Analysis of a subdivision scheme

The pure analysis scenario is that we are presented with a complete scheme; with a specific grid, arity, pattern of non-zero coefficients and values for those coefficients, and the task is to determine the properties of the scheme.

This is not wholly realistic; much more likely is that we have all the discrete decisions made, but now wish to choose the coefficients. Our analysis techniques are then being applied to determine the coefficients which will pass the analysis tests.

The issues which we consider are:-

- support region
- positivity
- precision set
- boundary scheme
- convergence and continuity at extraordinary points
- artifacts

Support region

The non-zero region of the basis function is no longer just an interval, but a region of the parameter space which may have an interesting shape, which depends on the pattern of non-zero coefficients in the mask and on the arity.

If we take a given mask, its non-zero coefficients have a convex hull (the Newton polygon if we think of the generating function as a bivariate polynomial).

Using ideas similar to those used for defining the support of a univariate scheme, we can identify a frontier beyond which the influence of the mask cannot lie. We just expand the Newton polygon by the factor $a/(a-1)$ about the original vertex at which the mask is being applied.

This definitely gives an outer bound for the support. However, it is possible to invent schemes where the support does not fill this bound.

An example is given by the ternary mask

$$\begin{array}{cccccccc}
 & & & * & * & & & \\
 & * & * & * & * & * & * & \\
 & * & * & * & * & * & * & \\
* & * & * & * & * & * & * & * \\
* & * & * & * & * & * & * & * \\
 & * & * & * & * & * & * & \\
 & * & * & * & * & * & * & \\
 & & & * & * & & &
\end{array}$$

The convex hull of the mask is a 12-sided region. If we apply scaled versions of the outer bound on the support to the vertices after one iteration, they do not fill the original outer bound, thus giving us a tighter bound.

This effect happens if an edge of the convex hull of the mask has less than a vertices (where a is the arity) or if the vertices on such an edge are not evenly spread out. Essentially there must not be a gap in any edge greater than $1/(a-1)$ of the edge.

If there is such a gap the support of the basis function does not fill its outer bound; closer outer bounds can be established by looking at the bounds of the supports of the non-zero points at the next iteration.

In the particular case of a ternary scheme with a Newton polygon with an edge with no interior points, the support touches the scaled version of that edge in the Cantor set.

It is quite possible for the support region to have some edges which are fractal and others which are normal. Consider the ternary scheme with mask

$$\begin{array}{cccccccccccccccc}
 & & & & & * & & * & & * & & & & & & \\
 & & * & & * & & * & & * & & * & & * & & & \\
 & * & & * & & * & & * & & * & & * & & * & & \\
* & & * & & * & & * & & * & & * & & * & & * & \\
 & * & & * & & * & & * & & * & & * & & * & & . \\
* & & * & & * & & * & & * & & * & & * & & * & \\
 & * & & * & & * & & * & & * & & * & & * & & \\
 & & * & & * & & * & & * & & * & & * & & & \\
 & & & & & * & & * & & * & & & & & &
\end{array}$$

A binary scheme always has enough vertices on each of the edges of its Newton polygon, and so the support is never fractally bounded.

Positivity

Exactly as in the univariate case, if the non-zeroes of the mask are all positive, the limit function must be non-negative everywhere, because it is a column of a matrix which is the product of non-negative matrices.

If a scheme is positive, the limit surface will consist of pieces which lie within the convex hulls of appropriate groups of control points.

Even if a scheme is negative, we can always dilate the bounding box of the group of control points by an appropriate factor to give a box containing each piece.

Precision set

In the regular grid case this is exactly analogous to the univariate case. We can imagine that all the vertices have abscissae derived from the pattern of refinement, and we need to check their ordinates.

We need to check whether the new vertices all have the same value when the old vertices have the same value (constant precision); whether they all lie in the same linear polynomial when the old vertices are co-planar (linear precision), and whether they all lie in the same quadratic when the old vertices do (quadratic precision). There are again strict and relaxed forms of quadratic precision.

The difference in the bivariate case is that we have to check two independent linear polynomials, three independent quadratics, four independent cubics. However, the symmetry of the scheme may mean that one test satisfies two or more of these criteria.

This is all straightforward in regular regions. At extraordinary points, however, it is much harder to determine what the appropriate abscissae are to assign to the vertices. The clue to this comes from the eigenanalysis, and this point will be taken up again below.

The precision set conditions give us linear equations which the coefficients in the mask must satisfy, thus reducing the dimension of the space in which we are searching for good solutions.

Tensor product schemes inherit their precision sets from their generators.

Boundary scheme

If we are to deal with polygons with boundary (part of the reason for wanting to be able to deal with arbitrary topology) we need to know what univariate scheme is appropriate at the boundary.

This is straightforward: we just take a regular region and give it extruded values, by which all points on one grid line have their values set to one, and all others are set to zero. Then apply the mask to that data. Any grid line crossing the direction of extrusion will then give the mask appropriate to a boundary curve there.

This is the same as summing the mask in a given direction, adding together the terms which map to the same place in the old grid.

This can also be looked at in terms of the generating function. If we evaluate the generating function at $x = 1$ we get the same answer as if we add all the coefficients in a y=const direction. Taking the mean $(a(1)+a(-1))/2$ gives the sum of all the even coefficients. Such operations can save a fair amount of symbolic manipulation.

For example, in the butterfly scheme, which has the mask

$$\begin{array}{cccccccccccc}
 & & & & -1 & & -1 & & & & & \\
 & -1 & & 0 & & 2 & & 0 & & -1 & & \\
-1 & & 2 & & 8 & & 8 & & 2 & & -1 & \\
 & 0 & & 8 & & 16 & & 8 & & 0 & & /16 \\
-1 & & 2 & & 8 & & 8 & & 2 & & -1 & \\
 & -1 & & 0 & & 2 & & 0 & & -1 & & \\
 & & & & -1 & & -1 & & & & &
\end{array}$$

the old polyhedron edges have the scheme $[-1, 0, 9, 16, 9, 0, -1]/16$ projected on them, which is exactly the 4-point scheme.

If we know the support and the pattern of a scheme, it is possible to identify what standard univariate schemes might be relevant, choose one, and then use that choice to determine (linear) equations between the coefficients which ensure that it will be met.

This is all trivial for tensor products, since then the boundary scheme is just the univariate generator scheme.

Convergence and continuity at extraordinary points

The question of continuity at the extraordinary points is one which has been a central focus of research in subdivision surfaces for the last 22 years. Many schemes have been proposed which are not particularly good, but which can be significantly improved by careful design of the coefficients around extraordinary points.

- Regular continuity

The continuity throughout the regular regions can be addressed using the z-transform approach. We derive schemes which determine differences or derivatives from differences, and check that they are contractive. If a difference scheme is contractive the scheme gives a continuous limit: if a derivative scheme is contractive the scheme has a bounded derivative.

The difference here is that we now have to consider two first difference schemes, three second difference schemes, four third difference schemes,....

But again, the symmetry of the grid normally reduces the number of separate tests necessary.

This approach gives sufficient conditions and so we can be sure that if they are met the scheme really does have the claimed continuity and boundedness.

- Extraordinary points

We saw earlier that the structure of a subdivision surface was that it contained regions of regular structure, and in these regions there would be well-defined surface patches, often polynomial. Around extraordinary points the surface consists of nested rings of these regular patches, one ring for each level of subdivision[1].

This determines the continuity everywhere except at the extraordinary points themselves and their infinitesimal neighbourhoods.

We also saw that as the refinement proceeds, the extraordinary points become separated by a 'sea' of regularity, so that we can treat them as isolated.

The technique we apply to determine the nature of this tiny region is eigenanalysis.

Just as in the curve case we could determine some properties of the curve by looking at the eigenproperties of a matrix closely related to the subdivision matrix, the same thing is true in the surface case.

The first procedure was to reorder the rows and columns of the subdivision matrix so that the vertices closest to the centre of interest cam first. This gave the bi-infinite subdivision matrix a 'top left hand corner'.

The equivalent in the surface case is to take the vertices in successive rings around the centre of interest, and we note that because of the isolation of the singularities, the pattern after a subdivision step is the same as that before.

[1] Not quite. For high degree box-splines there can be more than one ring of polynomial pieces at each level, but the nesting structure remains.

When we do this we discover that the subdivision matrix has a **circulant block** structure[2], consisting of blocks each of which has the structure that a single row is rotated cyclically to give the full block. The block lower triangular structure also becomes visible, that beyond a certain point, there are no non-zero entries on or above the diagonal.

- Example: the square of the simplest scheme.

This has the mask

$$\begin{array}{cccc} & 1 & 1 & \\ 1 & 2 & 2 & 1 \\ 1 & 2 & 2 & 1 \\ & 1 & 1 & \end{array} /4,$$

and the corresponding refinement equation for a mark point at the centre of a four-sided face is

$$4\begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \\ \\ b_1 \\ b_2 \\ b_3 \\ b_4 \\ \\ c_1 \\ c_2 \\ c_3 \\ c_4 \\ \\ d_1 \\ d_2 \\ d_3 \\ d_4 \end{bmatrix} = \begin{bmatrix} 2&1&0&1&0&0&0&0&0&0&0&0&0&0&0&0 \\ 1&2&1&0&0&0&0&0&0&0&0&0&0&0&0&0 \\ 0&1&2&1&0&0&0&0&0&0&0&0&0&0&0&0 \\ 1&0&1&2&0&0&0&0&0&0&0&0&0&0&0&0 \\ \\ 2&0&0&1&1&0&0&0&0&0&0&0&0&0&0&0 \\ 1&2&0&0&0&1&0&0&0&0&0&0&0&0&0&0 \\ 0&1&2&0&0&0&1&0&0&0&0&0&0&0&0&0 \\ 0&0&1&2&0&0&0&1&0&0&0&0&0&0&0&0 \\ \\ 2&1&0&0&0&0&0&0&1&0&0&0&0&0&0&0 \\ 0&2&1&0&0&0&0&0&0&1&0&0&0&0&0&0 \\ 0&0&2&1&0&0&0&0&0&0&1&0&0&0&0&0 \\ 1&0&0&2&0&0&0&0&0&0&0&1&0&0&0&0 \\ \\ 2&0&0&0&1&0&0&0&1&0&0&0&0&0&0&0 \\ 0&2&0&0&0&1&0&0&0&1&0&0&0&0&0&0 \\ 0&0&2&0&0&0&1&0&0&0&1&0&0&0&0&0 \\ 0&0&0&2&0&0&0&1&0&0&0&1&0&0&0&0 \end{bmatrix} \begin{bmatrix} A_1 \\ A_2 \\ A_3 \\ A_4 \\ \\ B_1 \\ B_2 \\ B_3 \\ B_4 \\ \\ C_1 \\ C_2 \\ C_3 \\ C_4 \\ \\ D_1 \\ D_2 \\ D_3 \\ D_4 \end{bmatrix}$$

We may legitimately truncate the matrix at this point, since columns further to the right correspond to vertices whose support regions do not reach the centre. They cannot therefore influence the behaviour of the surface there.

The next transformation to be made is the equivalent of the odd/even split. We can permute the matrix again, so that it becomes **block circulant**, consisting of blocks (themselves block lower triangular) with the blocks arranged in a circulant pattern. The eigenvalues of the complete matrix can be determined efficiently from the eigenvalues of the individual blocks.

---

[2] This is technically true only for dual schemes. For primal ones we need to regard the centre vertex as being repeated n times, so that all the blocks become the same size. This is only a technical detail.

$$4\begin{bmatrix} a_1 \\ b_1 \\ c_1 \\ d_1 \\ \\ a_2 \\ b_2 \\ c_2 \\ d_2 \\ \\ a_3 \\ b_3 \\ c_3 \\ d_3 \\ \\ a_4 \\ b_4 \\ c_4 \\ d_4 \end{bmatrix} = \begin{bmatrix} 2&0&0&0 & 1&0&0&0 & 0&0&0&0 & 1&0&0&0 \\ 2&1&0&0 & 0&0&0&0 & 0&0&0&0 & 1&0&0&0 \\ 2&0&1&0 & 1&0&0&0 & 0&0&0&0 & 0&0&0&0 \\ 2&1&1&0 & 0&0&0&0 & 0&0&0&0 & 0&0&0&0 \\ \\ 1&0&0&0 & 2&0&0&0 & 1&0&0&0 & 0&0&0&0 \\ 1&0&0&0 & 2&1&0&0 & 0&0&0&0 & 0&0&0&0 \\ 0&0&0&0 & 2&0&1&0 & 1&0&0&0 & 0&0&0&0 \\ 0&0&0&0 & 2&1&1&0 & 0&0&0&0 & 0&0&0&0 \\ \\ 0&0&0&0 & 1&0&0&0 & 2&0&0&0 & 1&0&0&0 \\ 0&0&0&0 & 1&0&0&0 & 2&1&0&0 & 0&0&0&0 \\ 0&0&0&0 & 0&0&0&0 & 2&0&1&0 & 1&0&0&0 \\ 0&0&0&0 & 0&0&0&0 & 2&1&1&0 & 0&0&0&0 \\ \\ 1&0&0&0 & 0&0&0&0 & 1&0&0&0 & 2&0&0&0 \\ 0&0&0&0 & 0&0&0&0 & 1&0&0&0 & 2&1&0&0 \\ 1&0&0&0 & 0&0&0&0 & 0&0&0&0 & 2&0&1&0 \\ 0&0&0&0 & 0&0&0&0 & 0&0&0&0 & 2&1&1&0 \end{bmatrix} \begin{bmatrix} A_1 \\ B_1 \\ C_1 \\ D_1 \\ \\ A_2 \\ B_2 \\ C_2 \\ D_2 \\ \\ A_3 \\ B_3 \\ C_3 \\ D_3 \\ \\ A_4 \\ B_4 \\ C_4 \\ D_4 \end{bmatrix} =$$

This transformation is equivalent to carrying out a Fourier transformation on all of the data.

Let

$$a_\omega = \sum_i a_i cos(2\pi\omega i/n)$$

$$A_\omega = \sum_i A_i cos(2\pi\omega i/n)$$

$$b_\omega = \sum_i b_i cos((2\pi\omega i + \phi)/n)$$

$$B_\omega = \sum_i B_i cos((2\pi\omega i + \phi)/n)$$

$$c_\omega = \sum_i c_i cos((2\pi\omega i - \phi)/n)$$

$$C_\omega = \sum_i C_i cos((2\pi\omega i - \phi)/n)$$

$$d_\omega = \sum_i d_i cos(2\pi\omega i/n)$$

$$D_\omega = \sum_i D_i cos(2\pi\omega i/n)$$

where n is the valency of the centre.

The refinement equations then fall apart into separate groups, one group for each value of ω.

- Contributions from different frequencies

$\omega = 0$ gives a unit eigenvalue, whose eigenvectors indicate that the whole configuration converges to the centroid of the face.

$\omega = 1$ gives two equal eigenvalues[3], which we shall call λ.

The eigenvectors on the right[4], when multiplied by the configuration, give two displacements which span the tangent plane of the limit surface at the centroid of the face.

The eigenvectors on the left give the coordinates of the vertices in the **natural configuration** with respect to the coordinate system defined by those two displacements.

$\omega = 2$ normally gives two equal eigenvalues, which we shall call μ. If we want well behaved curvatures the first condition is that $\mu = \lambda^2$.

The eigenvectors on the left give the actual shape of the saddle-curvatures at the centroid, with the eigenvectors on the right, multiplied by the original configuration, giving the amplitude and phase of these contributions to the shape.

$\omega = 0$ normally gives another eigenvalue equal to μ. Its eigenvector on the left gives the shape of the cup-curvature at the centroid.

- precision set revisited

The condition that $\mu \leq \lambda^2$ and that all other eigenvalues are less than or equal to μ is a necessary condition for boundedness of the second derivative at the centre. It is not quite a sufficient condition, because the existence of geometric degeneracy (equal eigenvalues with equal eigenvectors) in the spectrum can give logarithmic divergence of the second derivative.

If $\mu < \lambda^2$, the curvature at the centre collapses to zero. This is technically continuous, but is not what we seek in a well-behaved scheme.

For actual continuity of non-zero curvature at the extraordinary point, we require that all other eigenvalues are strictly less than μ, and also that when the original configuration is laid out according to the natural configuration and all its vertices lie in a quadratic, its refinement should also lie in a quadratic.

This is effectively a quadratic precision condition, based on the vertices being parametrised according to the natural configuration.

In this particular scheme, there are a large number of eigencomponents with eigenvalue μ, and most of them do not have quadratic precision. We therefore know that the scheme is not going to give C2 surfaces unless the original configuration is highly symmetric in some way.

[3] these actually come from $\omega = \pm 1$, the equality of the two eigenvalues coming from the rotational symmetry of the scheme.

[4] This rather clumsy wording is my response to the fact that these eigenvectors are actually left eigenvectors in the standard nomenclature, but I find using that phrase very confusing, as it is counterintuitive when looking at the eigenfactors in matrix multiplication sequence.

- Design of a scheme

We normally have to choose coefficients for the mask in the vicinity of an extraordinary point, and these values, as well as being different in number because of the change in pattern, can change in value. It is most unusual for a scheme to give the coefficients near an extraordinary point unambiguously.

We therefore need to choose the values as a function of the valency of the singularity. The Fourier approach allows us to do this consistently, by treating n as an algebraic variable.

That covers all cases of valency higher than normal. For valencies less than normal (3 for quad grids and 3 or 4 for triangulations) the saddle curvature components arise in a different way, from further components within $\omega = 1$ and need special case treatment.

Artifacts

We have three kinds of artifacts to consider in the surface case.

- longitudinal artifacts.

These are the ripples which run in the same direction as the ridges we deliberately put on to the surface. They are directly analysed by considering extruded data as for the boundary analysis above, and considering the artifact behaviour of the boundary refinement scheme.

- lateral artifacts.

These are the ripples which run across the ridges we deliberately put on to a surface. If the boundary scheme is not the same for all edges in a given direction, lateral ripples are present.

An example is the binary scheme whose mask is

$$\begin{array}{ccccccccc}
 & & 1 & & 3 & & 1 & & \\
 & 3 & & 4 & & 4 & & 3 & \\
1 & & 4 & & 8 & & 4 & & 1 \\
 & 3 & & 4 & & 4 & & 3 & \\
 & & 1 & & 3 & & 1 & &
\end{array}$$

When this is applied to cardinal extruded data we find the new polyhedron is

$$\begin{array}{cccccccccc}
\ldots & 2 & & 3 & & 2 & & 3 & & \ldots \\
\ldots & & 7 & & 7 & & 7 & & 7 & \ldots \\
\ldots & 8 & & 10 & & 8 & & 10 & & \ldots \\
\ldots & & 7 & & 7 & & 7 & & 7 & \ldots \\
\ldots & 2 & & 3 & & 2 & & 3 & & \ldots
\end{array}$$

which has alternating 8 and 10 along the spine.

For this to be avoided, every row of the mask thought of as a separate univariate generating function must be divisible by $(1+z)$.

This in turn suggests that rather than choosing the coefficients for the mask itself, we should choose a smaller mask and then multiply it by $(1+x)(1+y)$, $(1+x)(1+y)(1+xy)$ or $(1+x)(1+y)(1+xy)(x+y)$, depending on the pattern of symmetry aimed for.

Longitudinal and lateral artifacts are both examples of a bivariate sampling filter. If we set the data to $z = cos(\omega_x x + \omega_y y)$, and then apply a single subdivision step, the resulting array of values will contain spatial frequencies other than (ω_x, ω_y) and these give a complete picture of the artifacts to be found in a regular grid.

- rotational artifacts

These are ripples which appear spreading radially from an extraordinary point.

Again we need to do a Fourier analysis, but in terms of rotation around the extraordinary point, rather than in terms of distance in any given direction.