Artifacts in recursive subdivision surfaces

Malcolm A. Sabin, Loic Barthe

Artifacts in Recursive Subdivision Surfaces

Malcolm Sabin and Loïc Barthe

Abstract. For a subdivision scheme to be effective it has to be possible for designers to provide relatively sparse data and achieve something close to their mental images. The difference between what is expected (or hoped for) from the subdivision scheme, and what actually emerges as a limit surface is an **artifact**. This aspect has not been much studied, and this paper provides an initial categorisation of five issues and identifies how much and how little we understand them. It is assumed that the reader is already familiar with eigenanalysis and z-transform methods for analysis of the limit surface.

§1. What Kinds of Artifacts Are There ?

We define an artifact to be any feature of the limit surface which cannot be controlled by the movement of control points at the current level of subdivision. This implies that the concept of spatial frequency is a key one. Spatial frequency components of a frequency greater than one cycle per two control points cannot be controlled, and so mechanisms which give them can be confidently called artifacts.

Preliminary work has identified two aspects of curve artifacts and five of surface artifacts which we can address:

- longitudinal artifacts on curves
- end-conditions on curves
- longitudinal artifacts on surfaces
- edge- and corner-conditions
- lateral artifacts
- radial artifacts
- rotational artifacts

The rest of this paper describes current knowledge, which is fairly complete, relating to the first five of these. The last two, although described here, will be addressed in much more detail in future papers.

Curve and Surface Fitting: Saint-Malo 2002
XXX (eds.), pp. 1–3.

ISBN XXX.

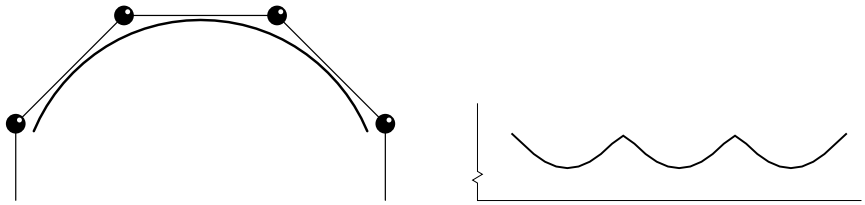

Fig. 1. (a) shows a piece of a cubic Bspline where the control points are at the vertices of a regular octagon, and (b) its curvature plot (curvature against distance). Although the curve looks smooth, the curvature plot shows up the higher frequency components. Note the broken vertical axis in the curvature plot; the peak to peak amplitude of the variation is about 0.1% of the average curvature.

§2. Spatial Frequencies

The Shannon-Whittaker theorem [15] which gives us the minimal number of samples necessary to exactly reconstruct a signal in terms of the extrema of its frequency components is just as applicable when the abscissa is a spatial dimension as when it is time.

The concept is already familiar in computer graphics under the name 'aliasing'. The subtlety that we need for surface work is that spatial frequency has not just a magnitude but a direction. If we regard the surface as being expressed as a set of components

$$P(u,v) = \sum_i A_i \cos(a_i u + b_i v + c_i)$$

the A_i are the amplitudes, $[a_i, b_i]$ the frequencies, and c_i the phases of the various terms.

Where high-frequency artifacts appear, they are not necessarily in the same direction as the dominant low frequency variations in the data.

§3. Longitudinal Artifacts on Curves

Longitudinal artifacts occur when smoothly positioned control points give a limit curve which has spatial frequency components at a frequency higher than the Shannon limit of one cycle per two samples. For B-spline curves this is always, because the Fourier transform of the B-spline basis function is not band-limited. The exercise is one of damage limitation, but fortunately the situation is not too bad quantitatively. Indeed it is often necessary to look at curvature plots before it is visible at all.

We can analyse this either by Fourier methods, by looking at the amplitude of one-cycle-per-control point frequency for data sets generated with n points round a circle (this can be measured easily by evaluating the limit curve near the control points and half-way in between them), or by z-transforms.

points	degrees		spline	degree	
per cycle	per point	2	3	4	5
3	120	0.111111	0.066667	0.030303	0.015873
4	90	0.029437	0.014059	0.003799	0.001437
5	72	0.011146	0.004726	0.000849	0.000266
6	60	0.005155	0.002040	0.000261	0.000073
8	45	0.001565	0.000576	0.000043	0.000010
10	36	0.000629	0.000224	0.000011	0.000002
12	30	0.000300	0.000105	0.000004	0.000001
16	22.5	0.000094	0.000032	0.000001	
20	18	0.000038	0.000013		
24	15	0.000018	0.000006		
32	11.25	0.000006	0.000002		
40	9	0.000002	0.000001		
48	7.5	0.000001			

Tab. 1. The figures in the rightmost four columns are the amplitudes of longitudinal artifacts measured for B-splines of degrees 2 to 5, tabulated for different numbers of control points evenly spaced around a unit circle.

The z-transform approach says that this artifact varies inversely with the $(d+1)^{\text{th}}$ power of the number of vertices per cycle, where $d+1$ is the number of $1+z$ factors in the symbol. Direct measurements of B-splines, tabulated in Table 1, suggest that the even degree B-splines in fact are better than this, giving variation with the $(d+2)^{\text{th}}$ power. This is probably due to the fact that the z-transform picks up errors along the curve as well as errors across it, and these dominate for the even degrees.

Two specific points to note are

- that the number of points needed is very low if the piece of curve turns through only a small angle. Even for quadratics, the error is only 1 part in a million at one vertex per 7.5 degrees of turn;
- that for very sparse polygons, increasing the degree does not help much.

There is hope that subdivision may be able to outperform simple B-splines in terms of longitudinal artifacts, by using the concept of geometric sensitivity. It is certainly possible to define a variant of the four-point scheme [5] in which each new vertex is positioned in such a way that the 3-point curvature estimator gives the mean of the old curvature estimates at the adjacent old points. Such a scheme is trivially proven to have circular precision — if enough consecutive control points lie on a circle, the limit curve will contain an arc of that circle. Like the circular-precision

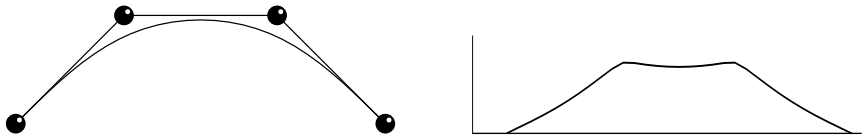

Fig. 2. The 'natural' end conditions have the effect of sending the curvature to zero at the ends. (a) shows a curve which uses this condition, and (b) its curvature plot. Note that the effect at the ends is much larger than the longitudinal variation in the interior of the curve.

scheme of Nasri and Farin [12] and unlike those of Dyn [4] and Warren [17], no prior information about radius or points/cycle is required.

§4. End-conditions on Curves

The usual end-condition applied to cubic subdivision curves can be described in three ways: as the extension of the control polygon by linear extrapolation of one extra control point at each end, as the modification of the rules near the end, or as the ad hoc appending of the original end control point after computation of the next polygon. All three give the same unsatisfactory result, which has zero curvature at the end points.

We need something more comparable to the (standard in practice) use of Bezier end-conditions, where an extra control point gives slope control at the end, and we retain all the convex hull properties. Unfortunately this amounts to using unequal intervals, and so does not fit into the subdivision world–view too easily.

The 'not-a-knot' conditions shown by Bejancu [1] to improve the approximation order in the semi-cardinal case look to be highly relevant. It is well known that this gives much better results in the interpolating spline context than the natural end conditions. A rather unsatisfactory modification of the cubic subdivision to give the not-a-knot condition is described by Nasri and Sabin [13]. Better formulations may be available, which give an effect similar to Bezier end-conditions for a spline in which the first knot interval is twice the normal, but this is still being explored.

§5. Longitudinal Artifacts on Surfaces

For quadrilateral grid schemes, these are just the tensor products of the curve artifacts. For triangle grid schemes, we can reduce the problem to a curve one by considering an initial polyhedron which is extruded in one of the mesh directions. This forces the spatial frequencies to be essentially univariate in a direction perpendicular to that of the extrusion.

What becomes interesting is what happens when the extrusion direction is not one of the mesh directions. This does not appear to have been considered at all yet. Clearly the approach of regarding the spatial frequency as a vector will be appropriate.

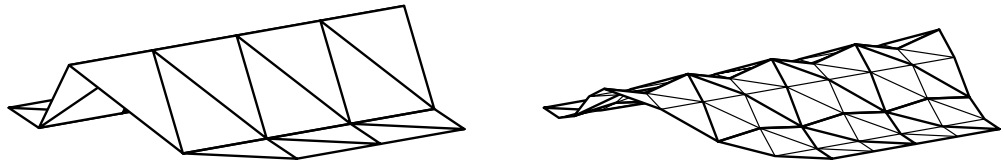

Fig. 3. (a) shows an extruded polyhedron. (b) shows the effect after one iteration. The 'dinosaur back' effect is clearly visible. There are spatial frequency components at twice the Shannon limit in the original direction of extrusion where there was no variation in the original at all.

§6. Edge- and Corner-conditions on Surfaces

In the quadrilateral grid case we merely use the tensor product structure. The main new difficulties with edge- and corner-conditions apply to schemes based on regular triangular grids, such as Loop and Butterfly [6].

For regular triangular grids there are two types of convex corner, and the two directions interact at either type of corner. Recent results indicate that it may still be possible to define 'not-a-knot' conditions on triangular grids, but the application of these to subdivision has not yet been carried out.

Levin [10] has described a scheme for controlling a triangular scheme in a way like Bezier edge-conditions, but the pattern of control points is not really intuitive, because additional vertices, interior to the surface, one per edge-vertex, are used to define the variation of tangent across the edge. Bejancu [1] has discovered that the not-a-knot condition can be applied to the 3-direction quartic box-spline (Loop) in the approximation context, and this may lead to a reasonable solution, although at present it looks as though achieving not-a-knot and the convex hull property at the same time may require changes in the rules over a region as wide as the support of the scheme. This could be unfortunate.

§7. Lateral Artifacts on Surfaces

These are an effect which has long been with us, unrecognised. It applies to the regular-grid parts of both subdivision and B-spline surfaces. While in a longitudinal artifact a spatial frequency in the original polyhedron gives an artifact component in the same direction, in lateral artifacts they cause an artifact component in a perpendicular direction.

This effect is fully understood. A lateral artifact will occur if the original polyhedron is extruded in a direction for which the symbol of the mask does not have a factor of $1 + z$. The example in Figure 3 was contrived by applying a version of the Loop scheme [11] whose mask was deliberately altered so that there was no $(1 + z)$ factor. (The mask is the set of influences of a given old vertex on the surrounding new ones.)

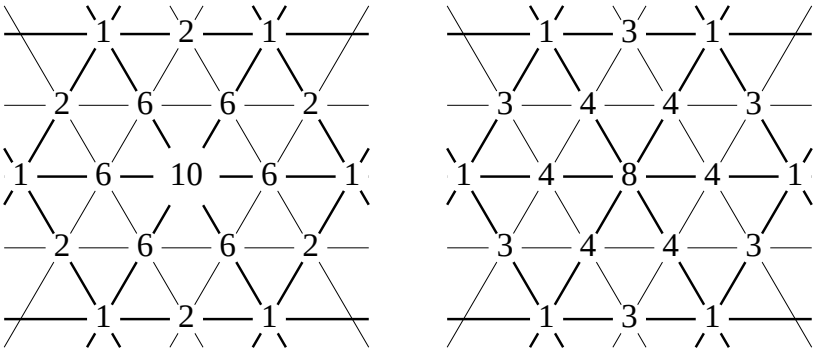

Fig. 4. The mask on the left is that of the standard Loop scheme, which has a $(1+z)^2$ factor in each mesh direction. That on the right has been altered so that although it still gives constant and linear precision, it has no $1+z$ factor in any direction.

This can be understood in the same breath as the fact that any attempt to run a feature skew to the mesh directions in a NURBS surface also causes undesired ripples. It is the same effect.

The effect of having a $(1+z)^2$ factor is that a cross-section which is linearly varying rather than merely constant gives no ripples. This is a visible difference between Simplest [14] and Velho [16].

This raises the question *'Is it better to have more directions in which* $1+z$ *factors exist, as in, for example the Simplest scheme, or Velho, or to have a higher power of* $(1+z)$ *in a few directions, as in Catmull-Clark ?'*

For schemes of arity other than 2, the relevant factors are $(1-z^a)/(1-z)$ rather than just $(1+z)$. This condition is closely related to the condition for non-fractal support described by Ivrissimtzis [8], in that any scheme with fractal support, such as the $\sqrt{3}$ schemes described by Guskov [7] and by Kobbelt [9] will have no directions free from lateral artifacts.

The lateral artifact story is thus pretty well complete. It is clear that the scheme designer needs to ensure that the mask is well-endowed with directions i with $(1+z_i)$ factors in the symbol of the scheme, and the surface designer needs to run those directions of the mesh along the ridge features of the required surface. This is pretty natural anyway. It is merely necessary to suppress the urge to try to be clever. This is perhaps one of the strongest arguments for subdivision surfaces as distinct from B-splines, that the mesh can be run locally along features, with extraordinary points where necessary in relatively featureless regions.

§8. Radial Artifacts on Surfaces

The last two effects so far identified were demonstrated by Jos Stam at a Dagstuhl meeting [3]. His original polyhedron consisted of a 14-gon upper face and a 14-gon lower face, joined by 14 rectangles, and he applied a

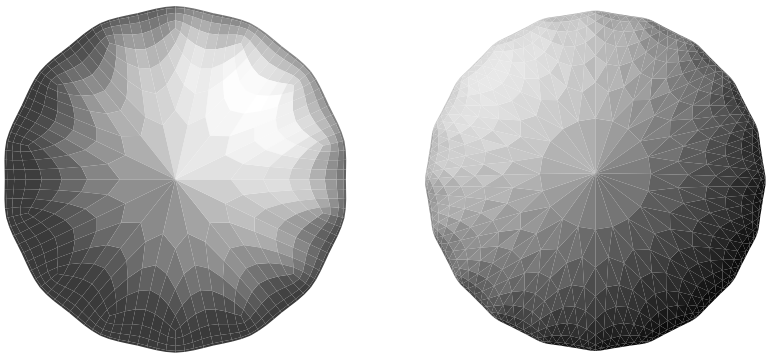

Fig. 5. The left figure shows the effect near the pole of the third iteration of the Catmull-Clark scheme on a regular 14-sided prism. The right figure shows that of the Loop scheme on an original polyhedron consisting of a vertex at each pole and 20 vertices around the equator, all joined up by 40 triangles. In both cases the facets near the pole are clearly much larger than those elsewhere. The number of vertices/faces round the equator was in each case chosen to show the artifacts most clearly.

few iterations of the standard Catmull-Clark [2] scheme. This showed two undesirable effects, long thin facets near the pole, and ridges running along lines of longitude. Very similar effects also occur in the Loop [11] scheme, and it appears likely that all schemes have the same potential problems related to extraordinary points. The question of continuity at such points is not the only important one.

Where the effect has spatial frequency components in directions radial to the extraordinary point, we call this a radial artifact; where it has spatial frequency components in directions around the extraordinary point, we call this a rotational artifact.

The specific radial artifact shown in Figure 5 is called the polar artifact, and is relatively easy to understand. What is happening is that while over the bulk of the shape each iteration halves the size of each triangle, near the extraordinary point the radial shrinkage factor is the value of the double eigenvalue λ associated with the natural configuration. If this is significantly larger than $1/2$, then after a few iterations the facets there are considerably larger than the average.

Interestingly, this is not a problem in the limit surface, although it gets worse at every iteration. It does mean that the rate of convergence is somewhat slower than quadratic, as was pointed out by Wang and Qin [18].

The polar artifact can be completely eliminated by adjusting the extraordinary point masks so that at every extraordinary point the second eigenvalue (λ) is the same as at ordinary vertices. Unfortunately the mask adjustments which give bounded curvature tend to increase λ.

We can also classify unbounded curvature at extraordinary points as a radial artifact.

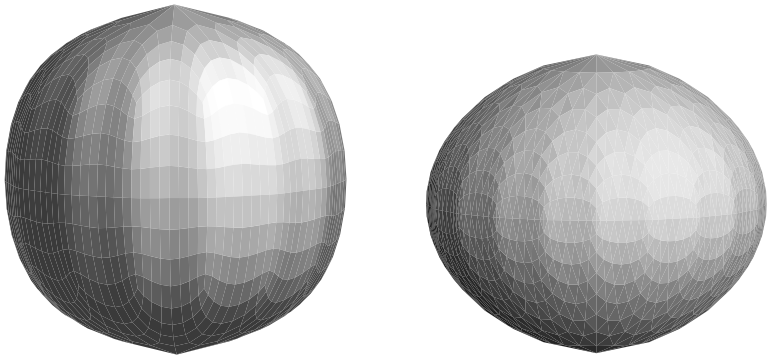

Fig. 6. The same data after 3 iterations of the two schemes (left Catmull Clark, right Loop) shows definite ridges. There is one ridge per original equatorial face/vertex, and so these ridges cannot be removed by repositioning those vertices.

§9. Rotational Artifacts on Surfaces

This effect was also shown at the Dagstuhl meeting. It is considerably harder to understand, because there are at least three potential causes:

- the high valency extraordinary vertex at the pole, and its eigenvectors.
- the low valency ones round the equator, and their eigenvectors.
- interaction between the two.

Preliminary experiments indicate that all three can contribute, and different schemes may be sensitive to different aspects. This will be reported in the detail it deserves in a separate paper.

§10. Conclusions

We are starting to understand the individual artifact mechanisms, and can in some cases control them. There still remains significant research to be done in the areas of edge conditions and rotational artifacts. More of a problem is that sometimes the adjustments we would like to make to the extraordinary point masks to control one artifact are exactly opposite to those we want to make to control another, and getting control of everything at once is the next big challenge. It may be necessary to focus on one artifact at the early stages, and on others at later stages. Geometric sensitivity may also be a useful idea to explore.

Acknowledgments. Much of the funding of the work described was provided by the EU through the Mingle project, HPRN–CT–1999–00117.

References

1. Bejancu, A., Semi-cardinal interpolation and difference equations: from cubic B-splines to a three-direction box-spline construction, Technical Report, University of Leeds, Department of Applied Mathematics; submitted for publication 2002.
2. Catmull, E. and J. Clark, Recursively generated B-spline surfaces on arbitrary topological meshes, Computer-Aided Design **10** (1978) 350–355.
3. Workshop on Subdivision in Geometric Modeling and Computer Graphics. Dagstuhl, Germany, 6-10 March 2000.
4. Dyn, N. and D. Levin, Stationary and non-stationary binary subdivision schemes, in *Mathematical Methods in Computer Aided Geometric Design II*, T. Lyche, and L. L. Schumaker (eds.), Academic press, 209-216 (1992).
5. Dyn, N., J. Gregory, and D. Levin, A four-point interpolatory subdivision scheme for curve design, Comput. Aided Geom. Design **4** (1987) 257–268.
6. Dyn, N., D. Levin and J. Gregory, A butterfly subdivision scheme for surface interpolation with tension control. ACM Trans. on Graphics **9** (1990) 160–169.
7. Guskov, I., Irregular subdivision and its applications, Princeton University, 1999.
8. Ivrissimtzis, I. P., M. A. Sabin and N. A. Dodgson, On the support of recursive subdivision, preprint, available as Tech. Rep't No. 544, Univ. of Cambridge Computer Laboratory (2002).
9. Kobbelt, L., $\sqrt{3}$-Subdivision, Proceedings of SIGGRAPH 2000, Computer Graphics Proceedings, (2000), 103-112.
10. Levin, A., Combined subdivision schemes for the design of surfaces satisfying boundary conditions, Comput. Aided Geom. Design **16** (1999), 345-354.
11. Loop, C. T., Smooth subdivision surfaces based on triangles, master's thesis, University of Utah, 1987.
12. Nasri, A. H. and G. Farin, A subdivision algorithm for generating rational curves, Journal of Graphics Tools **3** (2001) 35–47.
13. Nasri, A. H. and M. A. Sabin, Taxonomy of interpolation conditions on recursive subdivision curves, The Visual Computer **18** (2002) 259–272.
14. Peters, J. and U. Reif, The simplest subdivision scheme for smoothing polyhedra, ACM Trans. on Graphics **16** (1997), 420–431.

15. Shannon, C. E. and W. Weaver, *The Mathematical Theory of Communication*, University of Illinois Press, Urbana, (1949)
16. Velho, L., Quasi 4–8 subdivision, Comput. Aided Geom. Design **18** (2001) 345–357.
17. Warren, J. and H. Weimer, *Subdivision Methods for Geometric Design* Morgan Kaufmann, San Francisco (2002), chapter 4.4 *A smooth subdivision scheme with circular precision.*
18. Wang, H. and K. Qin, Estimating Subdivision Depth of Catmull-Clark Surfaces, submitted.

Malcolm Sabin
Numerical Geometry Ltd.
26 Abbey Lane, Lode
Cambridge, England CB5 9EP
malcolm@geometry.demon.uk

Loïc Barthe
RWTH-Aachen, Informatik VIII
Ahornstrasse 55, 52074 Aachen,
Germany
barthe@informatik.
rwth-aachen.de