Behaviour of recursive division surfaces near extraordinary points

D Doo and M Sabin*

BEST COPY AVAILABLE

NOTICE:
THIS MATERIAL
MAY BE PROTECTED
BY COPYRIGHT LAW
(TITLE 17 U.S. CODE)

The behaviour of the limit surface defined by a recursive division construction can be analysed in terms of the eigenvalues of a set of matrices. This analysis predicts effects actually observed, and leads to suggestions for the further improvement of the method.

A recursive division surface definition takes a polyhedron-like configuration of points, edges and faces (the faces need not be plane), and generates a surface as the limit of a 'chopping off the corners' process. There are many detailed variants of this process. All of them construct a new polyhedron at each step with more vertices and smaller faces than the original, the surface being the limit after many steps of division. They differ in the rules by which the new vertices are constructed.

After several steps, the bulk of the polyhedron consists of faces arranged in a regular lattice. There are, however, a finite number of extraordinary points at which the regularity of the lattice is disturbed. The properties of the regular parts of the surface are derived in terms of Cartesian product B-spline surfaces. This paper is concerned with the behaviour near the extraordinary points.

Because the number of extraordinary points does not increase with successive steps of the algorithm, the distance between them remains more or less constant. As the sizes of the faces shrink at every step, the number of faces between the extraordinary points has to grow, so they can be treated as being isolated from each other by regions of regular lattice.

CATMULL–CLARK CUBIC FORMULATION

In this formulation[1,2], which is used as an example, the extraordinary points are vertices at which n edges join, when n is not 4. Such points are separated by regions of regular rectangular lattice. The region of an extraordinary point may be represented diagrammatically as in Figure 1.

Applying one step of their algorithm gives Figure 2. Note that the configuration round the extraordinary point is topologically similar to Figure 1. The points are deliberately labelled to stress this similarity. In the original formulation (which for illustration has simpler

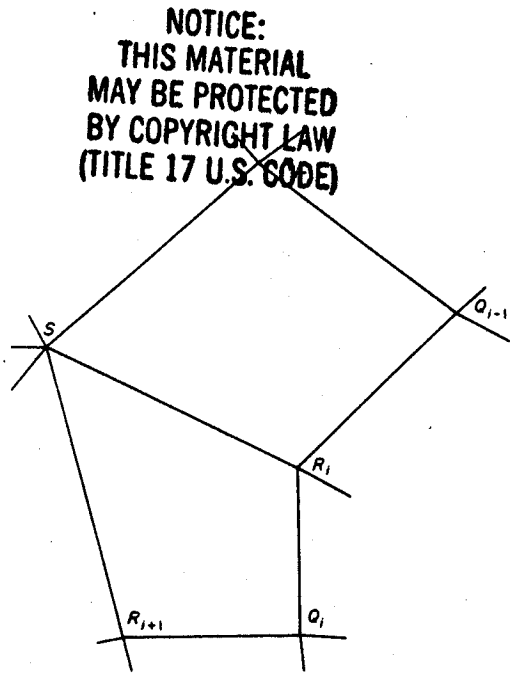

Figure 1. Region of an extraordinary point

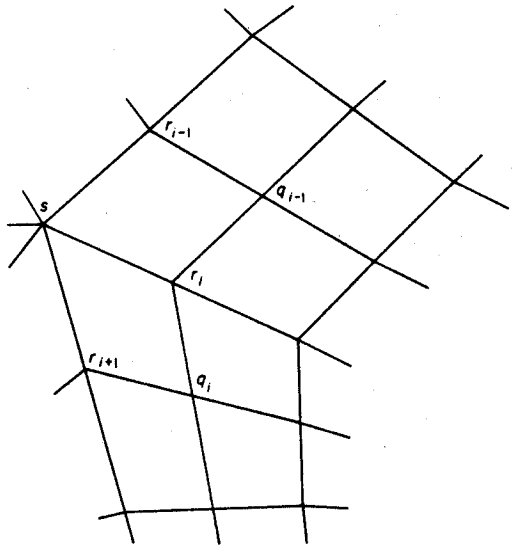

Figure 2. Applying one step of the Catmull–Clark algorithm to an extraordinary point

Brunel University, Uxbridge, UK.
*Kongsberg Limited, Maidenhead, Berks., UK.

 0010–3385/78/0603056–05 $02.00 © 1978 IPC Business Press

numbers than the most recent variant), the new points are computed from the old by the equations

$$q_i = (Q_i + R_i + R_{i+1} + S)/4 \quad i=1,n$$
$$r_i = (q_i + q_{i-1} + R_i + S)/4 \quad i=1,n \qquad (1)$$
$$s = (\sum_i \frac{q_i}{n} + \sum_i \frac{R_i}{n} + 2S)/4$$

The notation used here is slightly modified from the original, but conveys the same information.

MATRIX FORM

These equations are linear in the points, and so may be expressed in matrix form

$$\begin{bmatrix} q_1 \\ \bullet \\ \bullet \\ q_n \\ r_1 \\ \bullet \\ \bullet \\ r_n \\ s \end{bmatrix} = \begin{bmatrix} & & \\ & M & \\ & & \end{bmatrix} \begin{bmatrix} Q_1 \\ \bullet \\ \bullet \\ Q_n \\ R_1 \\ \bullet \\ \bullet \\ R_n \\ S \end{bmatrix}$$

At each step, the new surroundings of the extraordinary point are obtained by multipying the old by matrix M of constant elements.

It is a standard result in linear algebra that repeated multiplication of a vector by a constant matrix converges toward the eigenvector corresponding to the largest eigenvalue of the matrix. The behaviour of the limit surface is analysed, therefore, in terms of the eigenproperties of M.

FOURIER TRANSFORM

Although it is possible to evaluate eigenvalues and eigenvectors numerically, it provides considerably more insight to use the cyclic symmetry of Figure 1 to replace M by a number of smaller matrices for which algebraic manipulation is possible. The method used here applies a discrete Fourier transform to Q, R and S, and then separates the terms of different frequency.

Let

$$Q_j = \sum_{\omega=0}^{n/2} Q_\omega e^{2\pi ij\omega/n} \quad j = \sqrt{-1}$$

define the Fourier coefficients Q_ω, ω=0. $n/2$. Similar equations for R_j, q_j, r_j define corresponding sets of coefficients R_ω q_ω and r_ω.

Because Q_ω are complex and Q_j are real (Q_ω is real if $\omega = 0$ or $2\omega = n$), there are always exactly n degrees of freedom to satisfy the n equations.

S and s can also be forced into this pattern by setting

$$S_0 = S$$
$$S_\omega = 0, \ \omega \neq 0$$

Note that

$$Q_{j+1} = \sum_{\omega=0}^{n/2} Q_\omega e^{2\pi(j+1)j\omega/n}$$
$$= \sum_{\omega=0}^{n/2} Q_\omega e^{2\pi j\omega/n} e^{2\pi ij\omega/n}$$
$$= \sum_{\omega=0}^{n/2} Q_\omega e^{2\pi j\omega/n} e^{2\pi ij\omega/n}$$

and, writing a_ω for $e^{2\pi j\omega/n}$ and a_ω^* for $e^{-2\pi j\omega/n}$

$$Q_{j+1} = \sum_{\omega=0}^{n/2} a_\omega Q_\omega e^{2\pi ij\omega/n}$$

With this substitution made, equations (1) can be rewritten as a set of equations for each value of ω, and in each set we can cancel out the $e^{2\pi ij\omega/n}$. This procedure is closely analogous to the treatment of time series data in terms of components of different frequencies. In this case, the frequency ω may be thought of as the number of complete oscillations of, say, z per complete cycle through the Q_j.

This gives

$$q_\omega = (Q_\omega + (1+a_\omega)R_\omega + S_\omega)/4$$
$$r_\omega = ((1+a_\omega^*)q_\omega + R_\omega + S_\omega)/4 \qquad \omega = 0, n/2$$
$$s_0 = (q_0 + R_0 + 2S_0)/4$$

which, by substituting the equations for q_ω into those for r_ω and s_0, gives

$$\begin{bmatrix} q_0 \\ r_0 \\ s_0 \end{bmatrix} = \begin{bmatrix} 4/16 & 8/16 & 4/16 \\ 2/16 & 8/16 & 6/16 \\ 1/16 & 6/16 & 9/16 \end{bmatrix} \begin{bmatrix} Q_0 \\ R_0 \\ S_0 \end{bmatrix} \quad \omega=0$$

$$\begin{bmatrix} q_\omega \\ r_\omega \end{bmatrix} = \begin{bmatrix} 4/16 & 4(1+a_\omega)/16 \\ (1+a_\omega^*)/16 & ((1+a_\omega^*)(1+a_\omega)+4)/16 \end{bmatrix} \begin{bmatrix} Q_\omega \\ R_\omega \end{bmatrix} \quad \omega \neq 0$$

The eigenvalues of these equations are readily determined to be 1, 1/4, 1/16 at $\omega = 0$ and, writing

$(1 + a_\omega^*)(1 + a_\omega) = 2\phi_\omega$ $(=2(1 + \cos 2\pi\omega/n))$

$$\frac{4 + \phi_\omega \pm \sqrt{[\phi_\omega (8 + \phi_\omega)]}}{16} \quad \text{at } \omega \neq 0$$

INTERPRETATION: $n = 4$

Consider first the situation where $n = 4$. This is an ordinary point of the surface, and so it is known, for example, that it has continuity of first and second derivatives.

The relevant values of ω are 0, 1 and 2 only

- at $\omega = 1$ $\phi_\omega = 1 + \cos\frac{2\pi}{4} = 1$
- at $\omega = 2$ $\phi_\omega = 1 + \cos\frac{4\pi}{4} = 0$

Thus the eigenvalues of interest are

- $\omega = 0$ 1, 1/4, 1/16
- $\omega = 1$ 1/2, 1/8
- $\omega = 2$ 1/4, 1/4

The largest eigenvalue is the value 1.0 at $\omega = 0$. This indicates that as the iterations proceed, the configuration as a whole stays in the same place. This unit eigenvalue must have this value because all rows of the $\omega = 0$ matrix sum to unity, which will always happen when the new points at each step are calculated as weighted means of the old points. The effect of this eigenvalue can be cancelled by taking as coordinate system origin the position to which the extraordinary point converges, and so this eigenvalue will be ignored in the rest of this analysis.

The next eigenvalue is the value 1/2 at $\omega = 1$. This indicates that, as iteration proceeds, both the Q_i and the R_i converge to affine-regular n-gons (affine projections of regular n-sided polygons[3]) which halve in size without rotation at each step. This behaviour has been observed.

Convergence occurs because the relative contribution from other terms in the series shrinks, their eigenvalues being smaller.

Affine-regular n- gons appear because $\omega = 1$ for the dominant eigenvalue. If the unit frequency were the only term present, which holds in the limit, the components of Q_i and R_i have the form

$$x = a_x \sin(\frac{2\pi i}{n} + b_x)$$

$$y = a_y \sin(\frac{2\pi i}{n} + b_y)$$

$$z = a_z \sin(\frac{2\pi i}{n} + b_z)$$

which generate the vertices of such an n-gon when i is given the integer values from 1 to n.

Halving takes place because the value of the eigenvalue is 1/2. Rotation is absent because the eigenvalue is real, so that the phase shifts (values of b) remain constant from iteration to iteration, while the values of a halve.

If that X–Y plane is chosen towards which the plane of the n-gons converges, the behaviour of Z, on which the continuity of the limit surface then depends, is dominated by the third eigenvalue, in this case the values 1/4 at $\omega = 0$ and $\omega = 2$.

The $\omega = 0$ eigenvalue controls the behaviour of cup-shaped configurations; the $\omega = 2$ eigenvalue for saddle-shaped data. In both cases the out-of-plane components are divided by four every step.

CONTINUITY

Writing λ_ω for the dominant eigenvalue at frequency ω gives

$\lambda_0 = 1/4$ (excluding the unit eigenvalue)

$\lambda_1 = 1/2$

$\lambda_2 = 1/4$

The slope continuity may be investigated by setting up a one-sided estimate of the first derivative, and examining its limiting behaviour.

Such an estimate is kz/r ($r^2 = x^2 + y^2$) for any point, such as Q_i.

Because of the choice of axes, r will be multiplied by λ_1 at each step and z by λ_ω, so that the limit of kz/r depends on $\lambda_\omega / \lambda_1$.

If this is less than unity, the estimate will shrink at every step, converging to zero, and since this limit is independent of the direction taken in the X–Y plane, the surface is slope-continuous.

Having established this, z/r^2 can be used as an estimate of the second derivative. This estimate is multiplied by $\lambda_\omega / \lambda_1^2$ at every step. For $n = 4$ this ratio is 1.0 and so the curvature takes some finite nonzero value. (The continuity of curvature follows from the symmetry of the $\omega = 0$ and $\omega = 2$ configurations.)

INTERPRETATION: $n \neq 4$

The values of λ_ω, together with λ_ω/λ_1 and $\lambda_\omega/\lambda_1^2$ are tabulated for various values of n in Table 1.

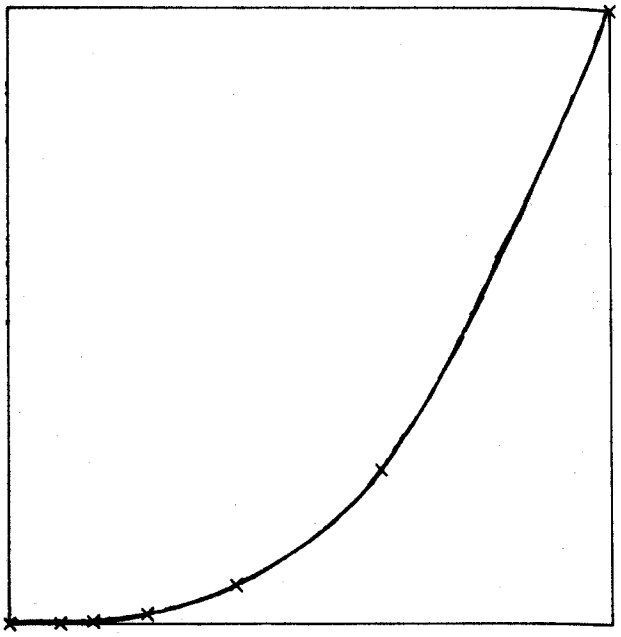

Figure 3. $\lambda_\omega/\lambda_1^2 < 1$

Table 1. Eigenvalues of original form

	$n = 3$	$n = 5$	$n = 6$	$n = 8$	$n \to \infty$
λ_0	0.250	0.250	0.250	0.250	0.250
λ_1	0.410	0.550	0.580	0.611	0.655
λ_2		0.340	0.410	0.500	0.655
λ_3			0.250	0.366	0.655
λ_4				0.250	0.655
λ_0 / λ_1	0.609	0.455	0.431	0.409	0.381
λ_0 / λ_1^2	1.487	0.826	0.744	0.670	0.584
λ_2 / λ_1		0.618	0.707	0.818	1.000
λ_2 / λ_1^2		1.124	1.220	1.339	1.528

It can be seen that λ_ω/λ_1 is less than unity for all n, and so slope continuity is achieved. $\lambda_\omega/\lambda_1^2$, however, is considerably greater than unity for $n = 3$, and so the estimate of curvature increases without limit as division proceeds, giving behaviour of the form

$$z = r^p \text{ with } 1 < p < 2$$

This behaviour may be described loosely as a discontinuity of a fractional derivative.

For $n > 4$, λ_0/λ_1^2 is less than unity, so that the curvature tends to zero, giving a local flat spot on the surface.

These types of behaviour are illustrated by Figures 3–5 which show the effects of $\lambda_\omega/\lambda_1^2 < 1$, $=1$ and >1. The points plotted use $\lambda_0 = 0.250$, $\lambda_1 = 0.611$ and $\lambda_2 = 0.500$, which are the values for the original Catmull–Clark cubic when $n = 8$. Figure 5 is directly comparable with Figure 10 of the Catmull–Clark paper[2]. Figure 4 illustrates the desirable behaviour with $\lambda_\omega = \lambda_1^2 = 0.373$.

IMPROVEMENTS

The behaviour described above has been observed by Catmull and Clark[2], and they have produced a modified version, in which the equation for s is weighted by the connectivity of S.

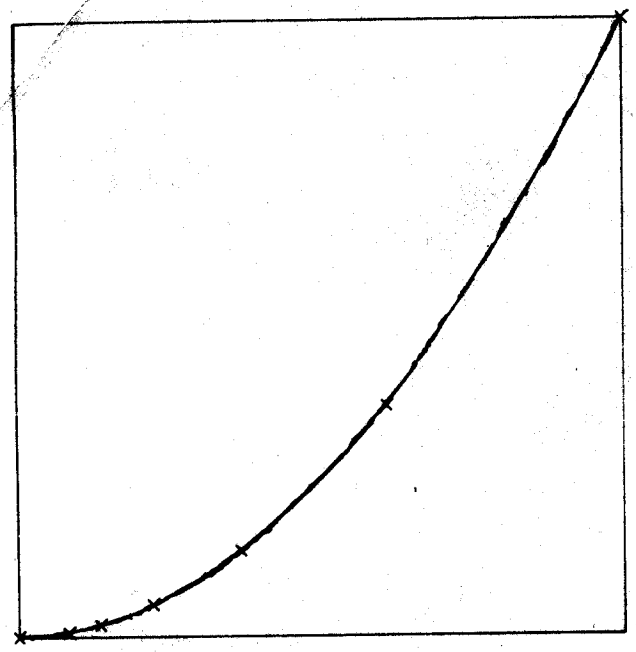

Figure 4. $\lambda_\omega/\lambda_1^2 = 1$

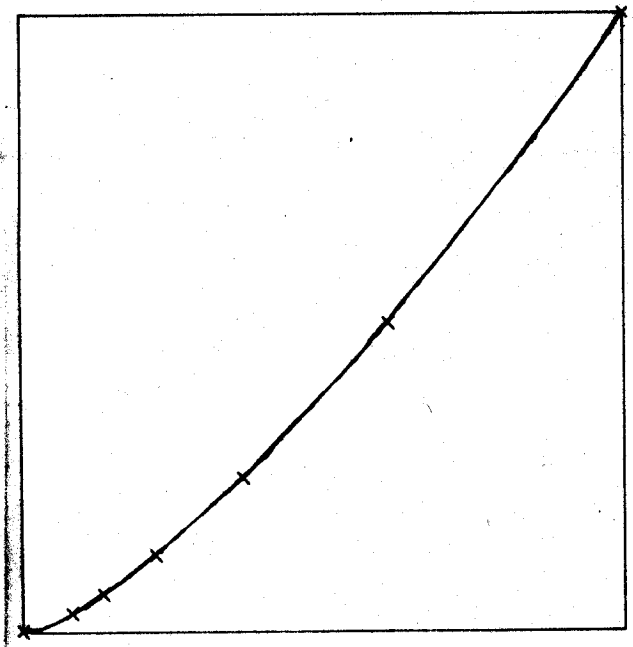

Figure 5. $\lambda_\omega/\lambda_1^2 > 1$

$$s = (\sum_i \frac{q_i}{n} + \sum_i \frac{R_i}{n} + (n-2)S)/n$$

This changes the $\omega = 0$ matrix and its eigenvalues, but does not alter the $\omega > 0$ properties (Table 2).

The value of λ_0/λ_1^2 for $n = 3$ is now much closer to unity, thus explaining the improvement seen on the tetrahedral data.

λ_0/λ_1^2 is also brought closer to unity for $n > 4$, but the value is now greater than unity instead of less. The local flat spots are replaced by local infinities of curvature.

It is clear that further improvement could be made by using as the equation for s

$$s = (\sum \frac{q_i}{n} + \sum \frac{R_i}{n} + (W-2)S)/W$$

where W is a precomputed function of n, chosen to give

$$\lambda_0/\lambda_1^2 = 1$$

The values of W could be stored as a table indexed by n for reasonable values of n, and as an asymptotic expression for very large values; so very little extra computing cost need be incurred.

Even this change would not affect the behaviour with data containing components of high ω value. A scheme of higher performance could use weighted means in all three equations, with precomputed weights chosen to give either

$$\lambda_0/\lambda_1^2 = 1 = \lambda_2/\lambda_1^2 = \lambda_3/\lambda_1^3$$

or $\lambda_0 = 1/4 \quad \lambda_1 = 1/2 \quad \lambda_2 = 1/4$

for all values of n

The nonlinear simultaneous equations need only be solved once for each n, because the values could again be stored in tables.

RELEVANCE

It may be asked whether such 'fine tuning' is important, when the only effect visible is a slight highlighting. The answer must be that for application in manufacturing, there are two reasons for avoiding fractional power derivative discontinuities.

The first is that concavities of zero radius cannot be machined, and so it is better if the numerical model does not include them.

The second is that many analysis procedures will function by using the closed-form equations rather than using the

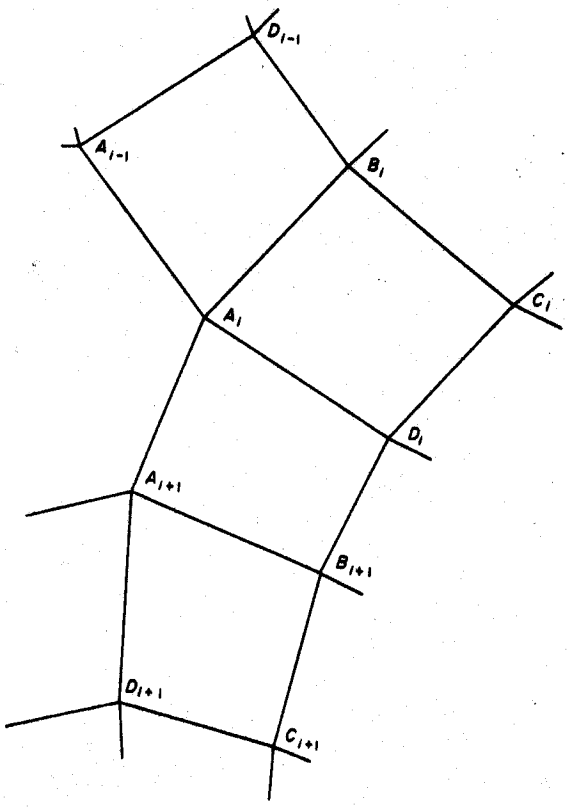

Figure 6. Quadratic form

Table 2. Eigenvalues of revised form

	$n=3$	$n=5$	$n=6$	$n=8$	$n\to\infty$
λ_0	0.167	0.322	0.375	0.443	0.655
λ_1	0.410	0.550	0.580	0.611	0.655
λ_0 / λ_1	0.406	0.586	0.647	0.725	1.000
λ_0 / λ_1^2	0.991	1.066	1.116	1.186	1.528

Table 3. Eigenvalues for quadratic formulations

	$n=3$	$n=4$	$n=5$	$n=6$	$n=8$	$n\to\infty$
λ_0	0.250	0.250	0.250	0.250	0.250	0.250
λ_1	0.375	0.500	0.577	0.625	0.677	0.750
λ_2		0.250	0.298	0.375	0.500	0.750
λ_0 / λ_1	0.667	0.500	0.433	0.400	0.369	0.333
λ_0/λ_1^2	1.778	1.000	0.750	0.640	0.546	0.444
λ_2/λ_1		0.500	0.516	0.600	0.739	1.000
λ_2 / λ_1^2		1.000	0.894	0.960	1.092	1.333

subdivision process itself. In this case, the recursive division proceeds locally only until an approximation can be generated for the infinite sequence of patches round each extraordinary point. It is likely that fractional power behaviour will be much less easy to approximate, thus delaying the level at which approximation is possible.

QUADRATIC FORMULATIONS

Two quadratic forms have been analysed, in which the extraordinary points are at the centres of n-sided faces. During the division step, each face is replaced by a new face, which is connected across the old edges and across the old vertices by other new faces.

Applying this process to Figure 6 gives Figure 7.

In such methods the vertices of each new face are functions only of the vertices of the corresponding old face.

$$a_i = \sum_{j=1}^{n} W_{ij} A_j$$

By symmetry, W_{ij} is a function of $|i-j|$ only,

and so a vector of $n/2+1$ values of W for each value of n defines the variant completely.

The original Catmull quadratic[1] uses

$W_{ij} = (4n+2)/8n \qquad |i-j| = 0$

$W_{ij} = (n+2)/8n \qquad |i-j| = 1$

$W_{ij} = 2/8n \qquad |i-j| > 1$

which results in the eigenvalues in Table 3.

This is considerably better than the original cubic.

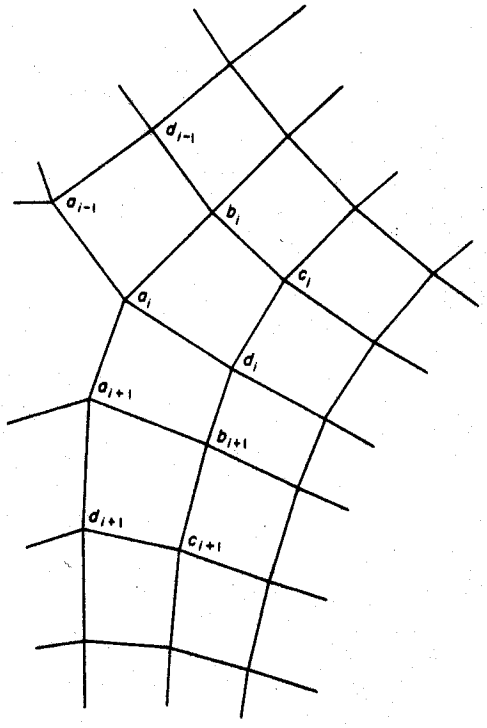

Figure 7. Quadratic form after division

The Doo–Sabin quadratic[4] uses

$W_{ij} = (n+5)/4n \qquad i=j$

$W_{ij} = (3 + 2\cos(2\pi(i-j)/n))/4n \qquad i \neq j$

which gives dominant eigenvalues

$\lambda_0 = 1/4, \lambda_1 = 1/2, \lambda_\omega = 1/4\ \omega > 1$

for all n, thus giving discontinuity of exactly the second derivative under all circumstances.

This ideal behaviour may not be significantly better than the Catmull quadratic in normal use, but that it can be achieved in the quadratic case encourages the search for an equally optimal cubic formulation.

REFERENCES

1 **Catmull, E and Clark, J** private communications (1978)

2 **Catmull, E and Clark, J** 'Recursively generated B-spline surfaces on arbitrary topological meshes' *Comput. Aided Des.* Vol 10 No 6 (November 1978) pp 000–000

3 **Bachmann, F and Schmidt, E** *n-gons* University of Toronto Press (1975)

4 **Doo, D** PhD thesis, Brunel University (to appear)