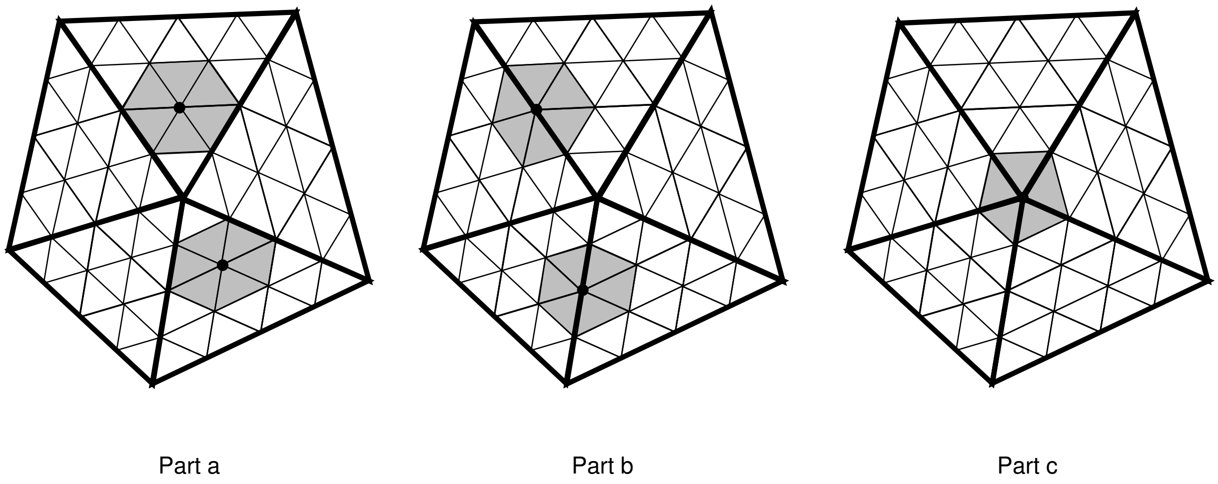

Figure 7.3 Portions of a triangulation that are self-similar under subdivision

7.2 Spectral conditions for irregular subdivision

7.2.1 Spectral analysis

As in the case of curves, our approach to analyzing the smoothness of $F[n,p](t)$ is to express this function locally as a linear combination of functions $F[n,x_i](t)$ where the x_i are eigenvectors of S. Again, the interesting spectral properties of S are captured by a finite submatrix of S.

Let the *bar* operator, $\bar{p}$, select the entries p_i from the infinite vector p such the n_i lie in the m-disc of the origin in T. By theorem 20, these are the only entries of p that affect the limit function in the 1-disc of the origin. The bar operator applied to the matrix S yields $\bar{S}$, the *square* matrix with entries S_{ij} where n_i lies in the m-disc of the origin in T and n_j lies in the m-disc of the origin in $D(T)$. The eigenstructure of S and $\bar{S}$ are related as follows.

Theorem 21 *Let $\lambda \neq 0$ be an eigenvalue of $\bar{S}$ with associated eigenvector y. Then S has eigenvalue λ with a unique associated eigenvector x such that $\bar{x} = y$.*

The proof of this theorem is simple and left to the reader. The eigenvectors of $\bar{S}$ with eigenvalue zero have no effect on the final limit function at the origin since after one round of subdivision the control points are mapped to zero. For the sake of simplicity, we assume that $\bar{S}$ has no zero eigenvalues.

If $\bar{S}$ does not have a full set of eigenvectors, then $\bar{S}$ is *defective*. A non-defective $\bar{S}$ has a full set of linearly independent eigenvectors $\bar{x}_0$, ..., $\bar{x}_k$. The extension of these eigenvectors, x_0, ..., x_k, can be used in the following manner.

Theorem 22 *Let $\bar{S}$ be non-defective matrix with eigenvectors $\bar{x}_0$, ..., $\bar{x}$. If*

$$\bar{p} = \sum_{i=0}^{k} c_i \bar{x}_i,$$

then, for all t in the 1-disc of the origin,

$$F[n,p](t) = \sum_{i=0}^{k} c_i F[n, x_i](t).$$

Proof: The vector $p - \sum_{i=0}^{k} c_i x_i$ is zero for entries in the m-disc of the origin. By theorem 20,

$$F[n, p - \sum_{i=0}^{k} c_i x_i](t) = 0$$

for all t in the 1-disc of the origin. The theorem follows by the linearity of the limit operator F. □

For defective $\bar{S}$, *generalized* eigenvectors can be used in place of eigenvectors. Each of the generalized eigenvectors, $\bar{x}_0$, ..., $\bar{x}_k$, is either an eigenvector of $\bar{S}$ or satisfies

$$\bar{S}\bar{x}_i = \lambda_i \bar{x}_i + \bar{x}_{i-1}.$$

These generalized eigenvectors can be extend to a set of infinite vectors, x_0, ..., x_k, satisfying

$$S x_i = \lambda_i x_i + x_{i-1}. \tag{7.2}$$

The proof of this fact is exactly the same are the proof of theorem 21. Since the generalized eigenvectors are linearly independent, theorem 22 also holds for these vectors. For subsequent theorems, we assume that $\bar{S}$ is non-defective. Where appropriate, we state the variant of the theorem that holds for defective $\bar{S}$ using generalized eigenvectors.

7.2.2 A spectral recurrence

As in the case of curves, the smoothness properties of a stationary subdivision scheme are tied to the spectral properties of its subdivision matrix S. In particular, the limit function associated with an eigenvector of a stationary subdivision scheme satisfies a fundamental relation.

Theorem 23 *Let x be a vector satisfying $Sx = \lambda x$. Then,*

$$\lambda F[n,x](t) = F[n,x](\frac{t}{2}). \tag{7.3}$$

Proof: The proof consist of simply recalling the definition of F.

$$\begin{aligned}
\lambda F[n,x](t) &= F[n,\lambda x](t), \\
&= F[n,Sx](t), \\
&= \lim_{j\to\infty} L[n^j, S^j(Sx)](t), \\
&= \lim_{j\to\infty} L[2n^{j+1}, S^{j+1}x](t), \\
&= \lim_{j\to\infty} L[n^{j+1}, S^{j+1}x](\frac{t}{2}), \\
&= F[n,x](\frac{t}{2}).
\end{aligned}$$

□

If x_i is a generalized eigenvectors as in equation 7.2, then

$$\lambda_i F[n,x_i](t) + F[n,x_{i-1}](t) = F[n,x_i](\frac{t}{2}). \tag{7.4}$$

Again, the proof is exactly as above.

7.2.3 Properties of the recurrence

The recurrence of theorem 23 is a powerful tool for analyzing stationary subdivision schemes. The following lemma illustrates several properties of such recurrences. (Note $|t|$ denotes the distance from t to the origin 0.)

Lemma 2 *Let $g(t)$ be a function non-zero away from the origin satisfying*

$$\lambda g(t) = g(\frac{t}{2}). \tag{7.5}$$

1. *If $g(t)$ is bounded on the annulus $\frac{1}{2} < |t| \leq 1$, then the limit as $t \to 0$ of $\frac{g(t)}{|t|^k}$ is zero.*
2. *If $\lambda = 1$ and $g(t)$ is continuous at the origin, then $g(t)$ is the constant function.*
3. *If $g(d) \neq 0$ for $d \neq 0$ and $|\lambda| > 1$ then the limit as $t \to 0$ of $g(t)$ diverges.*

Proof: For part one, let

$$v = \max_{\frac{1}{2}<|t|\leq 1} |g(t)|.$$

If $\frac{1}{2^{i+1}} < |t| \leq \frac{1}{2^i}$, then iterating equation 7.5 yields

$$\begin{aligned} |g(t)| &= |\lambda^i g(2^i t)|, \\ &\leq |\lambda|^i v. \end{aligned}$$

Dividing the lefthand side of this relation by $|t|^k$ and the righthand side by $\frac{1}{2^{k(i+1)}}$ respectively yields

$$\begin{aligned} \left|\frac{g(t)}{|t|^k}\right| &\leq 2^{k(i+1)}|\lambda|^i v, \\ &\leq (2^k|\lambda|)^i 2^k v \end{aligned}$$

Since $|\lambda| < \frac{1}{2^k}$, the limit as $i \rightarrow \infty$ and therefore as $t \rightarrow 0$ of $\frac{g(t)}{|t|^k}$ must be zero.

For part two, we observe that if there exists $d \neq e$ such that

$$g(d) - g(e) = c \neq 0,$$

then by equation 7.5

$$g(\frac{d}{2^i}) - g(\frac{e}{2^i}) = c \neq 0. \tag{7.6}$$

As i goes to infinity, $\frac{d}{2^i}$ and $\frac{e}{2^i}$ approach zero. Equation 7.6 contradicts that fact that $g(t)$ is continuous at the origin. Therefore, $g(d) = g(e)$ for all d and e.

For part three, we note that by hypothesis there exists $d \neq 0$ such that $g(d) \neq 0$. Since $g(\frac{d}{2^i}) = \lambda^i g(d)$ and $|\lambda| > 1$, $g(t)$ must diverge as $i \rightarrow \infty$ and $t = \frac{d}{2^i} \rightarrow 0$. $\square$

7.2.4 Necessary conditions for C^k subdivision

By theorem 22, we can restrict our smoothness analysis to those limit functions $F[n,x](t)$ where x is an eigenvector of S with eigenvalue λ. If $F[n,x](t)$ is a C^k continuous function, then x and λ must have special properties. Let $F^{(i)}[n,x](t)$ denote the ith derivative of $F[n,x](t)$ in any set of i directions.

Theorem 24 *Let $Sx = \lambda x$ with $|\lambda| \geq \frac{1}{2^k}$. If $F[n,x](t)$ is C^k continuous and not identically zero, then there exists $0 \leq i \leq k$ such that*

- $\lambda = \frac{1}{2^i}$,

- *$F[n, x](t)$ is a homogeneous polynomial of degree i.*

Proof: We first show that $F^{(k)}[n, x](t)$ is a constant function. Take the kth derivative of equation 7.3.

$$(2^k\lambda)F^{(k)}[n, x](t) = F^{(k)}[n, x](\frac{t}{2}).$$

If $F^{(k)}[n, x](t)$ is not a constant function, then $F^{(k)}[n, x](t)$ must diverge as $t \rightarrow 0$ either by part two of lemma 2 ($2^k\lambda = 1$) or part three of lemma 2 ($|2^k\lambda| > 1$). However, this contradicts the continuity of $F^{(k)}[n, x](t)$.

Since $F^{(k)}[n, x](t)$ is a constant function, $F[n, x](t)$ is a polynomial function of degree k. Since $F[n, x](t)$ is not identically zero, choose the minimal set of i directions such that $F^{(i)}[n, x](0) \neq 0$ and take the ith derivative with respect to these directions of equation 7.3.

$$\lambda F^{(i)}[n, x](t) = \frac{1}{2^i}F^{(i)}[n, x](\frac{t}{2}).$$

For lefthand and righthand side of this equation to agree at $t = 0$, λ must be $\frac{1}{2^i}$. Since $F^{(i)}[n, x](t)$ is continuous, by part two of lemma 2, $F^{(i)}[n, x](t)$ must be the constant function. Since all lower order derivatives are zero at the origin, $F[n, x](t)$ is a homogeneous polynomial of degree i. □

If a generalized eigenvector (equation 7.2) produces a C^k limit curve, then its eigenvalue must have modulus less than $\frac{1}{2^k}$. Consider the continuous (C^0) case. If $\lambda_0 = 1$, then $F[n, x_0](t)$ is the constant function. If x_1 were a generalized eigenvector satisfying

$$Sx_1 = \lambda_1 x_1 + x_0,$$

then $\lambda_0 = \lambda_1 = 1$. In terms of limit functions,

$$\begin{aligned} \lambda_1 F[n, x_1](t) + F[n, x_0](t) &= F[n, x_1](\frac{t}{2}), \\ F[n, x_1](t) + 1 &= F[n, x_1](\frac{t}{2}). \end{aligned}$$

As $t \rightarrow 0$, $F[n, x_1](t)$ must diverge.

Theorem 24 (in conjunction with theorem 28) yield necessary conditions for a bivariate subdivision scheme to produce C^k continuous limit functions. If there exist initial data p such that the all possible directional derivatives of $F[n, p](t)$ up to order k are nonzero, then there must exist eigenvectors of S that reproduce each monomial $(t_1)^i(t_2)^j$ for all $i + j \leq k$.

Specifically, the subdivision scheme defined by S can reproduce any polynomial up to degree k. The spectrum of S includes

$$1, \frac{1}{2}, \frac{1}{2}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{8}, \ldots$$

If the basis functions associated with the scheme are linearly independent, then these eigenvalues must also be unique. If a full span of derivatives at the origin does not exist, then the spectrum may be missing certain powers of two. If the scheme has linearly dependent basis functions, several powers of two may be repeated.

7.3 Convergence conditions for irregular subdivision

To show that an irregular subdivision scheme is convergent or has a particular order of smoothness, we derive a difference operator Δ that annihilates a set of eigenvectors X corresponding to low degree polynomials and then build a subdivision operator D for that difference scheme satisfying

$$\Delta S = D\Delta.$$

7.3.1 Difference schemes

In the case of C^0 continuity, the difference operator annihilates the eigenvector $x_0 = \mathbf{1}$. Using essentially the same proof as in chapter 3, one can show that the difference scheme defined by D uniformly converges to zero if and only if the scheme defined by S uniformly converges to a continuous function.

A bivariate function $f(t) = f(t_1, t_2)$ is C^1 continuous if the partial derivatives of $f(t)$ in two independent directions are themselves C^0 functions. To show that a subdivision scheme produces C^1 continuous functions, we must construct a subdivision scheme for the directional derivatives and then show that the resulting scheme is continuous. By the necessary conditions, a C^1 continuous scheme with non-zero derivatives at the origin must reproduce the linear functions t_1 and t_2.

To simplify the subsequent analysis, we assume the t_1 and t_2 components of the knot vector n uniquely reproduce the linear functions t_1 and t_2. Stated another way, these two components are x_1 and x_2, the two eigenvectors of S with associated with the eigenvalue $\frac{1}{2}$. Note that this condition is one of convenience, not necessity. Reasonable subdivision schemes are possible that do not satisfy this condition. We simply know of no irregular subdivision schemes that do not satisfy this property.

The difference operator Δ that computes the directional derivative of $F[n,p](t)$ in the t_2 direction should satisfy

$$\begin{aligned} \Delta\mathbf{1} &= \mathbf{0}, \\ \Delta x_1 &= \mathbf{0}, \\ \Delta x_2 &= \mathbf{1}. \end{aligned}$$

In particular, we construct a Δ whose ith row, Δ_i, has nonzero entries corresponding to the three vertices of the ith triangle in $T = T^0$. These three entries are chosen so that $\Delta_i p^0$ return the directional derivative of $L[n^0, p^0](t)$ in the t_2 direction over the ith triangle of T^0. Since each triangulation T^j is similar to T^0, Δ also acts as a difference operator on T^j. We need only scale Δ by 2^j to reflect the fact that $\frac{1}{2^j}x_2$ reproduces t_2 on T^j.

Given Δ, we desire a subdivision matrix D that satisfies

$$(2\Delta)S = D\Delta. \tag{7.7}$$

Figure 7.4 illustrates this relation in the case of piecewise linear subdivision. This relation states that subdividing the original scheme and then taking the discrete derivative is equivalent to taking the discrete derivative and then subdividing the derivative scheme. Given such matrices Δ and D, the following theorem holds.

Theorem 25 *Let the sequence $L[n^j, D^j q](t)$ uniformly converge to a continuous function for all bounded $||q||$. Then, for all bounded $||p||$,*

$$\lim_{j\to\infty} L[n^j, D^j(\Delta p)](t)$$

is the directional derivative of $F[n,p](t)$ with respect to t_2.

Proof: Recall that $F[n,p](t) = g(t_1, t_2)$ can be expressed as the limit of the functions

$$g_j(t_1, t_2) = L[n^j, S^j p](t).$$

The directional derivatives, $g'_j(t_1, t_2)$, of the $g_j(t_1, t_2)$ are piecewise constant functions over the triangulation T^j with piecewise values $(2^j\Delta)S^j p$. By the construction of D,

$$(2^j\Delta)S^j = D^j\Delta.$$

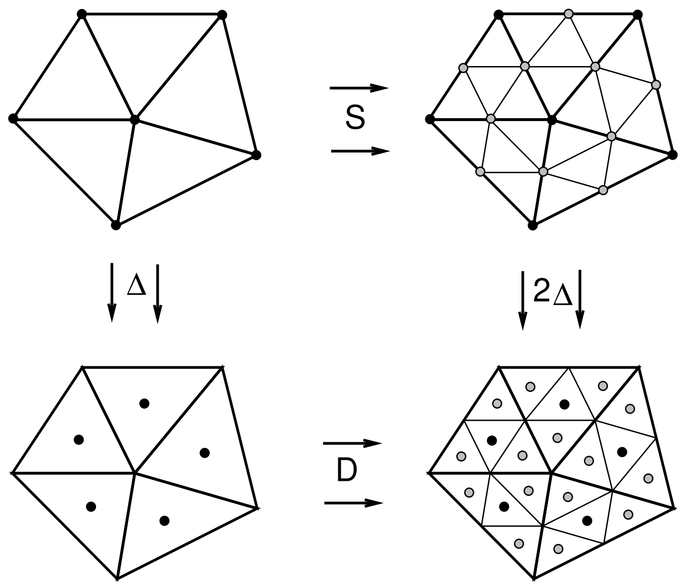

Figure 7.4 A derivative scheme for piecewise linear interpolation

So, the $g_j'(t_1, t_2)$ have piecewise values $D^j(\Delta p)$.

By hypothesis, the sequence

$$h_j(t_1, t_2) = L[n^j, D^j(\Delta p)](t)$$

uniformly converges to a continuous function, call it $h(t_1, t_2)$. The difference between the function $g_j'(t_1, t_2)$ and $h_j(t_1, t_2)$ uniformly converge to zero. Therefore, the sequence $g_j'(t_1, t_2)$ uniformly converges to $h(t_1, t_2)$.

We next prove that $\int_0^{t_2} g_j'(t_1, s)ds$ is point-wise convergent to $\int_0^{t_2} h(t_1, s)ds$ for any t_2. Fix t_2. By the uniform convergence of the $g_j'(t_1, t_2)$ to $h(t_1, t_2)$, for all $\epsilon > 0$, there exist an n such that for all $j \geq n$

$$|g_j'(t_1, t_2) - h(t_1, t_2)| < \epsilon.$$

In terms of the integrals, for all $j \geq n$,

$$\begin{aligned} \left| \int_0^{t_2} g_j'(t_1, s)ds - \int_0^{t_2} h(t_1, s)ds \right| &\leq \int_0^{t_2} |g_j'(t_1, s) - h(t_1, s)|ds, \\ &\leq \epsilon t_2. \end{aligned}$$

Therefore, $g_j(t_1, t_2) - g_j(t_1, 0)$ converges to $\int_0^{t_2} h(t_1, s)ds$ for any t_2. However, $g_j(t_1, t_2) - g_j(t_1, 0)$ also converges to $g(t_1, t_2) - g(t_1, 0)$. So,

$$g(t_1, t_2) = \int_0^{t_2} h(t_1, s)ds + g(t_1, 0)$$

for any t_2. Thus, $h(t_1, t_2)$ is the derivative of $g(t_1, t_2)$ with respect to t_2. □

To show that S produces C^1 schemes, one must still show that directional derivative scheme D defines a continuous scheme. This requires showing that the difference scheme associated with D uniformly converges to zero. If one wishes to verify the smoothness of the scheme defined by S directly, the following method can be used.

Construct Δ that annihilates x_0, x_1, and x_2 and whose rows are supported over each pair of edge adjacent triangles in T. If the subdivision matrix D satisfies

$$(2\Delta)S = D\Delta, \tag{7.8}$$

then the following theorem holds.

Theorem 26 *If the sequence $L[n^j, D^j q](t)$ uniformly converges to zero for all bounded $||q||$, then $F[n,p](t)$ is C^1 continuous for all bounded $||p||$.*

Proof: Let Δ^1 be the directional derivative operator in the t_2 direction. Δ^1 annihilates x_0 and x_1 and maps x_2 to $\mathbf{1}$. Let Δ^0 be the difference operator for the directional derivative scheme that annihilates $\mathbf{1}$ and whose rows are supported over pairs of edge adjacent triangles in T.

Note that $\Delta^0\Delta^1$ annihilates x_0, x_1, and x_2 and has rows whose support correspond to pairs of adjacent triangles in T. If T has no edges parallel to the t_2 direction, then none of these rows are identically zero. (Otherwise, we take the derivative in a different direction.) Therefore, the rows of Δ and $\Delta^0\Delta^1$ must agree up to multiplication by a constant

$$\Delta = C\Delta^0\Delta^1,$$

where C is a diagonal matrix.

Substituting into equation 7.8 yields that

$$(2C\Delta^0\Delta^1)S = D(C\Delta^0\Delta^1).$$

If we let $D0 = C^{-1}DC$, then

$$(2\Delta^0\Delta^1)S = D0(\Delta^0\Delta^1). \tag{7.9}$$

By construction, there exists a subdivision scheme $D1$ for the derivative in the x_2 direction satisfying

$$(2\Delta^1)S = D1\Delta^1.$$

Substituting into equation 7.9 yields

$$(\Delta^0 D1)\Delta^1 = (D0\Delta^0)\Delta^1.$$

So, for all input $\Delta^1 p$, $D0$ is the difference scheme for $D1$.

To conclude, D uniformly converging to zero implies that $D0$ uniformly converges to zero and $D1$ converges to a continuous function. Repeating this proof for the direction t_1, S defines a C^1 scheme. $\Box$

Note that this theorem does not state that the intermediate directional derivative schemes are continuous for all input. These schemes are guaranteed to be continuous only for input $\Delta^1 p$.

7.3.2 A local construction for difference schemes

The previous section assumed the existence of D given Δ and S. Next, we give a very general construction for building such difference schemes D. The input is a stationary subdivision matrix S and a finite set of eigenvectors X satisfying

$$SX = X\Lambda.$$

Given a Δ such that $\Delta X = 0$, we wish to construct a subdivision matrix D for the difference scheme such that

$$\Delta S = D\Delta.$$

(Note that any extra scaling factor for Δ can be absorbed into D.)

The following theorem gives a sufficient condition on the rows of Δ for the matrix D to exist. If Δ_i is the ith row of Δ, then this theorem allows the ith row of D, D_i to be be constructed locally.

Theorem 27 *Let Θ be the support of $\Delta_i S$. Let Ω be the set of k such that the support of Δ_k is in Θ. If the restriction of vectors in X to Θ are linearly independent and*

$$RowRank(\Delta_\Omega) = |\Theta| - |X|,$$

then there exists D_i with support Ω such that

$$\Delta_i S = D_i \Delta.$$

Proof: The non-zero entries of the rows of Δ_Ω form a $|\Theta|$-dimensional space. Since the vectors in X are linearly independent on Θ and $\Delta_\Omega X = 0$, the rows of Δ_Ω lie in a $|\Theta| - |X|$-dimensional subspace. By the hypothesis, these rows span that space.

Since $\Delta_i S$ has non-zero entries in the same position and

$$(\Delta_i S)X = \Delta_i X \Lambda = 0,$$

$\Delta_i S$ must also lie in the subspace spanned by the rows of Δ_Ω. Therefore, there exists a linear combination, D_i, of these rows that reproduce $\Delta_i S$. $\square$

This theorem explains the choice of supports for rows of Δ in the previous section. These support were

- An edge in T if Δ annihilates x_0,
- A triangle in T if Δ annihilates x_0 and x_1,
- A pair of edge adjacent triangles in T if Δ annihilates x_0, x_1, and x_2.

In each of these cases, Δ has the appropriate row rank for any Θ consistent with S having compact support. For these Δ, the matrix D always exists independent of the size of the support of the rows of S.

If S has compactly supported subdivision rules, then its associated difference scheme also has compactly supported rules. Figure 7.5 gives an example of this theorem applied to the subdivision rules for the three direction, C^2 quartic box-spline. The difference rule is the directional derivative rule supported over the triangles of T. The top portion of the figure shows the subdivision rules for this scheme. The middle portion of the figure shows the neighborhood Θ in T arising from a row of Δ for $D(T)$. There are two types of rows, those corresponding to triangles in $D(T)$ that contain a vertex of T (right) and those corresponding to triangles in $D(T)$ that do not contain a vertex of T (left). Each neighborhood gives rise to a different subdivision rule for the derivative scheme. The lower portion of the figure shows the two types of subdivision rules associated with the derivative scheme. This derivative scheme is the C^1 cubic half-box spline described in [Goo90, pp. 359].

Theorem 27 gives a very general method for constructing difference schemes. Of course, we still must prove that these schemes converge to the appropriate set of derivatives. Unfortunately, we have a proof of such convergence only in the C^0 and C^1 cases. One might

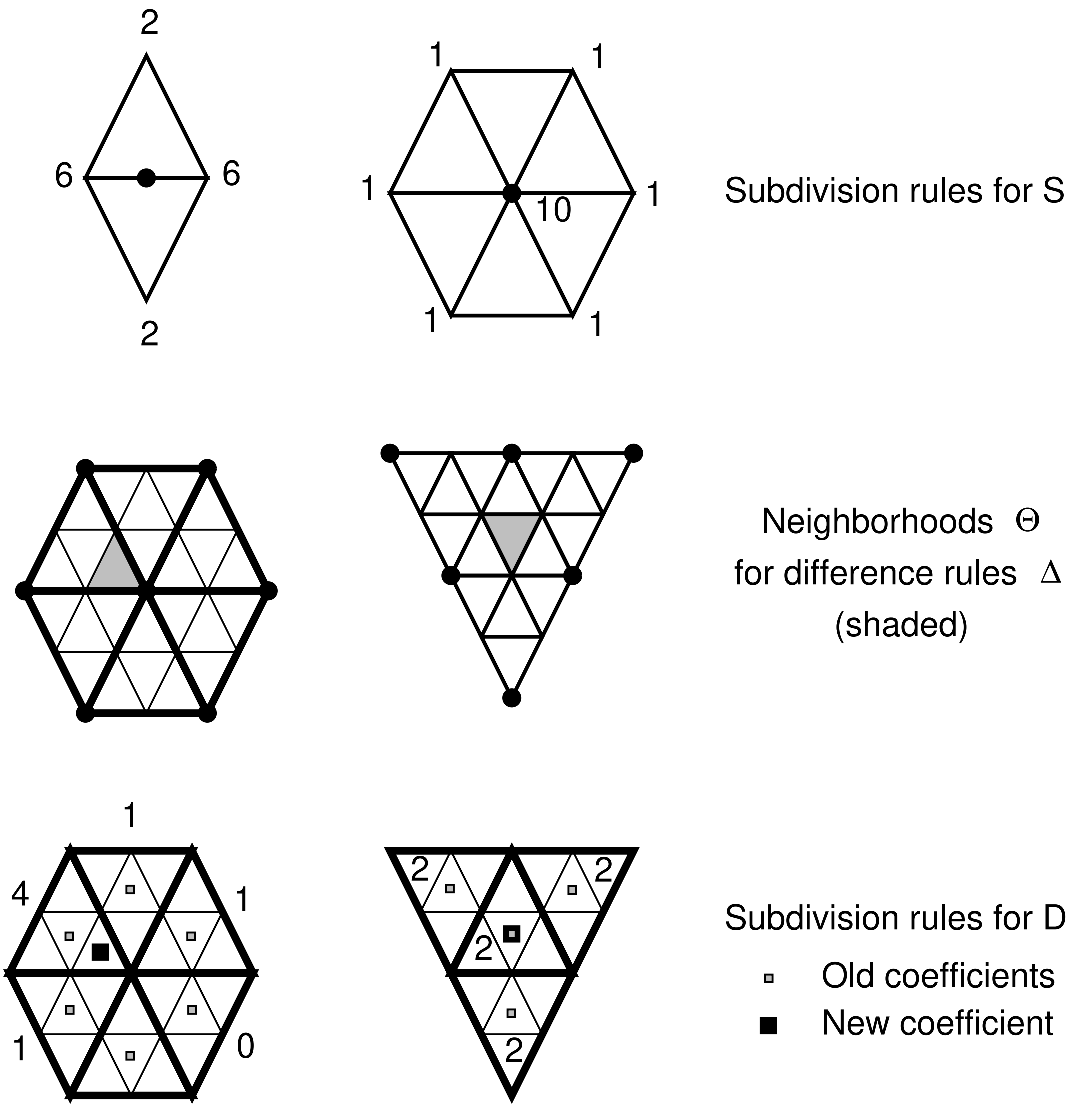

Figure 7.5 Subdivision rules for directional derivatives

try to iterate this process of taking directional derivatives in theorem 25. Unfortunately, the directional derivative of a triangular scheme has control points over the dual of the triangulation. In other words, the mesh for the derivative scheme has a hexagonal connectivity. The directional derivative of such a mesh is unclear.

7.4 An approximating C^1 scheme

We conclude this chapter by constructing an approximating scheme that is C^1 for a large class of irregular triangulations. To the best of the author's knowledge, this scheme is the first instance of a smooth functional subdivision scheme over irregular triangulations.

The building block for this scheme is piecewise linear interpolation. The subdivision matrix for piecewise linear interpolation nearly satisfies the necessary conditions for C^1 continuity. In the valence six case, this matrix is

$$\begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 & 0 & 0 \\ \frac{1}{2} & 0 & 0 & \frac{1}{2} & 0 & 0 & 0 \\ \frac{1}{2} & 0 & 0 & 0 & \frac{1}{2} & 0 & 0 \\ \frac{1}{2} & 0 & 0 & 0 & 0 & \frac{1}{2} & 0 \\ \frac{1}{2} & 0 & 0 & 0 & 0 & 0 & \frac{1}{2} \end{pmatrix}.$$

The scheme has linear precision and a spectrum of the form $1, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \ldots$. Our approach is to perturb S so that $|\lambda_3| < \frac{1}{2}$ while maintaining linear precision.

7.4.1 Perturbation using Δ

If Δ is the difference operator that annihilates x_0, x_1, and x_2, then perturbing the rows of S by a combination of rows of Δ maintains linear precision . The support for the ith row of Δ, Δ_i, is a pair of triangles on either side of the ith edge of T. We normalize each row of Δ to be of the form in figure 7.6. a and b are chosen so that $\Delta_i x_1 = 0$ and $\Delta_i x_2 = 0$. For regular, three direction meshes, a and b are zero.

The subdivision rules for the new scheme are now as follows:

- The subdivision rule for the midpoint of the ith edge of T is the linear subdivision rule for edge i plus $\frac{1}{8}\Delta_i$.

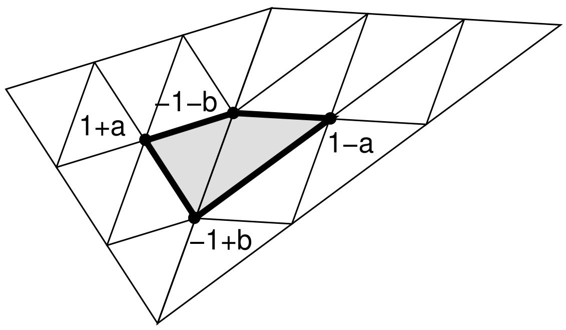

Figure 7.6 Difference mask for convergence to a C^1 scheme

- The subdivision rule at a vertex v of T is the linear subdivision rule for vertex v plus $\frac{1}{16}\sum_i \Delta_i$ where i varies over the indices of those edges incident on v.

If T is a regular, three direction mesh, then these rules reproduce the rules for the C^2 quartic box-spline with direction vectors $\{(1,0),(0,1),(1,0),(0,1),(1,0),(0,1)\}$.

7.4.2 Proof of C^1 continuity

We next characterize the class of triangulations for which this scheme produces C^1 limit functions. The smoothness analysis decouples into three case.

Interior of faces in T^0 In the interior of faces of T^0, the meshes T^j are three direction meshes (see part a of figure 7.3). As observed above, the subdivision rules define the three direction, C^2 quartic box-spline.

Interior of edges in T^0 Along the interior of an edge e of T^0, the mesh T^j consists of two regular, three direction meshes separated by the edge e (see part b of figure 7.3). If the pairs of triangles sharing the edge e have a difference mask as in figure 7.6, then the subdivision matrix S for this scheme is

$$\frac{1}{16}\begin{pmatrix}
10 & 1+b & 1+a & 1+a & 1-b & 1-a & 1-a & 0 & 0 & 0 & 0 & 0 & 0\\
6-2b & 6+2b & 2+2a & 0 & 0 & 0 & 2-2a & 0 & 0 & 0 & 0 & 0 & 0\\
6 & 2 & 6 & 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
6 & 0 & 2 & 6 & 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
6+2b & 0 & 0 & 2+2a & 6-2b & 2-2a & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
6 & 0 & 0 & 0 & 2 & 6 & 2 & 0 & 0 & 0 & 0 & 0 & 0\\
6 & 2 & 0 & 0 & 0 & 2 & 6 & 0 & 0 & 0 & 0 & 0 & 0\\
2 & 6 & 6 & 0 & 0 & 0 & 0 & 2 & 0 & 0 & 0 & 0 & 0\\
2 & 0 & 6 & 6 & 0 & 0 & 0 & 0 & 2 & 0 & 0 & 0 & 0\\
2 & 0 & 0 & 6 & 6 & 0 & 0 & 0 & 0 & 2 & 0 & 0 & 0\\
2 & 0 & 0 & 0 & 6 & 6 & 0 & 0 & 0 & 0 & 2 & 0 & 0\\
2 & 0 & 0 & 0 & 0 & 6 & 6 & 0 & 0 & 0 & 0 & 2 & 0\\
2 & 6 & 0 & 0 & 0 & 0 & 6 & 0 & 0 & 0 & 0 & 0 & 2
\end{pmatrix}$$

Here, S is defined over a neighborhood large enough to include any non-uniformities in the resulting difference scheme. (Figure 7.7 shows this neighborhood (shaded) and its indexing into S.) The difference matrix Δ for this neighborhood is

$$\begin{pmatrix}
-1-b & -1+b & 1+a & 0 & 0 & 0 & 1-a & 0 & 0 & 0 & 0 & 0 & 0\\
-1 & 1 & -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
-1 & 0 & 1 & -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
-1+b & 0 & 0 & 1+a & -1-b & 1-a & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
-1 & 0 & 0 & 0 & 1 & -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0\\
-1 & 1 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0\\
1 & -1 & -1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0\\
1 & 0 & -1 & -1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0\\
1 & 0 & 0 & -1 & -1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0\\
1 & 0 & 0 & 0 & -1 & -1 & 0 & 0 & 0 & 0 & 1 & 0 & 0\\
1 & 0 & 0 & 0 & 0 & -1 & -1 & 0 & 0 & 0 & 0 & 1 & 0\\
1 & -1 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 1
\end{pmatrix}$$

Note that each of the twelve rows of Δ correspond to a pair of adjacent triangles. Since the rows of Δ are linearly dependent, the difference matrix D is not unique. However, using

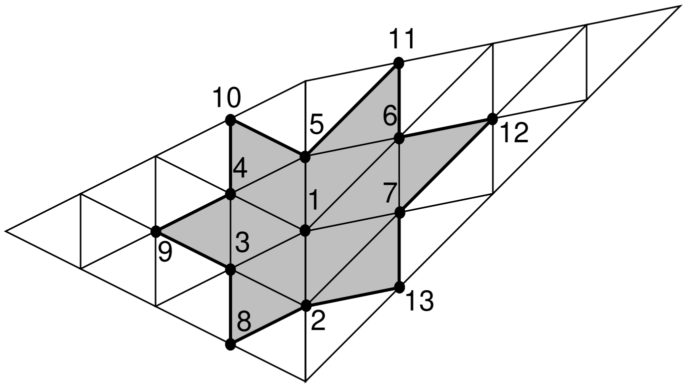

Figure 7.7 The neighborhood of S (shaded)

Mathematica, one can show that the matrix D

$$\frac{1}{8}\begin{pmatrix}
3+b & 0 & 0 & 1-b & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 4 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
-1 & 0 & 4 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1+b & 0 & 0 & 3-b & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
-1 & 0 & 0 & 1 & 4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & -1 & 0 & 4 & 0 & 0 & 0 & 0 & 0 & 0 \\
-1 & 0 & 2 & 1 & 0 & 0 & 2 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 1 & 0 & 0 & 0 & 2 & 0 & 0 & 0 & 0 \\
1 & 2 & 0 & -1 & 0 & 0 & 0 & 0 & 2 & 0 & 0 & 0 \\
1 & 0 & 0 & -1 & 0 & 2 & 0 & 0 & 0 & 2 & 0 & 0 \\
1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 2 & 0 \\
-1 & 0 & 0 & 1 & 2 & 0 & 0 & 0 & 0 & 0 & 0 & 2
\end{pmatrix}$$

satisfies $(2\Delta)S = D\Delta$.

If pairs of triangles sharing the edge of T^0 form a convex quadrilateral, then it is straightforward to show that $-1 \leq a, b \leq 1$. For a and b in this range, the matrix D has norm $\frac{3}{4}$. Therefore, the scheme associated with D uniformly converges to zero. By theorem 26, the rules of S produces C^1 functions locally. Note that in this case, the resulting subdivision rules are also non-negative.

Vertices of T^0 Analyzing the smoothness of the subdivision scheme at vertices of T^0 takes a different approach. Parameterizing S and Δ by the local mesh geometry and solving symbolically for D appears to be beyond the capabilities of packages such as Mathematica. Instead, we prove a generalization of theorem 14 for the bivariate case.

By the previous analysis, the subdivision scheme is C^1 continuous everywhere except at vertices of T^0. Local to these vertices, the subdivision scheme S is stationary. By theorem 22, the final limit function $F[n,p](t)$ can be expressed as a linear combination of limit functions $F[n,x](t)$ associated with eigenvectors of S. The eigenvectors associated with the three dominant eigenvalues reproduce polynomials. The next theorem characterizes the smoothness of the functions associated with remaining eigenvectors.

Theorem 28 *Let $Sx = \lambda x$ with $|\lambda| < \frac{1}{2}$. If $F[n,x](t)$ is C^1 continuous everywhere except at $t = 0$, then $F[n,x](t)$ is C^1 continuous everywhere.*

Proof: We first show that $F[n,x](t)$ exists and is continuous at $t = 0$. We note that $F[n,x](0)$ must be zero since $|\lambda| < \frac{1}{2}$. The limit as t goes to zero of $F[n,x](t)$ is also zero by part one of lemma 2. Therefore, $F[n,x](t)$ is continuous at $t = 0$.

We next show that the directional derivative of $F[n,x](t)$ in the t^1 direction, $F'[n,x](t)$, exists and is zero at $t = 0$. By definition,

$$\begin{aligned} F'[n,x](0,0) &= \lim_{t_1 \to 0} \frac{F[n,x](t_1,0) - F[n,x](0,0)}{t_1}, \\ &= \lim_{t_1 \to 0} \frac{F[n,x](t_1,0)}{t_1}, \end{aligned}$$

since $F[n,x](0,0)$ is zero. Since $|\lambda| < \frac{1}{2}$, by part one of lemma 2, the limit as t_1 goes to zero of $\frac{F[n,x](t_1,0)}{t_1}$ must also be zero.

Given that the partial derivative of $F[n,x](t)$ exists at $t = 0$, we can take the partial derivative of equation 7.3. Applying part one of lemma 2 shows that $F'[n,x](t)$ is continuous at $t = 0$. Repeating this argument with t_2 in place of t_1 finishes the proof. □

This theorem can be generalized in several ways. The theorem also holds for higher orders of continuity as in theorem 14. This theorem also holds for generalized eigenvectors (equation 7.2) whose eigenvalues have moduli less than $\frac{1}{2}$. The proof involves modifying part one of lemma 2 to use equation 7.4.

For the subdivision scheme at hand, the spectrum of S in the case of a regular mesh has eigenvalues λ_3, λ_4, ... whose moduli are less than or equal to $\frac{1}{4}$. Since S is a continuous function of the local mesh geometry, these eigenvalues are also a continuous function of the local mesh geometry. Thus, small perturbations of T away from the regular case do not affect the smoothness of the scheme. The difficulty with this analysis is that it gives no geometric characterization of those triangulations that produce C^1 schemes. Determining whether the scheme is C^1 requires the computation of eigenvalues of S.

In practice, the eigenvalue appear to well-behaved for a large range of triangulations. The author is currently investigating possible improvement to this situation. One possibility would be to weight the rows of Δ by different amounts, depending on the geometry of T. The result would be a scheme whose eigenvalues λ_3, λ_4, ... always have moduli less than $\frac{1}{2}$. Such schemes would be guaranteed to be C^1 at vertices of T^0. Another possibility would be to develop better tools for computing the eigenvalues of parameterized matrices and improve the previous analysis. Figures 7.8 and 7.9 show two examples of functional subdivision over irregular meshes.

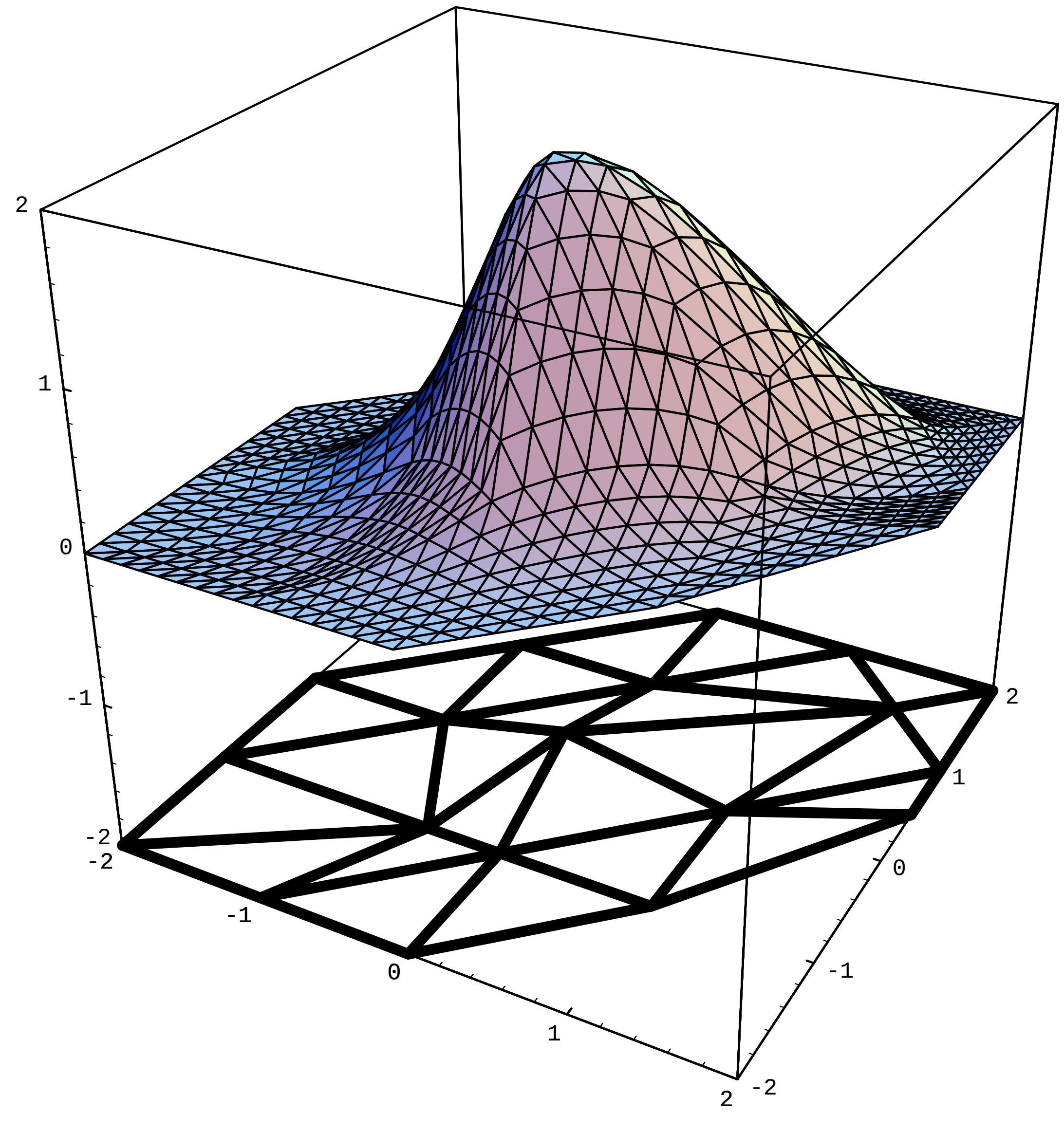

Figure 7.8 A basis function for a valence six vertex

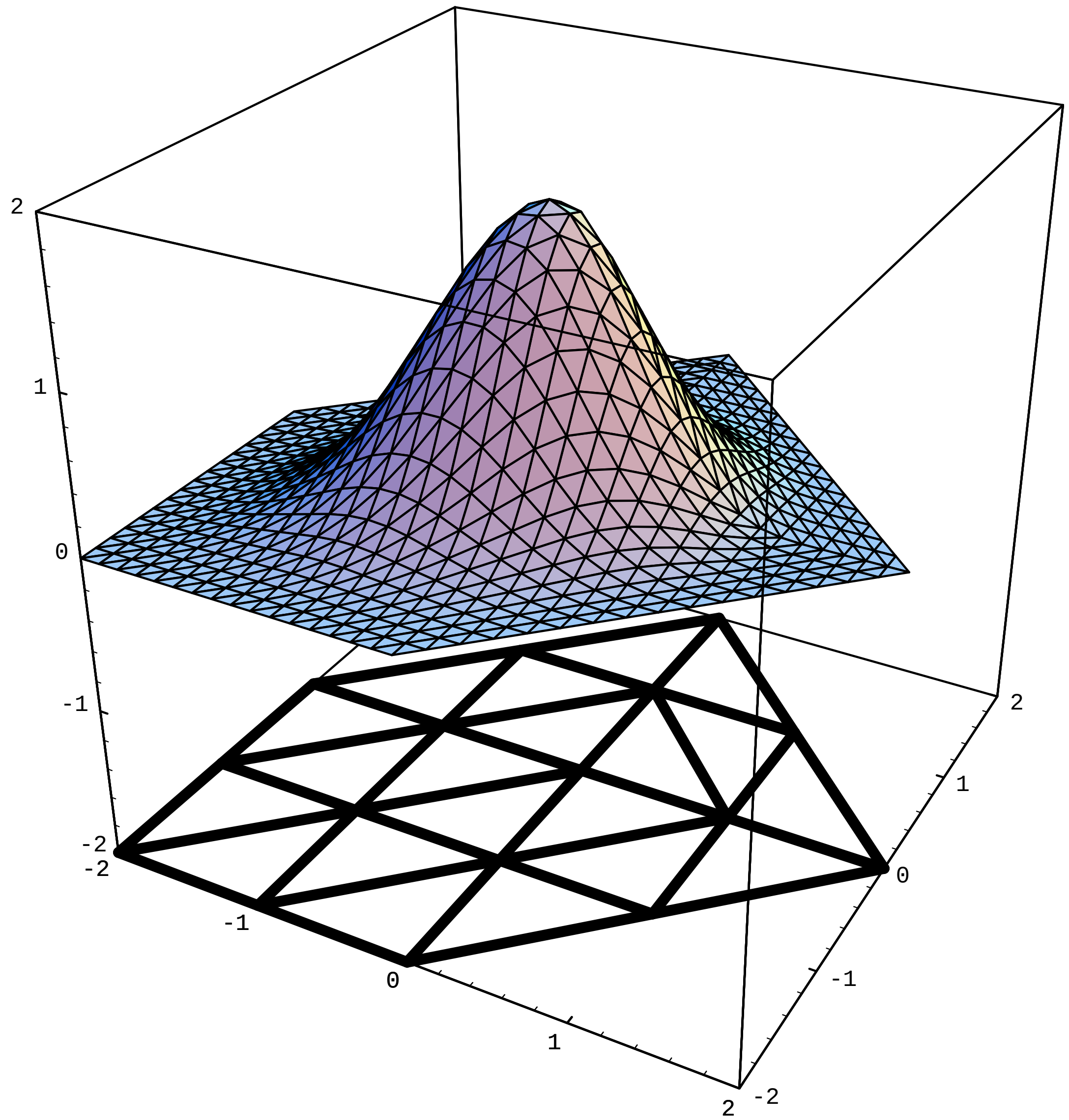

Figure 7.9 A basis function for a valence five vertex

Chapter 8

Subdivision schemes for triangular meshes

In the previous chapters, we have studied methods for creating C^k continuous functions over an infinite domain using subdivision. These functional techniques can be used to create parametrically defined, unbounded surfaces. Of course, these surfaces are just deformations of the plane. If one wishes to model closed surfaces (e.g. a sphere), then a purely functional approach will not always suffice. Fortunately, subdivision can be applied purely in the geometric domain with out recourse to globally defined functional domain. One might also like to model objects with boundaries. Using the notion of tagging, one can define special rules that produce boundary curves and vertices for surface patches. These extensions allows us to model complicated 3D shapes with a minimum amount of overhead.

8.1 C^k manifolds

One of our goals is to define closed, smooth surfaces purely through subdivision. In the functional case, we measured the smoothness of a function by noting the number of its partial derivatives that were continuous. In the geometric case, the notion of a global partial derivative is undefined. Instead, we measure the smoothness of the surface *locally.* A surface S is a C^k *manifold* if for every point $p \in S$ there exists an open neighborhood U_p of p such that $S \cap U_p$ is the graph of a C^k function. This definition applies to closed surfaces, that is those surface without boundaries. The next chapter deals with boundaries.

The next theorem provides our primary tool for showing that a surface is a locally the graph of a C^k function. If $f(t)$ is a vector valued function

$$f(t) = (f_1(t), f_2(t), .., f_m(t))$$

where $t = (t_1, ..., t_n)$ (with $m \geq n$), then the *differential* of $f(t)$, $Df(t)$, is the m by n matrix whose $(i,j)th$ entry is $\frac{\partial f_i(t)}{\partial t_j}$.

Theorem 29 *If $f(t)$ is C^k continuous and $Df(0)$ has full rank, then there exist a neighborhood U of 0 such that $\{f(t)|t \in U\}$ is locally the graph of a C^k function.*

Proof: If $Df(t)$ has full rank, then there exists an n by n submatrix with full rank. Without loss of generality, we assume one such submatrix consists of the first n rows. By the inverse function theorem, the transformation $\hat{f}(t) = (f_1(t), ..., f_n(t)$ has a local inverse $\hat{f}^{-1}(s)$ that is C^k continuous on a neighborhood $\hat{U}$ of $\hat{f}(0)$. Replacing $(t_1, ..., t_n)$ by $\hat{f}^{-1}(s)$ yields that

$$f(t) = (s_1, ..., s_n, f_{n+1}(\hat{f}^{-1}(s)), ..., f_m(\hat{f}^{-1}(s)))$$

on $\hat{U} \cap \hat{f}(U)$. Since the composition of two C^k functions is a C^k function, $f(t)$ is locally the graph of a C^k function. $\square$

Given this definition of smoothness, we next investigate methods for constructing such smooth surfaces via subdivision in the geometric domain.

8.2 Limitations of regular meshes

Given an initial triangular mesh T^0 in R^3, we can subdivide each triangular face of T^0 into four subfaces and position the new vertices based on some subdivision rules. (See figure 1.3.) The vertices of this new mesh, T^1, have valence six except for non-valence six vertices of T^0. If one uses the subdivision rules of the quartic box spline of section 6.4, then the limit of this subdivision process is a C^2 manifold everywhere except near the vertices of T^0. To see this, we note that in regions where the mesh has the connectivity of a three direction mesh, we can view each coordinate function as being graphed over a regular three direction mesh. Since the quartic box spline rules produce C^2 functions over a regular three direction mesh, each coordinate function is C^2 continuous for this particular parameterization. Therefore, in this neighborhood, the surface is a C^2 manifold.

The non-valence six vertices of T^0 are *extraordinary* points of the the mesh. One might ask if it is possible to avoid having exceptional points in the initial mesh? To answer this question, we must recall *Euler's* formula for polyhedra. It states that if v, e, and f are the number of vertices, edges, and faces in a closed polyhedron T, then

$$v - e + f = 2 - 2g,$$

where g is the *genus* of T. Roughly, g measures the number of distinct handles of T. For example, a sphere has zero handles and a doughnut (or coffee cup) has one handle.

If our initial mesh had no extraordinary points, then every vertex of the mesh must have valence six and the mesh is globally a three direction mesh. Thus, for every vertex there are three edges and two faces. So, the lefthand side of Euler's formula is zero. Therefore, the mesh must have genus one. If our initial mesh is topological a sphere (it has genus zero), then we must have extraordinary points.

8.3 C^1 subdivision methods for closed meshes

The solution to the problem of extraordinary vertices is to define special subdivision rules for these vertices. Of course, these rules should produce a limit surface that is locally the graph of a C^k function. In practice, creating a C^1 manifold at the extraordinary points is relatively straightforward. [Loo87] gives one such rule for use with the three direction quartic box spline rules mentioned above. [DS78, CC78] give rules for extraordinary points arising during the subdivision of quadrilateral meshes.

Proving that the resulting limit surface is a C^1 manifold is fairly difficult. A reliable proof for quadrilateral method was only recently given in [Rei94]. We next derive a class of subdivision masks for extraordinary points that includes Loop's rule. We then sketch a proof that these rules produce a C^1 manifold in a neighborhood of the extraordinary point.

Our rule for subdivision at an extraordinary vertex of valence n is as shown in figure 8.1. The coefficient for each adjacent vertex is multiplied by some weight a. The coefficient at the extraordinary vertex itself is multiplied by $1 - an$. For any choice of a, this rule has constant precision.

Consider a single extraordinary vertex surrounded by an infinite mesh of regular valence six vertices. The subdivision process centered at this extraordinary vertex is a stationary process since the rules used at each step of subdivision are the same. If we treat the coefficient vector and subdivision matrix as being infinite, then

$$p^{j+1} = Sp^j.$$

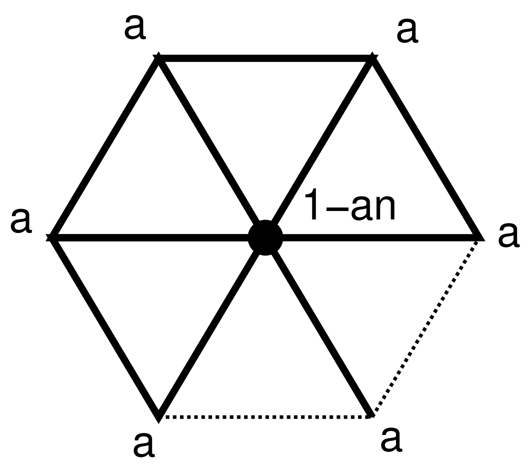

Figure 8.1 Subdivision for valence n extraordinary vertex

The restriction of S to the 1-disc around a valence five extraordinary vertex is

$$S = \begin{pmatrix} 1-5a & a & a & a & a & a \\ \frac{3}{8} & \frac{3}{8} & \frac{1}{8} & 0 & 0 & \frac{1}{8} \\ \frac{3}{8} & \frac{1}{8} & \frac{3}{8} & \frac{1}{8} & 0 & 0 \\ \frac{3}{8} & 0 & \frac{1}{8} & \frac{3}{8} & \frac{1}{8} & 0 \\ \frac{3}{8} & 0 & 0 & \frac{1}{8} & \frac{3}{8} & \frac{1}{8} \\ \frac{3}{8} & \frac{1}{8} & 0 & 0 & \frac{1}{8} & \frac{3}{8} \end{pmatrix}. \tag{8.1}$$

As observed in the previous chapter, the spectrum of a stationary subdivision process is directly related to the smoothness of the resulting scheme. We next characterize the spectrum of S. Let $\bar{S}$ be the restriction of the infinite matrix S to the the 1-disc of the extraordinary point as in equation 8.1. By inspection of S, the nonzero eigenvalues of S consist of eigenvalues of $\bar{S}$ and the eigenvalues $\frac{1}{8}$ and $\frac{1}{16}$, each with multiplicity n.

To determine the eigenvalues of $\bar{S}$, delete the first row and column of $\bar{S}$. The resulting matrix C is an n by n whose main diagonal is $\frac{3}{8}$ and whose adjacent diagonals are $\frac{1}{8}$. C is a *circulant* matrix, that is a a matrix in which

$$C_{i,1} = C_{(i+k-1 \ mod \ n)+1,k}$$

for all $0 \le k < n$. The spectral properties of circulant matrices are well-understood.

Theorem 30 *[Dav]*

Let ω be the nth root of unity. Then the circulant matrix C has eigenvalues $\sum_{i=1}^{n} C_{i,1}(\omega^j)^{i-1}$

with associated eigenvector

$$(1, (\omega^j), (\omega^j)^2, ..., (\omega^j)^{n-1})$$

for $1 \leq j \leq n$.

The nth root of unity can be expressed in terms of trigonometric functions as

$$\omega = \cos(\frac{2\pi}{n}) + \sin(\frac{2\pi}{n})\mathbf{i}.$$

The eigenvalue of C associated with ω $(j = 1)$ is

$$\lambda = \frac{3}{8} + \frac{1}{4}\cos(\frac{2\pi}{n}).$$

This eigenvalue appears with multiplicity two since ω^{n-1} also produces the same eigenvalue. In general, the eigenvalue associated with ω^j for $1 < j < n-1$ are real and lie in the range $\frac{1}{8}$ to λ. If we extend the eigenvectors of C associated with these eigenvalues by appending an intial zero, the new vectors

$$(0, 1, (\omega^j), (\omega^j)^2, ..., (\omega^j)^{n-1})$$

are also eigenvector of $\bar{S}$ for all $j \neq n$.

The remaining two eigenvalues of $\bar{S}$ are eigenvalues of the 2 by 2 system

$$\begin{pmatrix} 1 - an & an \\ \frac{3}{8} & \frac{5}{8} \end{pmatrix}$$

The eigenvalues of this matrix are 1 and $\frac{5}{8} - an$. The following theorem summarizes this analysis.

Theorem 31 *For an extraordinary vertex of valence n, the spectrum of S includes 1, λ, λ and $\frac{5}{8} - an$ with the remaining eigenvalues having modulus less than λ.*

If we restrict a to lie in the range

$$\frac{1 - \cos(\frac{2\pi}{n})}{4n} < a < \frac{4 + \cos(\frac{2\pi}{n})}{4n}, \tag{8.2}$$

then the spectrum of S has leading eigenvalues 1, λ, λ. As we shall see in the next section, this condition is sufficient to ensure that the limit surface is a C^1 manifold.

Several specific choices for a suggest themselves. Loop chooses a such that

$$a = \frac{1}{n}(1 - \lambda^2).$$

This choice forces $1 - an = \lambda^2$ and is intended to mimic the spectrum of C^2 subdivision process. For $n > 3$, a simpler choice that avoids the computation of trigonometric functions is

$$a = \frac{3}{8n}.$$

Under this rule, the weight for the extraordinary vertex is always $\frac{5}{8}$. Figure xxxx shows an example of a closed surface produced using this rule.

8.4 C^1 continuity at extraordinary vertices

Let x_1 and x_2 be the eigenvectors of S associated with eigenvalue λ. To define the limit surface produce by S at the extraordinary vertex, we will use x_1 and x_2 to define a parameterization $n = (x_1, x_2)$ associated with the coefficient vector p. However, before proceeding, we must verify that the parameterization given by (x_1, x_2) defines $1-1$ tiling of the parameter plane.

Theorem 32 *If a satisfies the bounds of equation 8.2, then triangulation produce by (x_1, x_2) is a $1-1$ covering of the parameter plane.*

Proof: Since proving this fact is remarkably involved, we sketch the major steps of this proof.

- Consider the annular portion A_k of the triangulation defined by (x_1, x_2) that is the difference of the $2^k - 1$-disc and the $2^{k+1} + 1$-disc centered at the origin. Applying S to A_k and scaling by $\frac{1}{\lambda}$ uniquely determines the next larger annulus A_{k+1} since x_1 and x_2 are eigenvectors of S.

- Find a k such that every 2-disc in A_k is within ϵ of lying on a regular mesh for a sufficiently small ϵ. To see that such a k exists, let p be position of the vertices of a 2-disc in A_k. If we express p in terms of the eigenvectors x_i, $p = \sum a_n x_n$, then subdividing p and scaling by $\frac{1}{\lambda}$ yields

$$\frac{a_0}{\lambda} x_0 + a_1 x_1 + a_2 x_2 + \sum_{i>2} \frac{\lambda_i}{\lambda} a_i x_i.$$

Multiplying $a_0 x_0$ by $\frac{1}{\lambda}$ induces a translation on the 2-disc away from the origin. The next two terms $a_1 x_1 + a_2 x_2$ form a regular 2-disc and are unaffected by subdivision. The magnitude of the remaining terms decreases since $\frac{\lambda_i}{\lambda} < 1$ for $i > 2$. As k goes to infinity, the contribution of these last terms becomes infinitesimal.

- If every 2-disc in A_k is within ϵ of being regular, then no pair of adjacent triangles in A_k can fold back on each other. Therefore, the triangulation in A_k must be $1-1$. A similar argument applies to larger annuli. Small annuli can be checked by hand.

□

Checking whether the triangulation forms a $1-1$ covering of the parameter plane is the equivalent of Reif's Jacobian condition in [Rei94]. We can now precisely characterize the final limit surface. Let $L[n,p](t)$ be the piecewise linear function defined by the parameterization $n = (x_1, x_2)$. We construct a sequence of related parameterizations n^j where $n^0 = n$ and

$$n^{j+1} = \lambda n^j.$$

If we take the limit of the piecewise linear functions associated with these parameterization $L[n^j, p^j](t)$, then the limit surface is

$$F[n,p](t) = \lim_{j \to \infty} L[n^j, p^j](t).$$

Theorem 33 *If a is in range of equation 8.2, then $F[n,p](t)$ is a C^1 function.*

Proof: Away from the origin, $F[n,p](t)$ is a C^2 manifold and can locally be reparameterization using the inverse function theorem to be the graph of a C^2 function. Next, we express $F[n,p](t)$ in terms of $F[n,x_i](t)$ where the x_i are eigenvectors of S (as in theorem 22).

The first three eigenvectors x_0, x_1, and x_2 produce the associated limit functions 1, t_1 and t_2. The eigenvalues for the remaining eigenvectors have magnitude less than λ. A slight generalization of theorem 28 suffices to show that the remaining function $F[n,x_i](t)$ for $i > 2$ are C^1 at the origin. □

Theorem 33 shows that the coordinate functions are C^1 functions. By theorem 29, if the differential of $F[n,p](t)$ has full rank at $t = 0$, then the resulting parametric surface is a C^1 manifold. The entries of this differential are simply the coefficient vectors a_1 and a_2 of the eigenvectors x_1 and x_2 used in the expansion of p. For almost all choices of p, these two vectors are linearly independent.

To conclude, we make a few observation about the current state of research on subdivision methods for closed surfaces. Defining a rule for extraordinary vertices that yields a C^2 manifold for closed surfaces has resisted the best efforts of current researchers. The proof of theorem 33 illustrates the difficulty. The eigenvectors x_1 and x_2 were used to define the parameterization for the coefficient vector p. This choice automatically forced the subdivision method to have linear precision, a necessary condition for a C^1 scheme. However, for such as a scheme to be C^2, it must have quadratic precision. The eigenvectors x_3, x_4, and x_5 must produce quadratic function over the parametrization defined by x_1 and x_2. This is a much more stringent condition that is very difficult to satisfy.

8.5 Subdivision along boundaries

We extended the subdivision methods developed in the functional case to the purely geometric domain. The motivation for this extension was practical, not all geometric objects can be described as the graph of a function. Likewise, realistic geometric objects often have boundaries. We next describe some general techniques for introducing boundaries and measuring their smoothness.

The key to introducing a boundary during subdivision is to use different subdivision rules along the boundary. A simple approach to this process is *tagged* subdivision. Given a control polyhedron p^j, each control point in p^j is tagged as to whether it lies on a face, edge or vertex of the final limit surface. The rules for producing a vertex of the new, refined polyhedron p^{j+1} depend on the tags associated with its ancestors. The tags for the vertices of p^{j+1} also depend on the tags of its ancestors.

The other guiding principal is defining subdivision rules for boundaries is the final limit curves and vertices along the boundaries should depend only on the initial data along the boundaries. This property insures that if two initial polyhedron share the same boundary data, then the corresponding limit objects share the same boundary.

8.5.1 Boundaries for curves

We start with a simple example of a curve segment in two dimensions. For example, consider the control polygon in figure 8.2. The two endpoints of the polygon are tagged as *corner* vertices, they correspond to endpoints of the final limit curve. The remaining vertices are tagged as *edge*vertices, they correspond to interior points of the limit curve. One possible set of subdivision rules for this curve segment is:

- The new corner vertices agree with the old corner vertices.
- A new edge vertex is introduced midway between a corner vertex and its edge neighbor.
- Two edge vertices are introduced $\frac{1}{4}$ and$\frac{3}{4}$ of the way between a pair of adjacent edge vertices.

These rules are exactly the subdivision rules for a uniform quadratic B-spline with double knots at the endpoints. The limit of these subdivision process is a C^1 curve that interpolates the boundary vertices. Since the limit curve interpolates the boundary vertices, connecting two such limit curves requires only that the corresponding boundary curves coincide.

If we restrict our attention to subdivision schemes of the type in chapter 4, the introduction of a boundary corresponds to restriction of the parameter domain to $t \geq 0$. Using this tagged subdivision, the subdivision process is still locally stationary. The subdivision process at a boundary vertex satisfies

$$p^{j+1} = Sp^j.$$

The difference here is p^j is infinite vector with entries p_i^j for $i \geq 0$. The knot vectors n^j are similarly indexed with $n_i^j \geq 0$ for all $i, j \geq 0$.

If the subdivision process produces C^k continuous curves away from the boundary vertex, then necessary and sufficient conditions on S for the limit curve to be C^k continuous at the boundary vertex are very similar to those of section 5.3. Let λ_i be an eigenvalue of S with

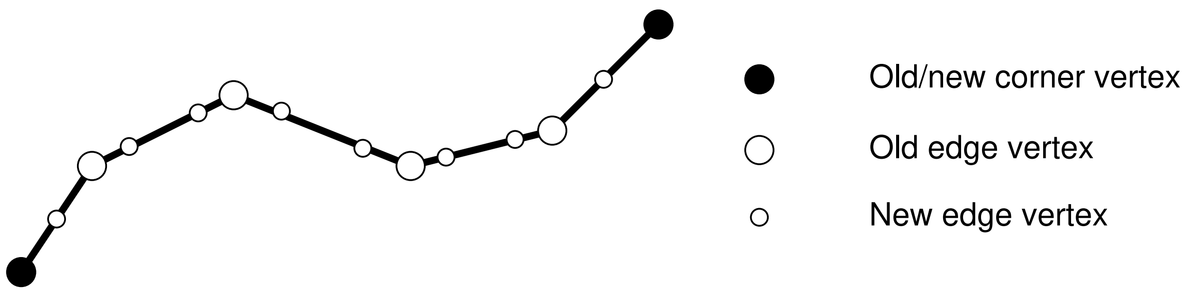

Figure 8.2 Tagged subdivision for a curve segment

associated eigenvector x_i (with $|\lambda_i| \geq |\lambda_{i+1}|$ for all i). If the subdivision process produces a full span of derivatives at the endpoint and linearly independent basis functions, then necessary and sufficient conditions for S to produce C^k limit curves at the boundary vertex are:

1. $\lambda_i = \frac{1}{2^i}$ for all $0 \leq i \leq k$,
2. $|\lambda_i| < \frac{1}{2^k}$ for all $i > k$.
3. The eigenvectors x_0, x_1, ..., x_k reproduce constant multiples of the polynomials 1, t, ..., t^k on the parameter range $t \geq 0$.

The proof of this result follows those of theorems 13 and 14 with the modification that the parameterization vector n spans only half of the parameter domain $t \geq 0$.

8.5.2 Boundaries for surfaces

Tagging can be used to incorporate boundaries in higher dimensions. Each tag reflects the dimension of boundary element that the tagged vertex lies on. In the case of a surface patch, each vertex of a control polyhedron is tagged as a *face* vertex (dimension two), an *edge* vertex (dimension one) or a *corner* vertex (dimension zero). (See figure 8.3 for an example.) New vertices are formed by taking a affine combination of the positions of parent vertices. The tag for this new vertices usually corresponds the highest dimension tag of its parents. For example, an affine combination of edge and corner vertices yields a new edge vertex.

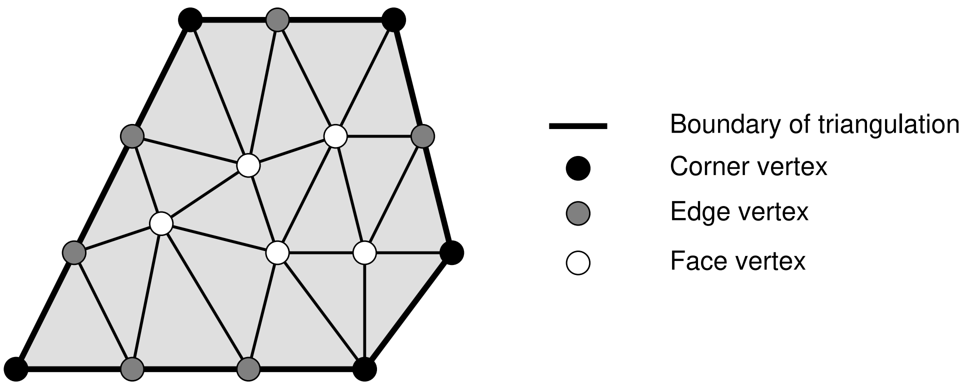

Figure 8.3 Tagged subdivision for a 2D region

Typically, the subdivision rules are chosen so that the patch interpolates its corner vertices. Moreover, the subdivision rules for an edge of the patch are chosen to depend only on the vertices along that edge. For a fixed subdivision scheme, this restriction guarantees that if two patches share the same vertices along an edge, then the corresponding limit surface share a common limit edge.

Analyzing the smoothness of such subdivision schemes involves many variables. If we restrict ourselves to the functional setting of chapter 7, then the analysis of section 7.2.4 is applicable. Let the subdivision process at particular boundary point be locally stationary

$$p^{j+1} = Sp^j,$$

and produce C^k limit functions. If the subdivision scheme produces a space of functions with a full span of derivative of up to order k, then the spectrum of S include the eigenvalue $\frac{1}{2^j}$ with at least multiplicity j for $0 \leq j \leq k$. Moreover, the limit functions corresponding to the associated eigenvectors span the space of all polynomials of degree k.

It is important to note that the spectrum of a C^1 stationary subdivision matrix S need not always have leading eigenvalues $1, \frac{1}{2}, \frac{1}{2}$ or even $1, \lambda, \lambda$. For example, consider a tensor product C^1 subdivision scheme in which binary subdivision is applied along one axis and ternary subdivision is applied along the other axis. The leading eigenvalues for this scheme will be 1, $\frac{1}{2}$, and $\frac{1}{3}$. In general, it is possible to have stationary subdivision scheme is which the parameterization vector n is non-uniformly scaled by a factor of λ in one direction and γ in another direction. Such scheme can be C^1 and have leading eigenvalues 1, λ, and γ.

To show that a set of subdivision rules leads to a C^k limit surface, we fall back on the approach of section 7.3. The idea is to build an appropriate difference scheme and then show that the difference scheme uniformly converges to zero. To illustrate this approach, we extend the irregular C^1 subdivision scheme of section 7.4.2 to allow boundaries. We then prove that the modified rules produce C^1 functions along the boundaries.

Let T^0 be a triangulation whose boundary is simple polygon. As mentioned previously, we tag vertices of the triangulation. Vertices of T^0 interior to T^0 are tagged as being face vertices. Vertices on the boundary of T^0 are tagged as being either edge vertices or corner vertices. An exterior vertex of T^0 is tagged as an edge vertex if its incident exterior edges are colinear. Otherwise, the vertex is tagged as a corner vertex. Figure 8.3 obeys this tagging rule.

The general rule for determining the irregular C^1 subdivision rules was to perturb the rules for linear interpolation by the difference masks generated by pairs of edge adjacent triangles. The subdivision rules for face vertices still follows this rule. There are three types of new subdivision rules for vertices on the boundaries:

- New corner vertices interpolate old corner vertices.
- New edge vertices are introduced midway between adacent pairs of boundary (corner or edge) vertices.
- Old edge vertices are replaced by new edge vertices using the following rule. If v is an edge vertex whose neighbors on the boundary are v_l and v_r, then position the new edge vertex at
$$\frac{1}{4(d_r+d_l)}(d_r v_l + 3(d_l+d_r)v + d_l v_r)$$
where d_r and d_l are the distances from v to v_r and v_l respectively.

The subdivision rules for interior vertices yields a C^1 limit function exactly as characterized in section 7.4.2. Smoothness along the boundary can characterized using similar techniques to those in that section. We first consider the smoothness of the final limit surface on an exterior edge between two vertices of the initial grid T^0. Locally, the subdivision process centered at one these vertices is shown in figure 8.4. Note that the triangulation is locally a three-direction grid. Along the edge, d_r and d_l are equal and therefore, the subdivision rules along that edge are those of a cubic B-spline. The subdivision rules for the interior edges are the three direction, quartic box spline rules. A finite portion of S numbered as in figure 8.4 is
$$\begin{pmatrix}
\frac{3}{4} & \frac{1}{8} & 0 & 0 & \frac{1}{8} & 0 & 0 & 0 \\
\frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 & 0 & 0 & 0 \\
\frac{3}{8} & \frac{1}{8} & \frac{3}{8} & \frac{1}{8} & 0 & 0 & 0 & 0 \\
\frac{3}{8} & 0 & \frac{1}{8} & \frac{3}{8} & \frac{1}{8} & 0 & 0 & 0 \\
\frac{1}{2} & 0 & 0 & 0 & \frac{1}{2} & 0 & 0 & 0 \\
\frac{1}{8} & \frac{3}{8} & \frac{3}{8} & 0 & 0 & \frac{1}{8} & 0 & 0 \\
\frac{1}{8} & 0 & \frac{3}{8} & \frac{3}{8} & 0 & 0 & \frac{1}{8} & 0 \\
\frac{1}{8} & 0 & 0 & \frac{3}{8} & \frac{3}{8} & 0 & 0 & \frac{1}{8}
\end{pmatrix}.$$

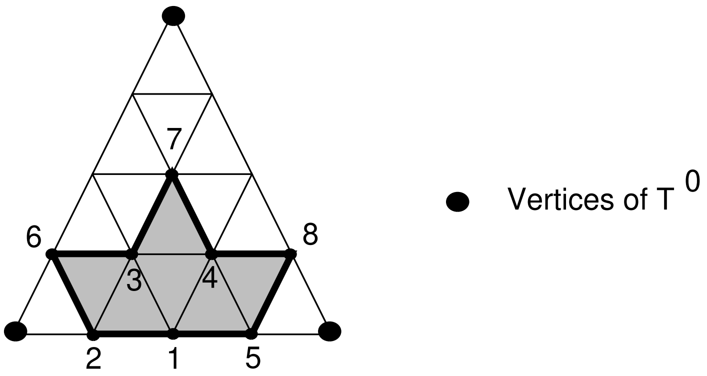

Figure 8.4 Subdivision along the interior of a boundary edge

To show that the limit surface is C^1, we can apply theorem 26. Given the difference rules Δ for this neighborhood

$$\begin{pmatrix} 1 & -1 & -1 & 0 & 0 & 1 & 0 & 0 \\ -1 & 1 & -1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & -1 & -1 & 0 & 0 & 1 & 0 \\ -1 & 0 & 1 & -1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & -1 & -1 & 0 & 0 & 1 \end{pmatrix},$$

we need only show that subdivision matrix D satisfying

$$(2\Delta)S = D\Delta$$

has row norm less than one. Using Mathematica, we can solve for D

$$\begin{pmatrix} \frac{1}{4} & 0 & 0 & \frac{1}{4} & 0 \\ 0 & \frac{1}{2} & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{4} & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{2} & 0 \\ 0 & \frac{1}{4} & 0 & 0 & \frac{1}{4} \end{pmatrix}.$$

The row norm of D is $\frac{1}{2}$. Therefore, the subdivision scheme is C^1 on exterior edge between original vertices of T^0.

At exterior vertices of T^0, the analysis is again similar to that of section 7.4.2. The subdivision process at these vertices is stationary and can be characterized by a subdivision matrix S. The smoothness of the resulting functions can be captured by theorem 28. The limit functions are C^1 if and only if S has leading eigenvalues 1, $\frac{1}{2}$, $\frac{1}{2}$ with remaining eigenvalues of smaller moduli. As in the unbounded case, we have no geometric characterization of when this spectral condition is satisfied.

In the pure geometric case, more general types of subdivision rules are possible along boundaries. Hoppe et al. [HDD$^+$94] give an interesting extension of Loop's method. A simple chain of edges on the boundary of a triangular mesh is tagged. Subdivision rules for cubic B-splines are applied on the interior of this chain. The endpoints are interpolated. Hoppe et al. show that if the standard Loop rules are used for the interior of the mesh, then the resulting surface is C^1 along the resulting boundary.

A chain of boundary edges may also introduce in the interior of the triangular mesh by treating the mesh on each side of a chain of edges as separate meshes. The resulting limit surface has a sharp crease corresponding to the limit curve associated with this chain of edges. Figure 1.5 gives an example of this method applied to a distributor cap. White edges on the initial polyhedron at the right yield sharp creases on the smooth limit surface at the left.

Chapter 9

Multiresolution analysis based on subdivision

Multi-resolution analysis (MRA) produces a hierarchical, orthogonal basis for representing functions. This basis can be used to improve the efficiency of many algorithms for computing with those functions. Traditionally, these basis functions are translates and dilates of a single function. Next, we outline a generalization of MRA to functions defined by subdivision over irregular triangulations.

9.1 Overview

Although the mathematical underpinnings of MRA are somewhat involved, the resulting algorithms are quite simple. We start with a brief intuitive description of how the method can be applied to decompose the polyhedral object shown in Figure 9.1(a).

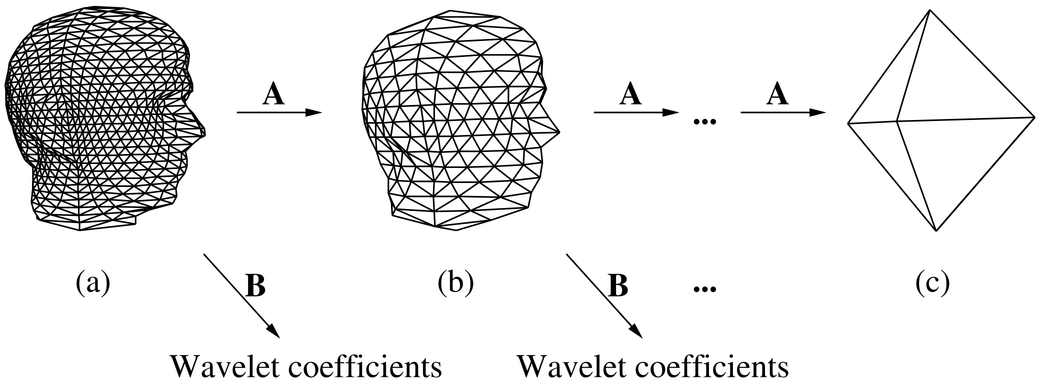

Figure 9.1 (a) Polyhedron in V^4, (b) Projection into V^3, (c) Projection into V^0.

The main idea behind MRA is the decomposition of a object, in this case a polyhedron, into a low resolution part and a "detail" part. The low resolution part of the polyhedron in Figure 9.1(a) is shown in Figure 9.1(b). The vertices in (b) are computed as certain weighted averages of the vertices in (a). These weighted averages essentially implement a low pass filter denoted as **A**. The detail part consists of a collection of fairly abstract coefficients, called wavelet coefficients, that are also computed as weighted averages of the vertices in (a), the weights forming a high-pass filter **B**. The decomposition process, technically called analysis, can be used to further split (b) into an even lower resolution version and corresponding wavelet coefficients. This cascade of analysis steps is often referred to as a filter bank algorithm.

The use of multi-resolution representations for curve editing was recently demonstrated by Finkelstein and Salesin [FS94]. The idea is to allow for changes in the overall sweep of the curve by modifying broad-scale wavelet coefficients; fine-scale changes can similarly be made by modifying only fine-scale wavelet coefficients, as shown in figure 9.2.

9.2 Nested spaces

We next derive the general components of multi-resolution analysis. Traditionally, MRA has been formulated by taking translates and dilates of a single basis function. [?, ?] give a mathematical introduction to this approach. [?] give a more applied introduction. However, our goal to define a variant of MRA that works without resort to translation and dilation. Our motivation in this case is be able to apply MRA to function spaces defined over irregular triangulations.

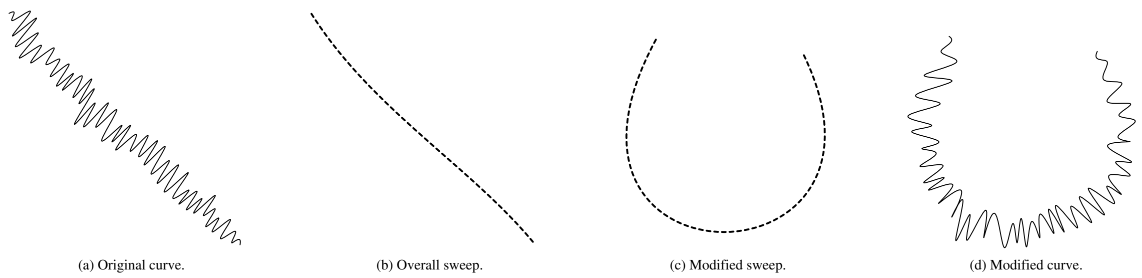

Figure 9.2 Multi-resolution editing

Our starting point is setting of chapter 7. Given an initial triangulation $T = T^0$, we associate a sequence of triangulation T^j related by

$$T^{j+1} = D(T^j)$$

where D splits each triangle into four similar triangles. Associated with each triangulation T^j is a a set of basis functions. These basis functions

$$\phi_i^j(t) = F[n^j, e^i](t)$$

are often referred to as *scaling* functions. For a fixed j, these functions, $\Phi^j(t)$, are the basis functions associated with the jth level of the subdivision process. The span of these basis function is a spline space, V^j,

$$V^j := \text{Span}(\Phi^j(t))$$

The subdivision process forces these spaces to be nested; that is,

$$V^0 \subset V^1 \subset \cdots.$$

The result is a hierarchy of linear spaces defined over the initial triangulation T^0. The basis functions for these spaces are related by the matrix equation:

$$\Phi^j(t) = \Phi^{j+1}(t) S^j \tag{9.1}$$

where S^j is shorthand for $S[T^j]$.

We next wish to form a basis for V^{j+1} that is an extension of $\Phi^j(t)$, the basis for V^j. To this end, we write $\Phi^{j+1}(t)$ in block form as

$$\Phi^{j+1}(t) = (\mathcal{O}^{j+1}(t)\ \mathcal{N}^{j+1}(t)). \tag{9.2}$$

were the $\mathcal{O}^{j+1}(t)$ consists of all scaling functions $\phi_i^{j+1}(t)$ associated with the "old" vertices of T^j and $\mathcal{N}^{j+1}(t)$ consists of the remaining scaling functions associated with the "new" vertices of T^{j+1} added at midpoints of edges of T^j. Equation 9.1 can also be expressed in block form:

$$\Phi^j(t) = (\mathcal{O}^{j+1}(t)\ \mathcal{N}^{j+1}(t)) \begin{pmatrix} S_{\mathcal{O}}^j \\ S_{\mathcal{N}}^j \end{pmatrix}. \tag{9.3}$$

Instead of using $\mathcal{O}^{j+1}(t)$ and $\mathcal{N}^{j+1}(t)$ as the basis for V^{j+1}, we use $\Phi^j(t)$ and $\mathcal{N}^{j+1}(t)$ as the new basis. This basis is *hierarchical* since a function $f^{j+1}(t)$ in V^{j+1} is expressed as

$$f^{j+1}(t) = f^j(t) + f_*^j(t)$$

where $f^j(t)$ is in the span of $\Phi^j(t)$ and $f^j_*(t)$ is in span of $\mathcal{N}^{j+1}(t)$. Using this hierarchical representation, projection of $f^{j+1}(t)$ in to the lower detail space V^j consists of forming $f^j(t)$.

9.3 Orthogonal spaces

Hierarchical bases provide a convenient means of building multi-resolution approximations to a function. Recall that the goal of MRA is to provide a low resolution version of the object that is a good approximation to the original object with the magnitude of each wavelet coefficient measuring the error introduced by that coefficient. If the "detail" space is orthogonal to the low resolution space, then the low resolution approximation is "best" in a least squares sense. Let us briefly explain why.

The inner product of a pair of functions f, g is

$$\langle f, g\rangle := \int_t f(t)g(t)dt$$

Given a high resolution space V^{j+1} and a low resolution space V^j, let the "detail" space be the space orthogonal to V^j in V^{j+1},$V^j_\perp$,

$$V^j_\perp \quad = \quad \{f \in V^{j+1} \ \mid \langle f, g\rangle = 0 \ \ \forall g \in V^j\}.$$

For $f^{j+1}(t)$ is in V^{j+1}, denote the projection of $f^{j+1}(t)$ into the space V^j and $V^j_\perp$,

$$f^{j+1}(t) = f^j(t) + f^j_\perp(t).$$

$f^j(t)$ is the best approximation to $f^{j+1}(t)$ in the sense that it minimizes the least squares residual

$$\langle f^{j+1} - f^j, f^{j+1} - f^j\rangle.$$

To ensure the "best" projection in V^j, we orthogonalize our hierarchical basis. Specifically, we replace the basis functions $\mathcal{N}^{j+1}(t)$ by their projection into $V^j_\perp$. The resulting functions $\Psi^j(t)$ form a basis for $V^j_\perp$. Expressed in matrix form,

$$\Psi^j(t) = \mathcal{N}^{j+1}(t) - \Phi^j(t)\boldsymbol{\alpha}^j. \tag{9.4}$$

The resulting functions $\Psi^j(t)$ are *pre-wavelets* since they span $V^j_\perp$ but they are not mutually orthogonal. If $f^{j+1}(t)$ is expanded in terms of the $\Phi^j(t)$ and the $\Psi^j(t)$, then the restriction of $f^{j+1}(t)$ to the $\Phi^j(t)$ is guaranteed to be the best approximation to $f^{j+1}(t)$ in V^j in a least squares sense.

The coefficients $\boldsymbol{\alpha}^j$ are the solution to the linear system formed by taking the inner product of each side of equation 9.4 with $\Phi^j(t)$.

$$\begin{aligned} \langle \Phi^j(t), \Phi^j(t) \rangle \boldsymbol{\alpha}^j &= \langle \Phi^j(t), \mathcal{N}^{j+1}(t) \rangle, \\ &= (S^j)^T \langle \Phi^{j+1}(t), \mathcal{N}^{j+1}(t) \rangle. \end{aligned} \tag{9.5}$$

The second line follows from the first by equation 9.1 and the linearity of inner products. $\langle \Phi^j(t), \Phi^j(t) \rangle$ is a matrix whose entries are inner products of pairs of elements in $\Phi^{j+1}(t)$. $\langle \Phi^{j+1}(t), \mathcal{N}^{j+1}(t) \rangle$ is a similar matrix. [DLW94] give a direct method for computing entries of these matrices.

9.4 Filter banks

The analysis filters and their inverse synthesis filters can be conveniently expressed using block matrix equations. Let $\Psi^j(t)$ denote the row matrix of pre-wavelets spanning $V_\perp^j$. Expand $\Phi^{j+1}(t)$ into $(\mathcal{O}^{j+1}(t)\ \mathcal{N}^{j+1}(t))$ as in equation 9.2. By equations 9.3 and 9.4, these bases must related by:

$$\left(\Phi^j(t)\ \Psi^j(t) \right) = \left(\mathcal{O}^{j+1}(t)\ \mathcal{N}^{j+1}(t) \right) \begin{pmatrix} S_\mathcal{O}^j & -S_\mathcal{O}^j \boldsymbol{\alpha}^j \\ S_\mathcal{N}^j & \mathbb{1} - S_\mathcal{N}^j \boldsymbol{\alpha}^j \end{pmatrix}. \tag{9.6}$$

The synthesis filters S^j and Q^j are the columns of the change of basis matrix. The rows of the inverse of this matrix are exactly the analysis filters A^j and B^j.

From a practical standpoint, it is critical that the analysis and synthesis matrices are sparse. To achieve linear time decomposition and reconstruction, they must each have a constant number of non-zero entries in each row. If S^j and $\boldsymbol{\alpha}^j$ are sparse, then Q^j is sparse. Unfortunately, the analysis filters derived from the inverse of the matrix need not be sparse. For interpolating subdivision schemes such as linear subdivision and the C^1 "butterfly" scheme of Dyn et. al. [DGL90], the situation is much improved. Such interpolating schemes have the property that $S_\mathcal{O}^j$ is exactly the identity matrix. In this case, equation 9.6 simplifies greatly. The resulting synthesis filters are:

$$\begin{pmatrix} S^j & Q^j \end{pmatrix} = \begin{pmatrix} \mathbb{1} & -\boldsymbol{\alpha}^j \\ S_\mathcal{N}^j & \mathbb{1} - S_\mathcal{N}^j \boldsymbol{\alpha}^j \end{pmatrix}.$$

The inverse analysis filters A^j and B^j are:

$$\begin{pmatrix} A^j \\ B^j \end{pmatrix} = \begin{pmatrix} \mathbb{1} - \boldsymbol{\alpha}^j S^j_{\mathcal{N}} & \boldsymbol{\alpha}^j \\ -S^j_{\mathcal{N}} & \mathbb{1} \end{pmatrix}$$

If S^j and $\boldsymbol{\alpha}^j$ are sparse, then all of these filters are also sparse. The situation is less desirable for B-spline like schemes such as Loop's scheme and Catmull-Clark surfaces. For these schemes, the synthesis filters are sparse, but the analysis filters are dense. Making these schemes efficient for multiresolution analysis is a topic of future research.

Having determined the analysis filters, they can be used to decompose a function $f^{j+1}(t)$ in V^{j+1} given by

$$f^{j+1}(t) = \sum_i f_i^{j+1} \phi_i^{j+1}(t) \tag{9.7}$$

into a lower resolution part in V^j plus a detail part in $V^j_\perp$

$$f^{j+1}(t) = \sum_i f_i^j \phi_i^j(t) + \sum_i g_i^j \psi_i^j(t)$$

as follows. Let F^j and G^j denote the matrices of coefficients corresponding to the f_i^j and the g_i^j. We now write Equation 9.7 in matrix form and substitute the definition of the analysis filters:

$$\begin{aligned} f^{j+1}(t) &= \Phi^{j+1}(t)\; F^{j+1} \\ &= \begin{pmatrix} \Phi^j(t) \; \Psi^j(t) \end{pmatrix} \begin{pmatrix} A^j \\ B^j \end{pmatrix} F^{j+1} \\ &= \Phi^j(t) A^j\; F^{j+1} + \Psi^j(t) B^j\; F^{j+1} \end{aligned}$$

and therefore

$$F^j = A^j\; F^{j+1} \qquad\qquad G^j = B^j\; F^{j+1}.$$

Of course, the analysis filters A^{j-1} and B^{j-1} can now be applied to F^j to yield F^{j-1} and G^{j-1} and so on. A similar argument shows that F^{j+1} can be recovered from F^j and G^j using the synthesis filters:

$$F^{j+1} = S^j\; F^j + Q^j\; G^j.$$

Bibliography

[Bar93] M. Barnsley. *Fractals Everywhere.* Academic Press, 1993.

[CC78] E. Catmull and J. Clark. Recursively generated b-spline surfaces on arbitrary topological meshes. *Computer-aided Design*, 10:350–355, 1978.

[CDM91] A. Cavaretta, W. Dahmen, and C Micchelli. *Stationary Subdivision*, volume 453 of *Memoirs of the AMS.* AMS, 1991.

[Dav] P. Davis. *Circulant Matrices.*

[dBHR93] C. de Boor, K. Höllig, and S. Riemenschneider. *Box Splines.* Springer Verlag, 1993.

[DGL87] N. Dyn, J. Gregory, and D. Levin. A 4-point interpolatory subdivision scheme for curve design. *Computer Aided Geometric Design*, 4:257–268, 1987.

[DGL90] N. Dyn, J. Gregory, and D. Levin. A butterfly subdivision scheme for surface interpolation with tension control. *ACM Transactions on Graphics*, 9:160–169, 1990.

[DGL91] N. Dyn, J. Gregory, and D. Levin. Analysis of uniform binary subdivision schemes for curve design. *Constructive Approximation*, 7:127–147, 1991.

[DGL94] N. Dyn, J. Gregory, and D. Levin. Piecewise uniform subdivision schemes. in preparation, 1994.

[DLW94] T. DeRose, M. Lounsbery, and Joe Warren. Multiresolution analysis for surfaces of arbitrary topological type. Technical Report 93-10-05b, University of Washington, 1994.

[DS78] D. Doo and M. Sabin. Behavior of recursive subdivision of surfaces near extraordinary points. *Computer-aided Design*, 10:356–360, 1978.

[Dyn92] N. Dyn. Subdivision schemes in computer aided geometric design. In W. Light, editor, *Advances in Numerical Analysis II*, pages 36–104. Oxford University Press, 1992.

[FS94] Adam Finkelstein and David Salesin. Multiresolution curves. *Computer Graphics*, 28(3), 1994.

[Goo90] T. Goodman. Polyhedral splines. In W. Dahmen, M. Gasca, and C. Micchelli, editors, *Computation of curves and surfaces*, pages 347–382. Kluwer Academic, 1990.

[GQ92] J. A. Gregory and R. Qu. A subdivision algorithm for non-uniform b-splines. In S. Singh, editor, *Approximation Theory, Spline Functions and Applications*, pages 423–436. NATO ASI Series C, 1992.

[GR94] J. A. Gregory and R.Qu. Non-uniform corner cutting. *Computer Aided Geometric Design*, 1994. To appear.

[HDD+94] H. Hoppe, T. DeRose, T. Duchamp, M. Halstead, H. Jin, J. McDonald, J. Schweitzer, and W. Stuetzle. Piecewise smooth surface reconstruction. *Computer Graphics*, 28:295–302, 1994.

[Loo87] C. Loop. Smooth subdivision based on triangles. Master's thesis, University of Utah, 1987.

[LR80] J. M. Lane and R. F. Riesenfeld. A theoretical development for the computer generation and display of piecewise polynomial surfaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-2(1):35–46, January 1980.

[Rei94] U. Reif. A unified approach to the analysis of subdivision schemes. To appear in CAGD, 1994.

[Rie75] R. F. Riesenfeld. On Chaikin's algorithm. *Computer Graphics and Image Processing*, 4:304–310, 1975.

[Tay55] A. Taylor. *Advanced Calculus*. Blaisdell Publishing, 1955.