Subdivision versus NU(R)BS

I'm not going to try saying why Subdivision surfaces are better. There are arguments on both sides, so what I can do is list some of the advantages that each has. Neither list should be taken as definitive.

First, however, the question of the R.

The algebraic geometers have long known that even in the plane, curves expressed in polynomial implicit form $f(P) = 0$ cannot all be represented as rational polynomial parametrics $P = F(t)/g(t)$ if the degree of the implicit is greater than two. If the degree is equal to 2, there are curves like the circle which can be represented by the rational form $P = F(t)/g(t)$ but not by the more limited form $P = F(t)$. In fact the only polynomial quadratics are all parabolae !

When the salesmen of the early CAD systems tried to sell into the aircraft industry, they found that the system used for outer mould lines was called 'conic lofting' they rapidly added the use of the denominator to their B-spline based surface systems, in order to be able to claim that 'they could do conics'. In fact such systems could not match the actual conic lofting definitions, but the CAD systems got the R added anyway. (Boeing bought CATIA, a system which at the time didn't have the R.)

Adding the R allows a bigger space of possible shapes, but the only people I know who understand well enough how to exploit this are professors at Austrian Universities. For the rest of us it is just extra hassle. However, if you just ignore it and set all the weights to 1 it doesn't make things any worse. If a surface you want to treat as the domain for analysis has non-unit weights, the IGA basis functions are not significantly different in any of the important (e.g. convergence) properties, and so you can still quietly ignore them. Yes, the weights do have to be positive, but that is an easy restriction.

Subdivision surfaces are very similar to NURBS in the fact that they are defined by a collection of control points, and the actual surface is defined by a set of basis functions multiplying those control points. So you can include weights in the formulation if you want do. I have not seen anybody bother except Tom Cashman, whose mandate was to make a superset of NURBS and subdivision. The R is not an issue in the B-spline versus Subdivision debate.

Similarities

As mentioned above, both approaches can be described in terms of control points and basis functions. In fact NURBS have a very well defined subdivision construction.

Differences

NURBS are specified by rectangular arrays of control points - marvellous for extruded things like wings if you ignore the tips or for fuselages if you ignore the front and back. They have a well defined parametrisation, so there are lots of nice algorithms which can address them as single entities, rather than as the collection of biparametric polynomials (Bézier patches) which you can also look at them as.

Subdivision looks quite different when you want to specify one: the control points are linked by faces which are not always quads and can meet at places which do not have four faces around them. The definition of the surface itself is expressed in terms of 'how to make a denser representation' and doing that an infinite number of times. We don't actually do that, except for computer graphics.

This potential non-rectangularity means that

(i) You can represent shapes where the flow of the shape is not basically rectangular, and still have boundaries which are polygons of control points.

(ii) There may not be any single parametrisation of what you would like to treat as a single geometric object.

Triangulations have a similar structure and the same impossibility of always having a single parametrisation.

In fact we should probably be considering triangulations as well as nurbs and subdivision. There are also subdivision schemes which are based on the control polyhedron being a triangulation.

Note that although subdivision is articulated in terms of 'how to make a mesh denser' it actually has a structure of basis functions in a very similar way to NURBS, and we know how to evaluate points (and derivatives) anywhere on the limit surface without playing with the mesh. Although there is no global parametrisation there are local ones.

NURBS advantages

(i) There is only one definition of NURBS.

(ii) All the CAD systems use it.

There are some small differences, relating to whether they can accept surfaces singular in some way which nobody should ever have designed, but at the IGES file level they all have degrees in the two directions (almost always the same), knot vectors in the two directions (multiplicity up to the degree being permitted in the interior of each vector) and an array of control points. Multiplicity of degree+1 at the ends (referred to as *Bézier end conditions* is not usually enforced, but is the sensible default assumption.

(iii) you can treat a NURBS as a single parametric surface, so that there are reasonably efficient interrogation algorithms.

NURBS disadvantages

The limitation to a rectangular shape. This forces the use of trimming, which in turn leads to lack of geometric watertightness, and, more subtly, to forcing regularity on to shapes which don't really want to be. This causes the 'dinosaur back' effect in places where the knot lines run diagonally to the flow of the shape.

Subdivision disadvantages

(i) There is a very large zoo of different subdivision schemes, and new ones are regularly invented. For a simulation reference geometry we have the question *'which subdivision scheme ?'*, which is not going to be easy to choose.

(ii) At *Extraordinary points*, which are the limit points corresponding either to non-4-spoked control points or non-4-sided control faces, the limit surface consists of an infinite regression of biparametric pieces. For some schemes the curvature at such places is infinite and even the sign of the Gaussian curvature is not related to any estimator from the control polyhedron.

(iii) The mathematicians picked this up, but in fact the Hölder ontinuity which they go on about, is not the problem. The practical problem is that there are wriggles in the curvature of the surface in a non-infinitesimal region round the EV. Tuning can stop the divergence to infiinity and minimise the wriggles, but not reduce the wriggles to zero.

However, remember that if you join nurb corners breaking the regularity, the continuity is even worse. If you join trimmed regions in places where you want tangency you probably have tiny steps. Once upon a time the mould lines of an aircraft were defined as the inside skin, and there were lap joints between skin panels, but that is not usual nowadays.

Subdivision advantages

(i) You can fill a region which is not four-sided with something that is guaranteed to be continuous and tangent-plane continuous.

(ii) You can make the defining polyhedron follow the flow of the surface shape that you want, thus minimising the kind of wriggles that we call *lateral artifacts* or *dinosaur backs*.

Some of the possible subdivision schemes

Catmull-Clark

This was one of the first to be published. In the regular regions it gives exactly the same shapes as bicubic NU(R)BS. In the form published in 1978 it had two problems

(i) at EVs the curvature diverges, negative curvature diverging slightly faster than positive. As you approach the EV the ripples in the curvature get bigger and bigger, with the wavelength shrinking.

(ii) at the edges of the complete surface the curvature across the edge is zero, so that the Gaussian curvature cannot be positive anywhere on the edge. The boundary curves are actually the cubic B-splines defined by the boundary polygon.

Analysis-suitable Catmull-Clark

This is a name that I have coined, for what I currently regard as the front-runner.

The divergence to infinity is cured by appropriate choice of the coefficients in the construction. There are three coefficients to be chosen, and two of them can be used to ensure the *'bounded curvature'* condition. The third can be used to minimise the size of the ripples which then stay the same size (with shrinking wavelength) as the EV is approached. There are actually three more coefficients, but it is the first three which have most effect. Note that the ripples stay the same size on curvature: the amplitude of the surface ripples shrinks quadratically.

The zero curvature at the boundary is also fixed by slightly complicating the rules at the boundary to give Bézier edge conditions.

In the regular case this is exactly the bicubic B-spline with Bézier edge conditions.

The name is designed to sound safe in the community. The Catmull-Clark half is because it is essentially the same as the standard in computer animation, the Analysis-suitable is also a phrase familiar in this general context. However, if we try to push this as a standard for analysis modelling we will need to make the exact formulations publicly available and also do quite a lot of 'soft-selling'. I would rather soft-sell this than any of the other candidates.

There is a job for me here, to articulate exactly what the construction needs to be.

Doo-Sabin

Published in the same journal issue as Catmull-Clark, this was the third subdivision scheme historically. The first was a biquadratic by Daniel Doo described in a conference proceedings just a few months earlier. Doo-Sabin is a much better quadratic which behaves almost perfectly, except that because it is only quadratic you tend to get detectable bulges even when the mesh follows the flow (*'Longitudinal artifacts'*. The longitudinal artifacts from cubics are significantly smaller.

Loop

This is a subdivision scheme defined over triangulations. It should have the same tunability around EVs, but its behaviour at the boundaries of the surface are not so easily fixed. The curvature in two directions across the boundary is zero, so the Gaussian curvature is almost always negative and I don't know whether some equivalent of Bézier edge conditions can be invented. They will certainly be more awkward to implement.

Butterfly

This is an interpolating scheme defined over triangles. It is fractal everywhere, so there are no polynomial pieces. But its limit surface does pass through the control points.

4-8

This is a less well-known scheme which has a lot of nice properties. In the regular regions the limit surface consists of triangles in a St Andrew's cross arrangement in a quad grid. Unfortunately it is not directly NURBS compatible

Tom Cashman's NURBS compatible subdivision

If you really want Non-Uniformity and general degree, this is the only real option. However, it involves some arbitrary steps to avoid some nasty non-uniform behaviour and to make the standard EV theory applicable, and is probably a bit complicated to implement.

Finite Fillings

There is yet another contender, which is not a subdivision scheme. It is the use of quads surrounding an EV which are each themselves made up of 9 bicubic pieces. It gives G1 continuity rather than C2, but avoids the infinite regress around EVs which subdivision gives. This has some track record as being what the T-spline software uses aroung EVs rather than subdivision.

Context

We are looking at this kind of geometry because IGA wants to use the same basis functions as the geometric definition of the shape. If the basis functions are not the same, then we have been able to make finite element meshes for decades.

This new requirement brings criteria, challenges and opportunities.

One issue is that of Adaptivity. With Spline-based elements it is possible to get fully nested local refinement extremely simply. You can add one freedom at a time exactly where you want it. This means that the non-uniformity issue is not so important, because local refinement can be used before the analysis to get the spatial frequencies of the mesh matching what the likely solution field will need. Nested local refinement is trivial for subdivision because the original descriptions of the scheme are exactly the recipe. I think you can achieve it for finite fillings, but the regions round EVs will need to have a different sort of code, which may complicate implementation.