



ELSEVIER

Computer Aided Geometric Design 12 (1995) 153–174

COMPUTER
AIDED
GEOMETRIC
DESIGN

A unified approach to subdivision algorithms near extraordinary vertices

Ulrich Reif*

Mathematisches Institut A, Universität Stuttgart, Pfaffenwaldring 57, 70511 Stuttgart, Germany

Received April 1993; revised January 1994

Abstract

We present a unified approach to subdivision algorithms for meshes with arbitrary topology which admits a rigorous analysis of the generated surface and give a sufficient condition for the regularity of the surface, i.e. for the existence of a regular smooth parametrization near the extraordinary point. The criterion is easily applicable to all known algorithms such as those of Doo–Sabin and Catmull–Clark, but will also be useful to construct new algorithms like interpolatory subdivision schemes.

Keywords: Subdivision; Arbitrary topology; Regular surface; Characteristic map; Extraordinary vertex

1. Introduction

Subdivision algorithms are a powerful tool in CAGD and have been well understood for tensor product surfaces for a long time. Usually, the principle is to describe a given surface over a refined control net where “refined” means in general that the scale is bisected. When a subdivision process is iterated, the control net will contain more and more vertices which converge under well-known conditions to the originally described surface. On the other hand, one can define a subdivision process by an arbitrary linear map in the space of control points without knowing something a priori about a surface. If the sequence of control nets converges in a certain sense, one can use this procedure to generate surfaces.

A combination of these two ideas can help to deal with the problem of filling an n -sided hole or, equivalently, to define surfaces over meshes with arbitrary topology. The first algorithms of this type were given by Doo and Sabin (1978) and by Catmull

* Email: reif@mathematik.uni-stuttgart.de.

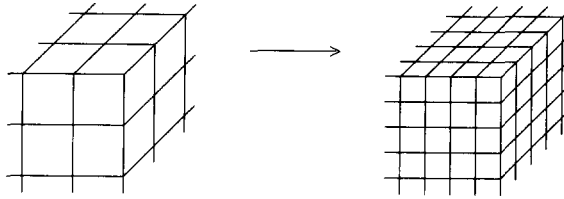


Fig. 1. Refining a mesh by subdivision.

and Clark (1978). The idea is the following: in general, a control net will consist of large regular regions, where standard C -conditions can be used to describe a smooth surface, and isolated irregular regions, where these conditions are too restrictive. Thus, the resulting surface has gaps, also called n -sided holes, corresponding to the irregular regions and these gaps can be filled gradually by subdivision. To achieve this, the key observation is that subdivision enlarges the regular regions of the control net and shrinks the irregular regions. Number and type of the irregular regions are not changed, but the scale is halved, cf. Fig. 1.

Now it is required that the subdivision process reproduces the original surface completely on the common domain. But, moreover, since parts of the former irregular regions are now regular, it is possible to prolong the original surface. In the space of control points, this means that most of the new ones are determined uniquely from the reproduction condition, but those corresponding to the new parts of the surface can be chosen rather arbitrarily. Usually, they are computed as a linear combination of old control points near the irregular region. The associated matrix is called the *subdivision matrix* and it turns out that a proper choice of this matrix leads in the limit to smooth surfaces when the subdivision process is iterated.

Doo–Sabin and Catmull–Clark suggested particular subdivision matrices where the coefficients were determined by some ad hoc generalizations of the regular case and after that some modified algorithms were proposed. But not before 1986, Ball and Storry (1986; 1988) made a first serious attempt to prove convergence to smooth surfaces for subdivision schemes of Catmull–Clark type. However, a detailed analysis of their proof reveals that it is incomplete as it is exclusively based on the analysis of the eigenproperties of the subdivision matrix and does not take the properties of the basis functions into account. But this is indispensable as will be shown in particular by the examples in the fourth section. The point is that the process of normalizing the sequence of normal vectors can fail if a certain regularity condition is violated. But even if this gap is filled (which can be done), the problem remains that Ball and Storry only try to prove tangent plane continuity near the extraordinary point. However, the existence of a limit for the normal vectors is a rather weak smoothness category since local self-intersections of the surface are still possible. Smoothness in the sense of differential geometry (and, of course, from the point of view of the designer) requires the existence of a regular smooth parametrization for the surface near the point in question.

Recently, Halstead et al. (1993) used Catmull–Clark surfaces to construct surfaces which interpolate the vertices of a given mesh of arbitrary topological type subject to the minimization of a certain fairness functional. In this paper, a short proof for the

smoothness of the surfaces is given, but the same remarks as for the Ball and Storry proof apply.

In this paper, we give a new unified approach to the topic which will be carried out for surfaces based on quadrilaterals, but will work in exactly the same way for surfaces based on triangles. Our main result is that all properties of a subdivision algorithm, especially the existence of a smooth regular parametrization for the generated surface near the limit point, can be detected from the leading eigenvalues of the subdivision matrix and an associated characteristic map.

The paper is organized as follows: In Section 2, we formulate the problem and define the setup; in Section 3, we present some theorems on the smoothness properties of the generated surfaces; and in Section 4, we give an example. In the appendix, we outline the proof of two lemmas needed in Section 3. Although they are crucial for the proof of our main theorem, the proofs were diverted to the appendix, since they are somewhat technical.

2. Definitions and formulation of the problem

In practice, subdivision is carried out in the space of control points corresponding to some basis functions. Nevertheless, it is more convenient to start our considerations in the space of patches which make up the surface.

Let X be a parametrically smooth spline surface¹ consisting of rectangular patches $x^j, j \in J$, with a single n -sided hole. For convenience, we shall assume that X is an infinite surface with the n -sided hole as its only boundary in order to avoid a formal discussion of the outer boundary of the surface. The *parameter space* $\Omega := \omega \times J$ consists of $\#J$ copies of the unit square ω provided with a neighborhood relation. A *representation* of Ω is a map

$$\Phi : (\omega, j) \mapsto \Phi^j(\omega) \in \mathbb{R}^2, \quad j \in J. \quad (2.1)$$

from Ω to \mathbb{R}^2 , where the functions Φ^j are smooth injective maps in \mathbb{R}^2 with the property that pieces of boundary curves identified by the neighborhood relation have a common image. Otherwise, the image is supposed to be unique. The resulting mesh $\Gamma := \Phi(\Omega)$ is *regular*, what means that every vertex in the interior of Γ is shared by exactly four patches. It is well known that this topology admits the generation of smooth surfaces with parametrically smoothness conditions between the patches. On the other hand, a mesh Γ as shown in Fig. 2 is maximal in the sense that one cannot add further patches inside the n -sided hole without breaking symmetry or regularity.

One possibility to fill the gap is the application of a *subdivision process* S : We replace every ω by four new squares and obtain a new parameter space $\tilde{\Omega}$; the new neighborhood relations are obvious, see Fig. 3.

The new mesh $\tilde{\Gamma} := \tilde{\Phi}(\tilde{\Omega})$ generated by some representation $\tilde{\Phi}$ has an important property: it can be prolonged by a layer $\text{pr}(\Gamma)$ of patches such that the mesh

¹ We do not distinguish explicitly between a surface viewed as a map or as a point set in \mathbb{R}^3 . It will be clear from the context what is meant.

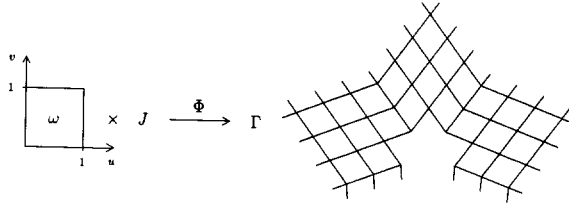


Fig. 2. Representation of the parameter space.

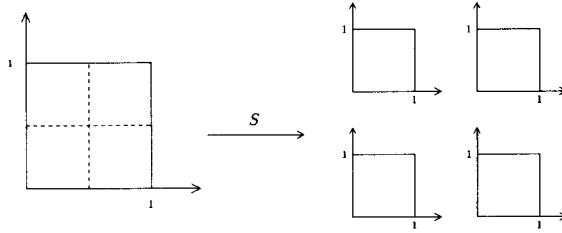


Fig. 3. Subdivision of the parameter space.

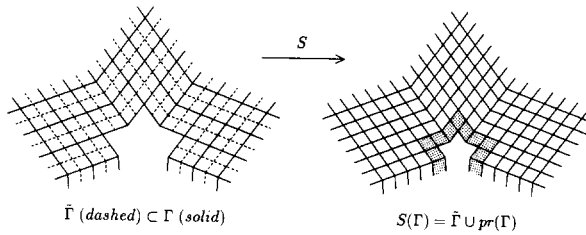


Fig. 4. Subdivision and prolongation.

$$S(\Gamma) := \tilde{\Gamma} \cup \text{pr}(\Gamma) \tag{2.2}$$

is still regular, cf. Fig. 4. In fact, it has exactly the same topological structure as the original mesh Γ . Thus, we can think of $S(\Gamma) =: S(\Phi)(\Omega)$ as a new representation $S(\Phi)$ of the original parameter space Ω reflecting the fact that the parameter space contains only information on the topology but not on the scale of the mesh.

Now it will be possible to find a parametrically smooth surface $S(X)$ over $S(\Gamma)$ with the property that the restriction $\tilde{X} := S(X)|_{\tilde{\Gamma}}$ of $S(X)$ to the mesh $\tilde{\Gamma}$ coincides with the original surface X (viewed as point sets in \mathbb{R}^3). The new part of the surface

$$\text{pr}(X) := S(X) \setminus X = S(X)|_{\text{pr}(\Gamma)} \tag{2.3}$$

is called a *prolongation* of X . Of course, it is not unique, since the only requirement a priori is that it joins smoothly with the original surface X , but once an algorithm to compute the prolongation has been established, it will not be changed (in a sense that will become clearer soon), when the subdivision process is iterated.

To keep things as simple as possible, assume that

$$\text{pr}(S(X)) = \text{pr}(\text{pr}(X)) = \text{pr}^2(X), \tag{2.4}$$

in other words, the prolongation of a surface should depend only on the inner layer of patches of the given surface. This is *not* an essential loss of generality since one can think of any finite number of layers of patches as a single layer of macro patches.

When the subdivision process is iterated, more and more new layers arise which, in the limit, will fill the n -sided hole completely if the prolongation procedure is properly chosen. The object in question is the surface

$$P := \bigcup_{m \in \mathbb{N}} S^m(X) \setminus X = \bigcup_{m \in \mathbb{N}} \text{pr}^m(X), \quad (2.5)$$

whose properties must be analyzed in dependence on the particular subdivision process. To this end, we start with some definitions.

Definition 2.1. A subdivision procedure S is said to be *convergent*, if there is a unique limit point p with

$$\lim_{m \rightarrow \infty} x_m = p$$

for any sequence of points $x_m \in \text{pr}^m(X)$.

This means that the closure

$$\bar{P} := \text{closure}(P) = P \cup p \quad (2.6)$$

is a surface without gaps.

Definition 2.2. The surface \bar{P} is called *tangent plane continuous*, if S converges and if there is a unique limit $n(p)$ for any sequence of normal vectors

$$\lim_{m \rightarrow \infty} n(x_m) = n(p), \quad x_m \in \text{pr}^m(X).$$

Remark. $n(p)$ is just the name of the limit of the normal vectors and not necessarily the normal vector of \bar{P} at p which might not even exist.

Definition 2.3. The surface P is called *regular* at p , if there is a regular smooth parametrization of \bar{P} near p .

It is important to notice that regularity is essentially more than tangent plane continuity. Interestingly enough, the question of regularity has (as far as we know) never been investigated for subdivision algorithms. To demonstrate the difference, consider the surface F shown in Fig. 5. This surface is tangent plane continuous with $\lim_{r \rightarrow 0} n(r, t) = (0, 0, 1)$, but not regular at the origin since the projection of F on the xy -plane is not injective.

Now we turn towards a more precise description of the surface P : First, note that every prolongation $\text{pr}^m(X)$ has the same parameter domain $\hat{\Omega} =: \omega \times \hat{J}$, where $\hat{J} \subset J$ is the subset of the index set J corresponding to the inner layer of patches, cf. Fig. 6. With the notations from above, we have $\hat{\Omega} = (S^m(\Phi))^{-1}(\text{pr}^m(\Gamma))$ for all $m \in \mathbb{N}$.

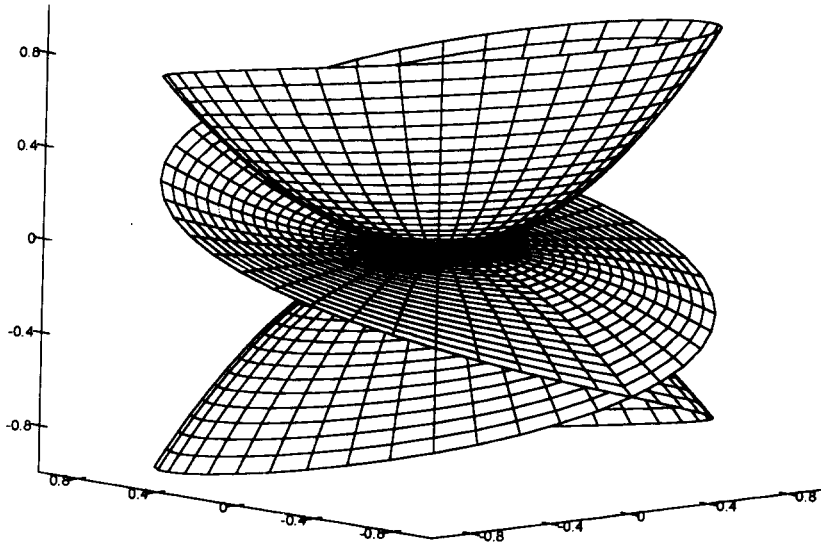


Fig. 5. $F : (r, t) \mapsto (r \cos 3t, r \sin 3t, r^2 \sin t)$, $(r, t) \in [0, 1] \times [0, 2\pi)$.

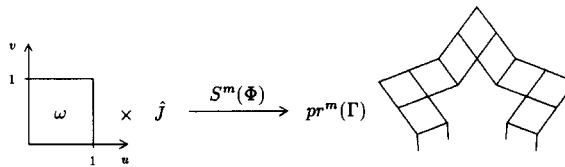


Fig. 6. Representation of the prolongation.

Further, it is required that the surfaces $pr^m(X)$ can be written in the form

$$pr^m(X) : (u, v, j) \mapsto \mathbf{x}_m^j(u, v) := \sum_{k=1}^K b_k(u, v, j) \mathbf{B}_m^k, \quad (u, v) \in \omega, \quad j \in \hat{J},$$

or in vector notation

$$\mathbf{x}_m^j(u, v) = b(u, v, j) \mathbf{B}_m \tag{2.7}$$

with a row vector $b(u, v, j)$ of functions and a column vector \mathbf{B}_m of points² in \mathbb{R}^3 .

The functions b_k are supposed to lie in $C^1(\hat{\Omega})$, the space of parametrically smooth functions over $\hat{\Omega}$ and to form a partition of unity, i.e.

$$\sum_{k=1}^K b_k(u, v, j) \equiv 1. \tag{2.8}$$

² The points itself are row vectors with three entries for the coordinates of the point. Thus, \mathbf{B}_m can be viewed either as a column vector of points in \mathbb{R}^3 or as a $K \times 3$ -matrix of reals.

We refer to the b_k as *basis functions* and to the $\mathbf{B}_m^k \in \mathbb{R}^3$ as *control points*. The crucial point is that the basis functions do not depend on m . This seems to be very restrictive at first sight, but many ways of generating surfaces, and in particular B-splines, fit into this concept.

Further, we assume that the subdivision process is linear and stationary, i.e. it is a linear map in the space of control points and does not depend on m . We write in vector notation

$$\mathbf{B}_{m+1} = A\mathbf{B}_m, \quad \mathbf{B}_m = A^m\mathbf{B}_0 \quad (2.9)$$

with some matrix A which is called the *subdivision matrix*. A basic property of A is that all rows have sum 1, since the subdivision process is supposed to be affine invariant. Then some entries of A are determined by the condition that consecutive layers must join smoothly, but other ones can be chosen arbitrarily reflecting the fact that the prolongation of a given surface is not unique. Usually, there are some other requirements on A such as positivity or periodicity, but we do not need them in the further analysis. The problem to be solved is to find easily applicable criteria correlating a given subdivision matrix with the regularity properties of the generated surfaces.

3. Sufficient smoothness conditions

In this section we shall show that the smoothness properties of the surface \mathbf{P} can be detected from the leading eigenvalues of the subdivision matrix A and an associated characteristic map Ψ . We start with a simple lemma from linear algebra.

Lemma 3.1. *Denote the eigenvalues of A by $\lambda_1, \lambda_2, \dots, \lambda_K$ with*

$$|\lambda_1| \geq |\lambda_2| \geq \dots \geq |\lambda_K|$$

and let v_1, v_2, \dots, v_K be a system of real vectors spanning the corresponding invariant subspaces. If $|\lambda_p| > |\lambda_{p+1}|$, then

$$A^m v_q = o(\lambda_p^m) \quad \text{as } m \rightarrow \infty \quad \text{for all } q > p.$$

Remark. Omitting for simplicity the usual notation of the absolute value in the argument of the order function, a sequence r_1, r_2, \dots in \mathbb{R}^K is said to be of order $o(\lambda^m)$, if

$$\lim_{m \rightarrow \infty} \frac{\|r_m\|}{|\lambda|^m} = 0$$

with respect to some norm $\|\cdot\|$ in \mathbb{R}^K . For sequences of functions the convergence is always uniform since all domains are compact in this context.

Proof. The proof follows immediately from the fact that there is a norm in \mathbb{R}^{K-p} , with respect to which the associated matrix norm differs from the spectral radius of the matrix by an arbitrarily small $\varepsilon > 0$. \square

Now we give a first theorem which provides a sufficient criterion for the convergence of a subdivision algorithm.

Theorem 3.2. *A subdivision algorithm converges, if $1 = \lambda_1 > |\lambda_2|$.*

Proof. All rows of A sum up to 1, so $\lambda_1 = 1$ is always an eigenvalue of A and $v_1 = [1, \dots, 1]^T$ is the corresponding eigenvector. We can write the sequence of vectors of control points in the form

$$B_0 = \sum_{i=1}^K v_i p_i, \quad B_m = v_1 p_1 + o(1) \quad (3.1)$$

with coefficients $p_1, \dots, p_K \in \mathbb{R}^3$ and obtain

$$\begin{aligned} x_0^j(u, v) &= \sum_{i=1}^K b(u, v, j) v_i p_i, \\ x_m^j(u, v) &= b(u, v, j) v_1 p_1 + o(1) = p_1 + o(1), \end{aligned} \quad (3.2)$$

where we have used that the functions in $b(u, v, j)$ form a partition of unity. Since the order function converges uniformly in (u, v) , we finally obtain

$$\lim_{m \rightarrow \infty} x_m^j(u, v) = p_1 \quad (3.3)$$

and $p := p_1$ is the limit point in the sense of Definition 2.1. \square

It turns out that the further smoothness properties of the surface \bar{P} can be derived from the leading eigenvalues of A and a map Ψ from $\hat{\Omega}$ to \mathbb{R}^2 which depends only on the corresponding eigenvectors of A and the basis functions. Therefore, Ψ is called the *characteristic map* of the subdivision process.

Definition 3.3. For a subdivision matrix A with $1 > |\lambda_2| = |\lambda_3| > |\lambda_4|$, the *characteristic map* $\Psi : \hat{\Omega} \mapsto \mathbb{R}^2$ is defined by

$$\Psi : (u, v, j) \mapsto b(u, v, j) V := b(u, v, j) [v_2, v_3],$$

where V is a $K \times 2$ matrix with the vectors v_2 and v_3 as its columns. Ψ is called *regular*, if $\Delta(u, v, j) := \det \partial \Psi(u, v, j) / \partial(u, v) \neq 0$ for all $(u, v, j) \in \hat{\Omega}$.

Note that the characteristic map can be viewed as a two-dimensional spline function with two-dimensional control points given by the rows of the matrix V . In Section 4, this fact will be used for visualization purposes.

Of course, the characteristic map is subject to the ambiguity in the choice of the vectors v_2 and v_3 , however, all of its crucial properties are well defined. This is shown in the following lemma.

Lemma 3.4. (i) *Injectivity and regularity of the characteristic map do not depend on the particular choice of the vectors v_2 and v_3 .*

(ii) If Ψ is regular, then

$$\mu(\Delta) := \inf_{\hat{\Omega}} |\Delta| > 0.$$

Proof. Denote by $\tilde{V} := [\tilde{v}_2, \tilde{v}_3]$ another feasible matrix. Its columns span the same linear space as those of the given matrix V and so it can be written as

$$\tilde{V} = VT$$

with T some regular 2×2 -matrix. This implies

$$\tilde{\Psi}(u, v, j) := b(u, v, j) \tilde{V} = \Psi(u, v, j)T \quad (3.4)$$

and

$$\tilde{\Delta}(u, v, j) = \Delta(u, v, j) \det T. \quad (3.5)$$

Now, the first statement follows from (3.4) and (3.5) whereas the second statement is an immediate consequence of the continuity of Δ for fixed $j \in \hat{J}$ and the compactness of ω . \square

The next theorem gives a sufficient condition for the tangent plane continuity of a subdivision algorithm.

Theorem 3.5. *If $\lambda_2 = \lambda_3$, $1 > |\lambda_2| > |\lambda_4|$, is a real eigenvalue with algebraic and geometric multiplicity 2 and if the characteristic map is regular, then \bar{P} is tangent plane continuous for almost every initial vector B_0 of control points.*

Proof. We have

$$B_m = v_1 p_1 + \lambda_2^m (v_2 p_2 + v_3 p_3) + o(\lambda_2^m) \quad (3.6)$$

and

$$\begin{aligned} x_m^j(u, v) &= p_1 + \lambda_2^m b(u, v, j) (v_2 p_2 + v_3 p_3) + o(\lambda_2^m), \\ x_{m,u}^j(u, v) &= \lambda_2^m b_u(u, v, j) (v_2 p_2 + v_3 p_3) + o(\lambda_2^m), \\ x_{m,v}^j(u, v) &= \lambda_2^m b_v(u, v, j) (v_2 p_2 + v_3 p_3) + o(\lambda_2^m). \end{aligned} \quad (3.7)$$

For the cross product of the partial derivatives we find

$$x_{m,u}^j \times x_{m,v}^j = \lambda_2^{2m} (\Delta(u, v, j) p_2 \times p_3 + o(1)). \quad (3.8)$$

Δ is nonzero and the cross product of p_2 and p_3 has a positive norm for almost every choice of control points B_0 . So, the normalized normal vector is given by

$$n_m(u, v, j) = \frac{x_{m,u}^j \times x_{m,v}^j}{\|x_{m,u}^j \times x_{m,v}^j\|} = \frac{p_2 \times p_3}{\|p_2 \times p_3\|} + \frac{o(1)}{|\Delta(u, v, j)| \|p_2 \times p_3\|} \quad (3.9)$$

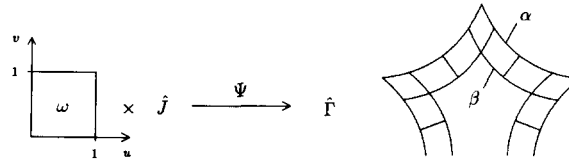


Fig. 7. Representation by the characteristic map Ψ .

and we obtain for the limit

$$\lim_{m \rightarrow \infty} \mathbf{n}_m^j(u, v) = \frac{\mathbf{p}_2 \times \mathbf{p}_3}{\|\mathbf{p}_2 \times \mathbf{p}_3\|} =: \mathbf{n}(\mathbf{p}) \tag{3.10}$$

since the order function converges uniformly in (u, v, j) and $|\Delta(u, v, j)| \geq \mu(\Delta) > 0$ by Lemma 3.4. \square

Theorem 3.5 is the generalization of some results for particular types of basis functions and subdivision matrices, but the indispensable regularity condition for the characteristic map has not yet been established. A much more significant gap in the theory of subdivision algorithms is filled by the following theorem which is our main result. It gives a sufficient condition for the regularity of the surface \bar{P} , i.e. the existence of a regular smooth parametrization for \bar{P} near the limit point \mathbf{p} .

Theorem 3.6. *If $\lambda_2 = \lambda_3$, $1 > |\lambda_2| > |\lambda_4|$, is a real eigenvalue with algebraic and geometric multiplicity 2 and if the characteristic map is regular and injective³, then P is regular at \mathbf{p} for almost every initial vector \mathbf{B}_0 of control points.*

The proof of the theorem is essentially based on the following two lemmas, the proofs of which are postponed to the appendix, since they are somewhat technical and would disturb the line of thoughts here. The first lemma is used to overcome the difficulty that the proof of the theorem requires some tools from topology and functional analysis which are not available in the present setup. We have no topology on $\hat{\Omega}$ nor can we differentiate without having many difficulties on the boundaries of the patches; in particular, the inverse function theorem is not at our disposal up to now. All these problems disappear, if we use the characteristic map Ψ of the subdivision algorithm, which is now always supposed to be regular and injective, as representation map for the parameter space $\hat{\Omega}$ (cf. Fig. 7).

Lemma 3.7. *If f is a parametrically smooth function over $\hat{\Omega}$, then $\tilde{f} := f \circ \Psi^{-1}$ lies in $C^1(\hat{\Gamma})$, the space of continuously differentiable functions over the compact set $\hat{\Gamma}$.*

The described transformation enables us to use all standard tools from real analysis. For convenience, we shall subsequently use the name f for both functions f and \tilde{f} . It

³ Here and subsequently, injectivity has to be understood modulo the neighborhood relation, i.e. points which have a common image under the representation map are identified.

will come clear from the arguments (typically ξ and η), when we refer to the transformed function. The characteristic map itself, for instance, will be written in the form

$$\Psi(\xi, \eta) = (\xi, \eta) \tag{3.11}$$

as it is transformed to the identity on $\hat{\Gamma}$. Further, we need the following lemma from functional analysis.

Lemma 3.8. *The set of all regular injective functions is open in $C^1(\hat{\Gamma})$.*

Now we can give the

Proof of Theorem 3.6. First, we transform the problem to the canonical form

$$p_1 = o, \quad p_2 = e_1, \quad p_3 = e_2$$

by an affine change of coordinates, where (e_1, e_2, e_3) are the unit vectors and o is the origin of the new coordinate system. This leads to

$$x_m^j(u, v) =: \begin{pmatrix} x(u, v, j, m) \\ y(u, v, j, m) \\ z(u, v, j, m) \end{pmatrix} = \begin{pmatrix} \lambda_2^m b(u, v, j) v_2 + o(\lambda_2^m) \\ \lambda_2^m b(u, v, j) v_3 + o(\lambda_2^m) \\ o(\lambda_2^m) \end{pmatrix} \tag{3.12}$$

or in the transformed form to

$$\begin{pmatrix} x(\xi, \eta, m) \\ y(\xi, \eta, m) \\ z(\xi, \eta, m) \end{pmatrix} = \begin{pmatrix} \lambda_2^m \xi + o(\lambda_2^m) \\ \lambda_2^m \eta + o(\lambda_2^m) \\ o(\lambda_2^m) \end{pmatrix}. \tag{3.13}$$

By Theorem 3.5, $n(p) = e_3$ is the limit of the normal vectors. Therefore, we try to parametrize the surface \bar{P} over the xy -plane in the form $z = h(x, y)$ in a neighborhood of the origin. This is possible, if there is an $M \in \mathbb{N}$ such that the map

$$F_M : (\xi, \eta, m) \mapsto (x(\xi, \eta, m), y(\xi, \eta, m)), \quad (\xi, \eta) \in \hat{\Gamma}, m \geq M \tag{3.14}$$

is injective, because then we can set $z = h(x, y) := z(F_M^{-1}(x, y))$.

Since both the functions $b(u, v, j)$ and their partial derivatives are bounded over the compact set $\hat{\Gamma}$, the order functions converge with respect to the C^1 -norm. Hence

$$\lim_{m \rightarrow \infty} \|\lambda_2^{-m} F_M(\cdot, m) - \Psi\|_{C^1(\hat{\Gamma})} = 0 \tag{3.15}$$

and Ψ is regular and injective by assumption. Thus, by Lemma 3.8, there is an $M \in \mathbb{N}$ with the property that F_M is injective for arbitrary but fixed $m \geq M$, i.e.

$$F_M(\xi, \eta, m) = F_M(\tilde{\xi}, \tilde{\eta}, m) \Rightarrow (\xi, \eta) = (\tilde{\xi}, \tilde{\eta}). \tag{3.16}$$

It remains to show that the images of the maps $F_M(\cdot, m)$ are essentially disjoint, where “essentially” means that the intersection of the images of consecutive maps consists exactly of the common boundary curve. To this end, we define a natural partial order

for closed curves without self-intersections in \mathbb{R}^2 . Denote by $I(c)$ the interior of such a curve c , then we set

$$c_1 < c_2 \Leftrightarrow c_1 \subset I(c_2). \quad (3.17)$$

Now \hat{F} has two closed disjoint boundary curves, we call the outer curve α and the inner curve β (cf. Fig. 7). The boundary of $F_M(\cdot, m)$ consists exclusively of the disjoint curves $F_M(\alpha, m)$ and $F_M(\beta, m)$. This follows from the regularity of $F_M(\cdot, m)$ and the inverse function theorem. Since $F_M(\alpha, m)$ and $F_M(\beta, m)$ are the only boundary curves of the compact and connected set $F(\hat{F}, m)$, we have either

$$F_M(\alpha, m) < F_M(\beta, m) \quad \text{or} \quad F_M(\beta, m) < F_M(\alpha, m)$$

and all function values of $F_M(\cdot, m)$ lie in the strip between the two curves. From the fact that $\beta < \alpha$ and

$$\lim_{m \rightarrow \infty} \|\lambda^{-m} F_M(\alpha, m) - \alpha\|_{C^1(\hat{F})} = \lim_{m \rightarrow \infty} \|\lambda^{-m} F_M(\beta, m) - \beta\|_{C^1(\hat{F})} = 0 \quad (3.18)$$

one easily concludes that the second case must hold if M is large enough. Now consider the sequence

$$F_M(\alpha, M), F_M(\beta, M) = F_M(\alpha, M+1), F_M(\beta, M+1) = F_M(\alpha, M+2), \dots$$

of boundary curves. It is strictly monotone decreasing and converges to the origin. Each strip corresponds exactly to the image of F_M for some fixed m and thus F_M is injective.

The image \mathcal{F}_M of F_M is the interior of the outermost curve $F_M(\alpha, M)$ with exception of the origin, i.e.

$$\mathcal{F}_M = I(F_M(\alpha, M)) \setminus \{(0, 0)\} \quad (3.19)$$

and $\mathcal{F}_M \cup \{(0, 0)\}$ is a neighborhood of the origin. There the surface \bar{P} can be parametrized by

$$z = h(x, y) := \begin{cases} z(F_M^{-1}(x, y)) & \text{for } (x, y) \in \mathcal{F}_M, \\ 0 & \text{for } (x, y) = (0, 0), \end{cases} \quad (3.20)$$

where the function h is continuously differentiable for $(x, y) \neq (0, 0)$ since P is a smooth surface. Finally, we obtain for the normal vector \mathbf{n} of P by Theorem 3.5

$$\lim_{(x,y) \rightarrow (0,0)} \mathbf{n}(x, y) = \lim_{(x,y) \rightarrow (0,0)} \frac{1}{\sqrt{1 + \|\nabla h(x, y)\|^2}} \begin{pmatrix} -h_x(x, y) \\ -h_y(x, y) \\ 1 \end{pmatrix} = \mathbf{e}_3. \quad (3.21)$$

Thus, the limit of the gradient

$$\lim_{(x,y) \rightarrow (0,0)} \nabla h(x, y) = (0, 0)$$

exists and consequently h is continuously differentiable for all $(x, y) \in \mathcal{F}_M \cup \{(0, 0)\}$ as a simple consequence of the mean value theorem. \square

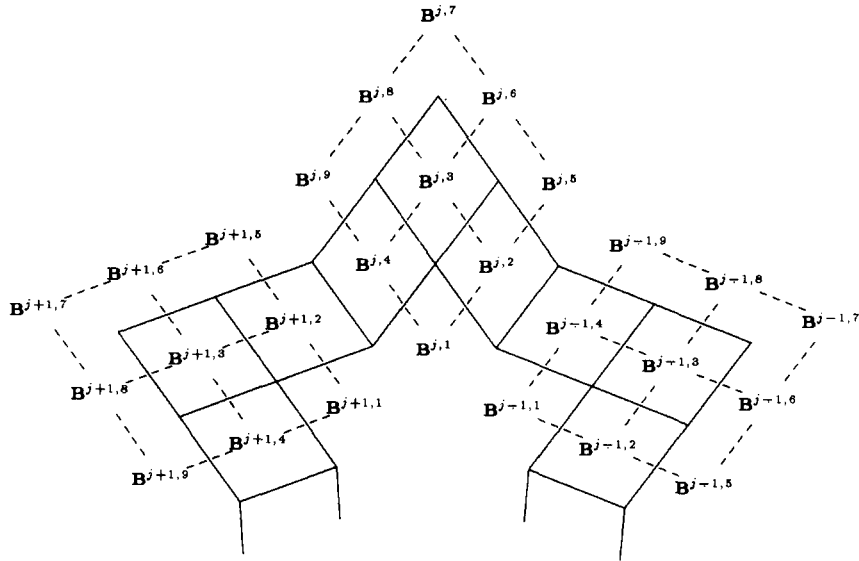


Fig. 8. Labeling of the control points.

4. An example

It is not the purpose of this section to discuss particular subdivision algorithms in detail, but to show some generic phenomena and how they fit into the theory which has been developed in the preceding sections. We skip many computational details and all proofs and restrict ourselves to the results and some hints, how they can be obtained. Providing the arguments should be easy for anyone who is familiar with B-spline theory and the discrete Fourier analysis of cyclic systems. The use of a computer algebra system is helpful, since the computations are rather extensive.

We shall examine some subdivision schemes for biquadratic B-splines near a six-sided hole. The general case of an n -sided hole is not more difficult, but we make this choice in order to keep things as simple and concrete as possible. For the same reason, we shall discuss only some particular subdivision matrices and avoid the use of parameters as much as possible. A more general discussion is beyond the scope of this paper.

The inner layer of patches of a biquadratic B-spline surface is defined by a vector B of control points which consists of 6 blocks B_1, \dots, B_6 with 9 elements each,

$$B := [B^0, \dots, B^5]^T, \quad B^k := [B^{k,1}, \dots, B^{k,9}]^T, \quad k = 0, \dots, 5, \quad (4.1)$$

cf. Fig. 8.

Note: Subsequently, the index k always runs from 0 to 5 and has to be understood modulo 6.

A subdivision process

$$B_0 := B, \quad B_{m+1} := AB_m \quad (4.2)$$

should have no preference for a particular side of the hole, hence the subdivision matrix is supposed to be cyclic in blocks of 9×9 matrices A^0, \dots, A^5 . We have

$$B_{m+1}^k = \sum_{j=0}^5 A^{k-j} B_m^j, \tag{4.3}$$

where

$$A_{1:9,1:4}^0 = \begin{pmatrix} a^0 & 0 & 0 & 0 \\ r & s & 0 & 0 \\ r & s & t & s \\ r & 0 & 0 & s \\ s & r & 0 & 0 \\ s & r & s & t \\ t & s & r & s \\ s & t & s & r \\ s & 0 & 0 & r \end{pmatrix}, \quad A_{1:9,1:4}^1 = \begin{pmatrix} a^1 & 0 & 0 & 0 \\ s & 0 & 0 & t \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ t & 0 & 0 & s \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad A_{1:9,1:4}^5 = \begin{pmatrix} a^5 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ s & t & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ t & s & 0 & 0 \end{pmatrix}, \tag{4.4}$$

$$A_{1,1}^2 = a^2, \quad A_{1,1}^3 = a^3, \quad A_{1,1}^4 = a^4$$

with $r = 9/16$, $s = 3/16$ and $t = 1/16$, all other entries being zero. Except for the first rows, the matrices are uniquely determined by the condition that consecutive layers must join smoothly. The weights r , s and t coincide with those of the subdivision scheme for biquadratic tensor product surfaces. The entries in the first rows are at our disposal and the first simplification is to set most of them to zero. The present form implies that the innermost ring of control points depends only on the innermost ring of the preceding stage. This will be illustrated by figures of the type shown in Fig. 9.

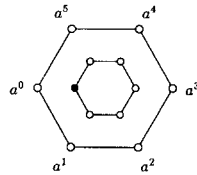


Fig. 9. Free parameters of the subdivision matrix.

For symmetry reasons we require

$$a^1 = a^5, \quad a^2 = a^4 \tag{4.5}$$

and since the rows of A must sum up to 1 we have

$$a^0 + 2a^1 + 2a^2 + a^3 = 1. \tag{4.6}$$

Applying the discrete Fourier transformation

$$\hat{p}_k = \sum_{j=0}^5 \omega^{-jk} p_j, \quad \omega := \exp(2\pi i/6) \tag{4.7}$$

to the system yields the diagonalized form

$$\hat{B}_{m+1}^k = \hat{A}^k \hat{B}_m^k \tag{4.8}$$

with

$$\hat{A}_{1:9,1:4}^k = \begin{pmatrix} \hat{a}^k & 0 & 0 & 0 \\ r + s\omega^{-k} & s & 0 & t\omega^{-k} \\ r & s & t & s \\ r + s\omega^k & t\omega^k & 0 & s \\ s + t\omega^{-k} & r & 0 & s\omega^{-k} \\ s & r & s & t \\ t & s & r & s \\ s & t & s & r \\ s + t\omega^k & s\omega^k & 0 & r \end{pmatrix}. \tag{4.9}$$

The eigenvalues μ_1^k, \dots, μ_9^k of \hat{A}_k are also eigenvalues of A . We find

$$\mu_1^k = \hat{a}^k, \quad \mu_2^k = 1/4, \quad \mu_3^k = \mu_4^k = 1/8, \quad \mu_5^k = \dots = \mu_9^k = 0, \tag{4.10}$$

where

$$\begin{aligned} \hat{a}^0 &= 1, & \hat{a}^1 &= \hat{a}^5 = a^0 + a^1 - a^2 - a^3, \\ \hat{a}^2 &= \hat{a}^4 = a^0 - a^1 - a^2 + a^3, & \hat{a}^3 &= a^0 - 2a^1 + 2a^2 - a^3. \end{aligned} \tag{4.11}$$

What happens if we apply different subdivision matrices to a B-spline surface with a 6-sided hole? The following four cases reveal a completely different behavior.

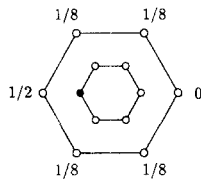


Fig. 10. Weights for Case 1.

Case 1 (see Fig. 10). We have $\hat{a}^0 = 1$, $\hat{a}^1 = \hat{a}^3 = \hat{a}^5 = 1/2$ and $\hat{a}^2 = \hat{a}^4 = 1/4$ and obtain for the leading eigenvalues of A

$$\lambda_1 = 1, \quad \lambda_2 = \lambda_3 = \lambda_4 = 1/2, \quad \lambda_5 = 1/4. \tag{4.12}$$

According to Theorem 3.2 we can expect a convergent algorithm, but no smooth surface. This is illustrated in Fig. 11.⁴ The hole is filled gradually, but the different shades near the limit point indicate that there is no limit for the normal vectors.

⁴ All surfaces shown in Figs. 11, 14, 17 and 20 were generated from the same set of initial data. The pictures on the left-hand side show the first few steps for the given algorithm and those on the right-hand side a magnification of the region near the limit point. Different shades correspond to different normal vectors, the white lines indicate boundaries of consecutive layers.



Fig. 11. Surface for Case 1 and 1000-fold magnification.

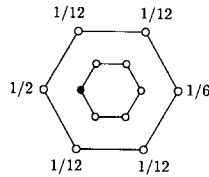


Fig. 12. Weights for Case 2.

Case 2 (see Fig. 12). We have $\hat{a}^0 = 1$, $\hat{a}^1 = \hat{a}^3 = \hat{a}^5 = 1/3$ and $\hat{a}^2 = \hat{a}^4 = 1/2$ and obtain for the leading eigenvalues of A

$$\lambda_1 = 1, \quad \lambda_2 = \lambda_3 = 1/2, \quad \lambda_4 = 1/3. \tag{4.13}$$

The eigenvectors $v_2 = [v_2^0, \dots, v_2^5]^T$ and $v_3 = [v_3^0, \dots, v_3^5]^T$ are given by

$$v_2^k = \text{Re}(\omega^{2k}\hat{v}), \quad v_3^k = \text{Im}(\omega^{2k}\hat{v}), \tag{4.14}$$

where

$$\hat{v} = [14, 21, 36, 21, 28, 45, 58, 45, 28] + 7\sqrt{3}i[0, -1, 0, 1, -2, -1, 0, 1, 2] \tag{4.15}$$

is an eigenvector of \hat{A}^2 and \hat{A}^4 . We find

$$v_2^k = v_2^{k+3}, \quad v_3^k = v_3^{k+3}, \tag{4.16}$$

from which follows that the characteristic map is not injective. As mentioned above, the rows of the matrix $V = [v_2, v_3]$ serve as two-dimensional control points for the characteristic map. Fig. 13 shows on the left-hand side a plot of these control points connected by dashed lines according to the structure defined in Fig. 8. The right-hand side shows the corresponding characteristic map Ψ where the images of a 7×7 grid parallel to the edges of the unit square ω are plotted with thin lines. The thicker lines indicate the patch boundaries. The image consists of 18 patches, but in fact one can see only 9 patches since always two copies of ω have a common image.

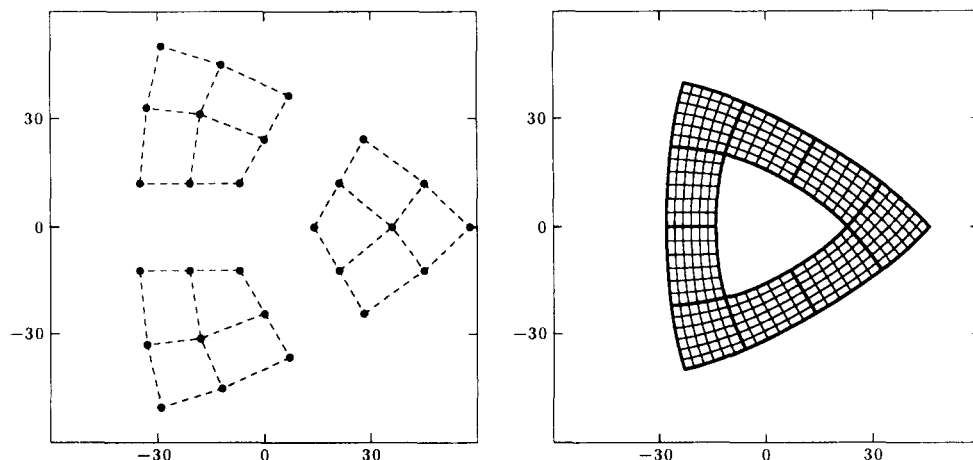


Fig. 13. Eigenvectors and characteristic map for Case 2.

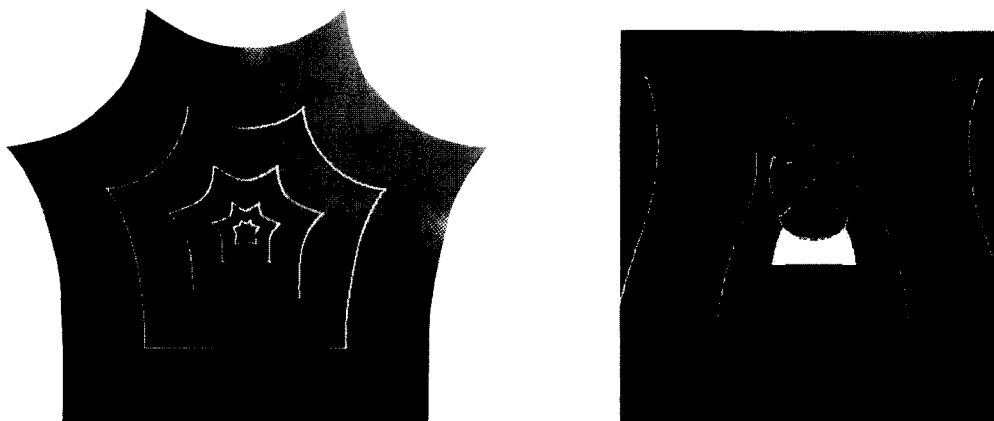


Fig. 14. Surface for Case 2 and 1000-fold magnification.

A numerical or analytical computation of the Jacobian of the characteristic map Ψ shows that Ψ is regular and thus the surfaces generated by the given algorithm are tangent plane continuous by Theorem 3.5. Nevertheless, self-intersections will lead to a non-regular limit point. This is shown in Fig. 14.

The non-injectivity of the characteristic map in Case 2 is rather obvious, since the dominant eigenvalues occur in the blocks $k = 2$ and $k = 4$ of the Fourier transformed system. But even if the dominant eigenvalues come from the blocks $k = 1$ and $k = 5$, things can go wrong. This is shown by the next case as well as the fact that the regularity of the characteristic map is an indispensable condition for generating smooth surfaces.

Case 3 (see Fig. 15). We have $\hat{a}^0 = 1$, $\hat{a}^1 = \hat{a}^5 = -3/4$, $\hat{a}^2 = \hat{a}^4 = 1/4$ and $\hat{a}^3 = 0$ and obtain for the leading eigenvalues of A

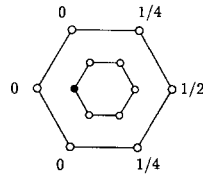


Fig. 15. Weights for Case 3.

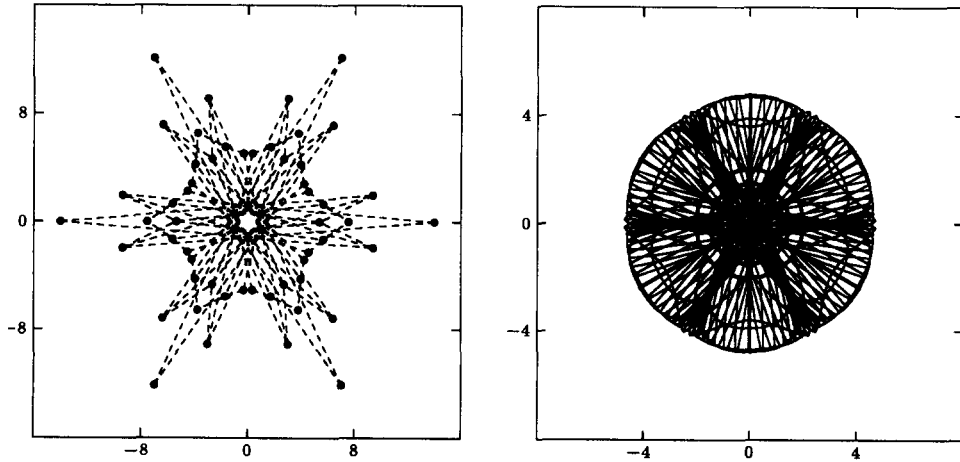


Fig. 16. Eigenvectors and characteristic map for Case 3.

$$\lambda_1 = 1, \quad \lambda_2 = \lambda_3 = -3/4, \quad \lambda_4 = 1/4. \tag{4.17}$$

In this case one can show that the characteristic map is not regular and indeed it looks rather strange, see Fig. 16.

The generated surfaces are *not* tangent plane continuous and far from being well shaped, as can be seen in Fig. 17.

One could object that the dominant eigenvalues are negative and that it is rather obvious that this algorithm cannot work. However, following the arguments of Ball and Storry it would have been found to be tangent plane continuous.

Finally, here is a suitable algorithm for generating regular surfaces:

Case 4 (see Fig. 18). We have $\hat{a}^0 = 1, \hat{a}^1 = \hat{a}^5 = 1/2$ and $\hat{a}^2 = \hat{a}^3 = \hat{a}^4 = 0$ and obtain for the leading eigenvalues of A

$$\lambda_1 = 1, \quad \lambda_2 = \lambda_3 = 1/2, \quad \lambda_4 = 1/4. \tag{4.18}$$

The eigenvectors $v_2 = [v_2^0, \dots, v_2^5]^T$ and $v_3 = [v_3^0, \dots, v_3^5]^T$ are given by

$$v_2^k = \text{Re}(\omega^k \hat{v}), \quad v_3^k = \text{Im}(\omega^k \hat{v}), \tag{4.19}$$

where

$$\hat{v} = [14, 35, 48, 35, 56, 67, 82, 67, 56] + 7\sqrt{3}i[0, -1, 0, 1, -2, -1, 0, 1, 2] \tag{4.20}$$

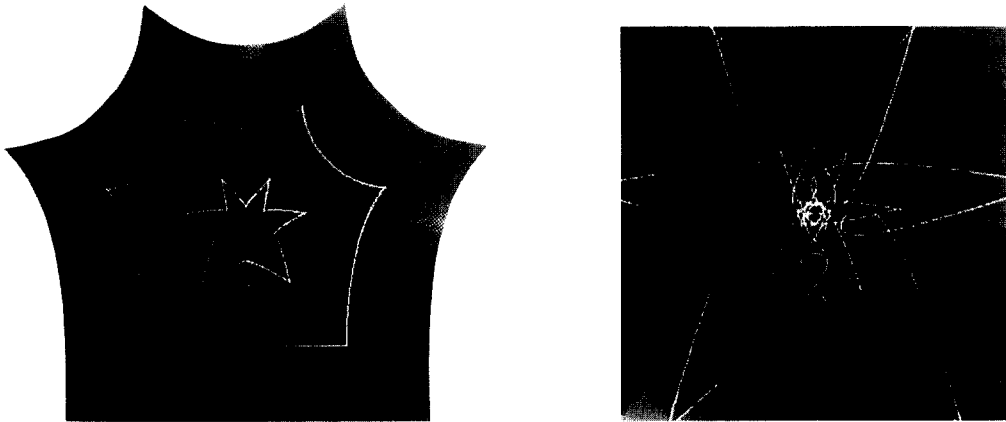


Fig. 17. Surface for Case 3 and 300-fold magnification.

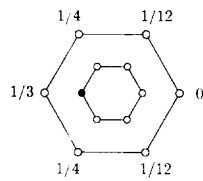


Fig. 18. Weights for Case 4.

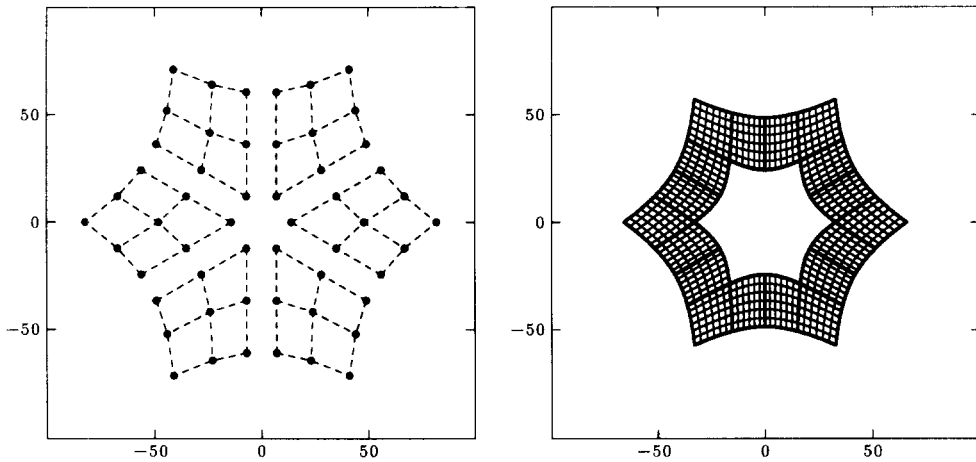


Fig. 19. Eigenvectors and characteristic map for Case 4.

is an eigenvector of \hat{A}^1 and \hat{A}^5 . It can easily be shown that the restriction of the characteristic map Ψ to a single patch is injective. To prove the global injectivity, write Ψ in Bézier form and use the convex hull property. The Bézier form is also convenient to show that Ψ is regular.

Thus, all assumptions of Theorem 3.6 are fulfilled and the algorithm generates smooth regular surfaces. Indeed, Fig. 20 illustrates that the algorithm lives up to the usual design

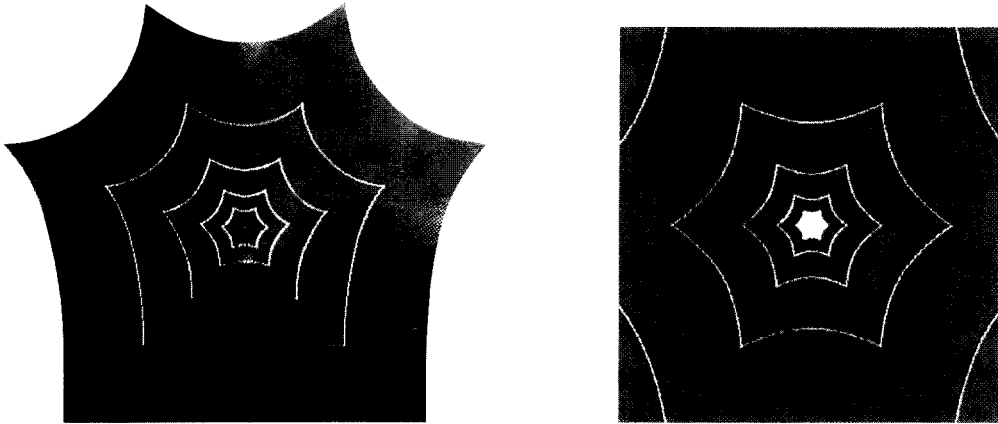


Fig. 20. Surface for Case 4 and 1000-fold magnification.

standards in CAGD.

Finally, we want to mention that exactly the same characteristic map as in the last case is found for the Doo–Sabin weights given in Fig. 21.

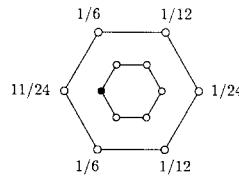


Fig. 21. Doo–Sabin weights.

Here we have $\hat{a}^0 = 1, \hat{a}^1 = \hat{a}^5 = 1/2$ and $\hat{a}^2 = \hat{a}^3 = \hat{a}^4 = 1/4$ and it can easily be seen that the eigenvectors v_2 and v_3 depend only on the value of $\hat{a}^1 = \hat{a}^5$.

Appendix

Lemma 3.7. *If f is a parametrically smooth function over $\hat{\Omega}$, then $\tilde{f} := f \circ \Psi^{-1}$ lies in $C^1(\hat{F})$, the space of continuously differentiable functions over the compact set \hat{F} .*

Proof. Parametrical smoothness means that the cross boundary derivatives of f at a common boundary of two patches are equal up to sign. More precisely, assume without loss of generality that for some patches j and k the boundary curves $([0, 1], 0, j)$ and $([0, 1], 0, k)$ are identified by the neighborhood relation. Then we have for $u \in [0, 1]$

$$f(u, 0, j) = f(u, 0, k), \quad Df(u, 0, j) = Df(u, 0, k)Q, \quad (\text{A.1})$$

where Df denotes the matrix of partial derivatives with respect to (u, v) and Q is a diagonal 2×2 -matrix with entries 1 and -1 . In particular, we obtain for the characteristic map

$$c(u) := \Psi(u, 0, j) = \Psi(u, 0, k), \quad D\Psi(u, 0, j) = D\Psi(u, 0, k)Q. \quad (\text{A.2})$$

Now it must be shown that $D\tilde{f}$ is a continuous function over \hat{F} . Ψ is smoothly invertible in the interior of the patches and the chain rule implies there

$$D\tilde{f}(x, y) = Df(\Psi^{-1}(x, y)) (D\Psi(\Psi^{-1}(x, y)))^{-1}. \quad (\text{A.3})$$

If (x, y) is approaching the boundary curve $c(u)$ from either the patch j or k , we obtain in the limit

$$D\tilde{f}(c(u)) = Df(u, 0, j)(D\Psi(u, 0, j))^{-1} \quad (\text{A.4})$$

or

$$D\tilde{f}(c(u)) = Df(u, 0, k)(D\Psi(u, 0, k))^{-1}. \quad (\text{A.5})$$

The right-hand sides coincide since

$$\begin{aligned} Df(u, 0, j)(D\Psi(u, 0, j))^{-1} &= (Df(u, 0, k)Q)(Q^{-1}(D\Psi(u, 0, k))^{-1}) \\ &= Df(u, 0, k)(D\Psi(u, 0, k))^{-1} \end{aligned} \quad (\text{A.6})$$

and thus, $D\tilde{f}$ is well defined on \hat{F} and continuous by the mean value theorem. \square

Lemma 3.8. *The set of regular injective functions is open in $C^1(\hat{F})$.*

Proof. The proof is based on the continuous embedding of $C^1(\hat{F})$ in $C^{0,1}(\hat{F})$, the space of Lipschitz continuous functions, i.e. there is a constant $M \in \mathbb{R}$ such that

$$\sup_{x, \tilde{x} \in \hat{F}} \frac{\|g(x) - g(\tilde{x})\|_2}{\|x - \tilde{x}\|_2} =: \|g\|_{C^{0,1}(\hat{F})} \leq M \|g\|_{C^1(\hat{F})} \quad (\text{A.7})$$

for all functions $g \in C^1(\hat{F})$. If a function $f \in C^1(\hat{F})$ is regular and injective, then it can be inverted and $f^{-1} \in C^1(f(\hat{F})) \subset C^{0,1}(f(\hat{F}))$. Thus,

$$\|f^{-1}\|_{C^{0,1}(f(\hat{F}))} := \sup_{y, \tilde{y} \in f(\hat{F})} \frac{\|f^{-1}(y) - f^{-1}(\tilde{y})\|_2}{\|y - \tilde{y}\|_2} =: K < \infty \quad (\text{A.8})$$

and consequently

$$\inf_{x, \tilde{x} \in \hat{F}} \frac{\|f(x) - f(\tilde{x})\|_2}{\|x - \tilde{x}\|_2} = 1/K. \quad (\text{A.9})$$

Now we find for some function $\tilde{f} := f + g$

$$\begin{aligned} \inf_{x, \tilde{x} \in \hat{F}} \frac{\|\tilde{f}(x) - \tilde{f}(\tilde{x})\|_2}{\|x - \tilde{x}\|_2} &\geq \inf_{x, \tilde{x} \in \hat{F}} \frac{\|f(x) - f(\tilde{x})\|_2}{\|x - \tilde{x}\|_2} - \sup_{x, \tilde{x} \in \hat{F}} \frac{\|g(x) - g(\tilde{x})\|_2}{\|x - \tilde{x}\|_2} \\ &\geq 1/K - \|g\|_{C^{0,1}(\hat{F})} \geq 1/K - M \|g\|_{C^1(\hat{F})}. \end{aligned} \quad (\text{A.10})$$

Thus \tilde{f} is injective, if $\|g\|_{C^1(\hat{r})} = \|f - \tilde{f}\|_{C^1(\hat{r})} < 1/(MK)$. One can show by a simple continuity argument that \tilde{f} is also regular, if $\|f - \tilde{f}\|_{C^1(\hat{r})}$ is small enough and so the proof is complete. \square

References

- Ball, A.A. and Storry, D.J.T. (1986), A matrix approach to the analysis of recursively generated B-spline surfaces, *Computer-Aided Design* 18, 437–442.
- Ball, A.A. and Storry, D.J.T. (1988), Conditions for tangent plane continuity over recursively generated B-spline surfaces, *ACM Trans. Graph.* 7(2), 83–102.
- Catmull, E. and Clark, J. (1978), Recursively generated B-spline surfaces on arbitrary topological meshes, *Computer-Aided Design* 10, 350–355.
- Doo, D. and Sabin, M.A. (1978), Behaviour of recursive division surfaces near extraordinary points, *Computer-Aided Design* 10, 356–360.
- Loop, Ch.T. (1987), Smooth Subdivision for Surfaces Based on Triangles, Master Thesis, Univ. of Utah, Salt Lake City.
- Nasri, A.H. (1987), Polyhedral subdivision methods for free-form surfaces, *ACM Trans. Graph.* 6(1) 29–73.
- Reif, U. (1993), Neue Aspekte in der Theorie der Freiformflächen beliebiger Topologie, Thesis, Universität Stuttgart
- Halstead, M., Kass, M. and DeRose, T.D. (1993), Efficient, Fair interpolation using Catmull–Clark surfaces, in: *SIGGRAPH '93*, ACM Press, New York, 35–44.