A Circle-Preserving Variant of the Four-Point Subdivision Scheme

Malcolm A. Sabin and Neil A. Dodgson

Abstract. The four-point curve subdivision scheme is one of the classic reference points of subdivision theory. It has effective C2 continuity, although the curvature at the data points actually diverges slowly to infinity as very large numbers of subdivision steps are taken. However, it has rather large longitudinal artifacts, so that points interpolated around a curve of almost constant curvature are fitted by a curve with significant variations of curvature. We describe here a geometry-sensitive variant of this scheme which does not have this problem. In fact circles are reproduced exactly with any spacing of the initial data.

§1. The Four-Point Scheme

The four point scheme is a uniform stationary subdivision scheme with the mask $[-1,0,9,16,9,0,-1]/16$. It was first described by Dyn, Levin and Gregory in [3], although the functional version had already been described by Dubuc in [1]. There is a new vertex at each old vertex, and also a new vertex associated with each edge of the control polygon, and these new vertices are given by the stencil $[-1,9,9,-1]/16$. Thus, in Figure 1, $16P = 9[B+C] - [A+D]$, applying to each coordinate independently.

It is well-known [2] that this scheme has a limit curve which is almost C2 ("almost" is due to a Jordan block in the eigenanalysis at the dyadic points, so that the second divided differences there increase at each subdivision step by a fixed amount proportional to the original fourth differences), but that the shape is somewhat prone to longitudinal artifacts (the ripples which occur at one cycle per data point, visible primarily in the curvature plots). The magnitude of this artifact (tabulated in Sect. 9) is about three times as large as that of the cubic B-Spline, and this scheme

Mathematical Methods for Curves and Surfaces: Tromsø 2004
M. Dæhlen, K. Mørken, and L. L. Schumaker (eds.), pp. 1–12.

ISBN 0-9728482-4-X

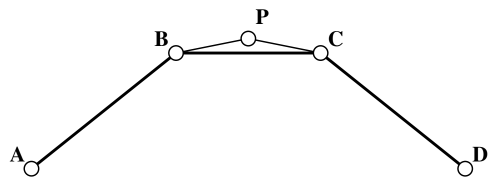

Fig. 1. The four-point construction.

has a smaller artifact than the quadratic B-spline only when there are more than eight points per full circle in the original polygon.

Figures 2 and 3 illustrate the artifact on a dataset with eight points per cycle. The irregularity at the left hand side of the curve plots and at the ends of the curvature plots are due to the fact that the data is closed, but not cyclic, and no particularly sophisticated end-conditions have been applied.

From the stencil we may determine that the second divided difference at P is the mean of the second divided differences of the original polygon at B and C:

$$\begin{aligned} 8[B+C-2P] &= 8[B+C] - 16P \\ &= 8[B+C] - 9[B+C] + [A+D] \\ &= [A+D] - [B+C] \\ &= [A+C-2B] + [B+D-2C] \\ \frac{[B+C-2P]}{1/4} &= \frac{[A+C-2B]+[B+D-2C]}{2} \end{aligned}$$

This fact was known to Floater, who observed in [5] that taking the harmonic mean instead of the arithmetic mean gave the scheme the property of convexity preservation. Other convexity preserving variants have been proposed by Le Mehauté and Utreras [8], and by Marinov, Dyn and Levin [9]. A recent paper by Kuijt and van Damme [7] has a good list of references to earlier papers.

§2. A Circle-Preserving Variant

Artifacts of the magnitude generated by the four-point scheme are unacceptable when dealing with curves and surfaces where fairness matters, and this motivated the search for a scheme which is essentially artifact-free.

We observe that the four-point scheme is essentially function-based. When used to fit functions which are polynomials of degree up to three it is perfect, reproducing the original function exactly.

The fact that taking the mean of second differences gives good behaviour on functions led to the idea that if we want curvature to vary

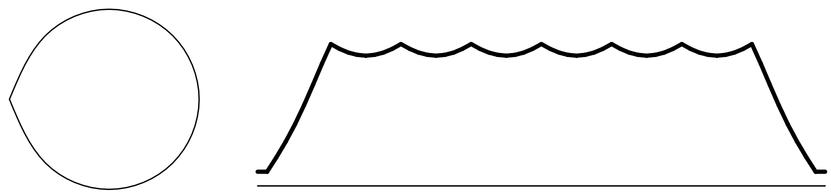

Fig. 2. (left) Cubic Bspline Curve (right) Curvature Plot.

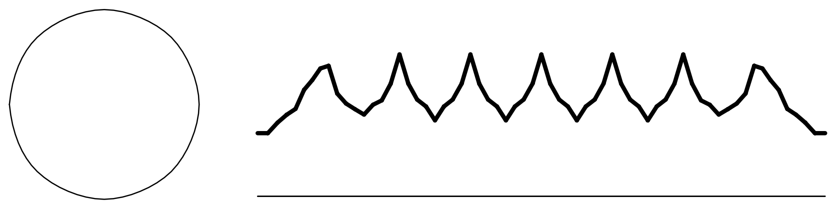

Fig. 3. (left) Four-Point Curve (right) Curvature Plot.

smoothly we should make the curvature at each new vertex equal to the mean of the curvatures at the adjacent old points.

The measure of curvature used is the Curvature Axis Vector defined as the vector perpendicular to the plane of the circle through three points, with a magnitude inversely proportional to the radius. It was necessary to produce a vector measure, because a scheme capable of defining twisted curves (non-planar curves in 3 dimensions) was required. The curvature vectors at B and C are given by

$$V_B = \frac{[C-B] \times [B-A]}{|C-B||B-A||A-C|}$$
$$V_C = \frac{[D-C] \times [C-B]}{|D-C||C-B||B-D|}$$

and that at P by

$$V_P = \frac{[C-P] \times [P-B]}{|C-P||P-B||B-C|}.$$

Note that this vector does not lie in the osculating plane, but perpendicular to it. This means that it remains constant as a point moves round a circle, rather than rotating to point always towards the centre.

Setting $V_P = [V_B + V_C]/2$ forces P to lie on a specific circular arc which passes through B and C. In order to produce a specific point P we need to identify which point of that arc to select. An obvious possibility,

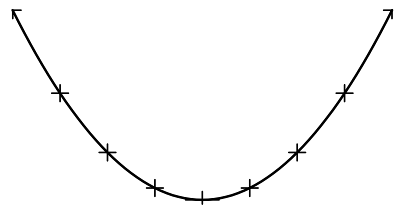

Fig. 4. Four-point applied to points on a quadratic function.

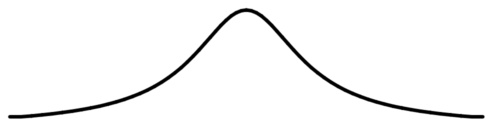

Fig. 5. Curvature plot thereof.

mentioned by Kuijt and van Damme [7], is to compute the intersection with the perpendicular bisector plane of BC, but this would result in the spacing of points remaining uneven as subdivision proceeded wherever a long edge was adjacent to a short one.

To make such unevennesses blend out as subdivision proceeds, we instead choose a new point closer to the end which has the shorter adjacent edge. The particular algorithm used is to apply the condition that if the lengths of the edges form a geometric sequence with ratio ρ before subdivision, the same will be true after subdivision, with a ratio of $\sqrt{\rho}$. It is clear that ρ will then converge towards unity. A computation which achieves this is to choose P so that

$$\frac{|P-B|}{|P-C|} = \sqrt{\frac{|C-A|}{|D-B|}}.$$

The locus of points whose distances to two fixed points are in a given ratio is a sphere, and so we have to find the point in which a given circular arc meets a given sphere. There is always a solution, because one end of the arc is inside the sphere and the other is outside.

In fact, because the ratio of these distances is known before the point has to be computed, to improve the solution still further, the arc is set to be that which has a curvature vector an appropriately weighted mean of the end curvatures, rather than just the mid-value.

§3. End Conditions

This interpolant supports a full range of end conditions. We construct it by setting the curvature vector at the end point, (A), in appropriate ways, in place of the 3-point formula which is no longer available at the ends, and by setting the ratio for division of AB to $\sqrt{|AB| : |BC|}$, because we can only estimate the geometric progression ratio from one side. Let the end point be A and the curvature vector there be V_A.

End-condition	computation of V_A
fixed direction, T_A:	$V_A := [B-A] \times T_A/([B-A].[B-A]\|T_A$
dummy point, Z:	$V_A := [B-A] \times [A-Z]/(\|AB\|\,\|ZA\|\,\|ZB\|)$
"natural" end condition:	$V_A := 0$
constant curvature:	$V_A := V_B$
"not-a-knot":	$U_A := \frac{(\|AB\|+\|BC\|)V_B-(\|AB\|)V_C}{\|BC\|}$
	$V_A := U_A - \frac{U_A.[B-A]}{\|AB\|^2}[B-A]$

Note that, in the "not-a-knot" condition, V_A has to be adjusted so that it is perpendicular to AB. It may also need to be reduced if its magnitude is greater than $|AB|$, as might happen if $|BC| > |AB|$ or $V_B.V_C < 0$.

§4. Behaviour in the Limit

Consider the case after many iterations, when consecutive vertices are locally almost equispaced along a straight line. Place the coordinate system at the midpoint of BC, with its x-axis pointing along BC, and its xz plane containing D. We may express the coordinates of the four points as

$$\begin{aligned} A &= (-3 + x_A, y_A, z_A) \\ B &= (-1 + x_B, 0, 0) \\ C &= (1 + x_C, 0, 0) = (1 - x_B, 0, 0) \\ D &= (3 + x_D, 0, z_D) \end{aligned}$$

where the values x_A, x_B etc. are all small relative to the distances between the vertices. Then to a first approximation in x_A, x_B etc.,

$$\begin{aligned} |AC| &= 4 + x_C - x_A \\ &= 4 - x_B - x_A \\ &= 4(1 - (x_B + x_A)/4) \\ |BD| &= 4 - x_B + x_D \\ &= 4(1 - (x_B - x_D)/4) \\ |AC|/|BD| &= (1 - (x_B + x_A)/4)/(1 - (x_B - x_D)/4) \\ &= 1 - (x_B + x_A)/4 + (x_B - x_D)/4 \\ &= 1 - (x_A + x_D)/4 \end{aligned}$$

$$
\begin{aligned}
\sqrt{|AC|/|BD|} &= 1 - (x_A + x_D)/8 \\
&= (x_P - (-1 + x_B))/((1 - x_B) - x_P) \\
&= ((1 - x_B) + x_P)/((1 - x_B) - x_P) \\
&= (1 - x_B)(1 + 2x_P/(1 - x_B)) \\
2x_P/(1 - x_B) &= -(x_A + x_D)/8(1 - x_B) \\
x_P &= -(x_A + x_D)/16,
\end{aligned}
$$

which is exactly the perturbation from $(B{+}C)/2$ predicted by the standard four point scheme.

Note that the magnitude of x_P is a factor of 8 less than the original x_A etc. while the spacing is only halving. The x's are therefore converging to zero cubically, so that the ignored terms (quadratic in the x's) are converging sextically.

We may therefore concern ourselves in the next stage of the proofs solely with P being a nominal midpoint, not bothering with the fact that in fact we bias the curvature calculation by taking a weighted mean. We also ignore the x's from this point in order to shorten the algebra.

$$
\begin{aligned}
V_B &= [2, -y_A, -z_A] \times [2, 0, 0]/(2 * 4 * 2) \\
&= [0, -z_A, y_A]/8 \\
V_C &= [2, 0, 0] \times [2, 0, z_D]/(2 * 4 * 2) \\
&= [0, -z_D, 0]/8 \\
V_P &= (V_B + V_C)/2 \\
&= [0, z_A + z_D, y_A]/8 \\
&= [1, y_P, z_P] \times [1, -y_P, -z_P]/(1 * 2 * 1) \\
&= [0, 2z_P, -2y_P]/2 \\
z_P &= -(z_A + z_D)/8 \\
y_P &= -y_A/8.
\end{aligned}
$$

Again, the deviations from the 4-point scheme must be converging as the sixth power of the shrinkage rate.

We thus see that both along BC and perpendicular to it, for small deviations from the straight, those deviations behave almost exactly as the 4-point scheme would require, and the quadratic terms which we have ignored above are converging much better than quartically.

Quartic convergence is sufficient to argue that the continuity behaviour of the scheme in the limit will be the same as the original four point scheme. The method of proof is shown by Dyn and Levin in [4].

§5. Implementation

This scheme can be implemented straightforwardly as a sequence of passes through the data. Let the given control points be numbered from 0 to n in sequence, and the spans numbered from 1 to n.

$$\begin{aligned}
\forall j \in 1 \ldots n : d_j &:= |P_j - P_{j-1}| \\
\forall j \in 1 \ldots n-1 : \{e_j &:= |P_{j+1} - P_{j-1}| \\
V_j &:= [P_{j+1} - P_j] \times [P_j - P_{j-1}]/(d_j * d_{j+1} * e_j)\} \\
V_0 &:= \text{ appropriate end conditions.} \\
V_n &:= \quad \text{(see section 3 above.)} \\
\forall j \in 2 \ldots n-1 : \{dl_j &:= \sqrt{e_{j-1}} \\
dr_j &:= \sqrt{e_j}\} \\
dl_1 &:= \sqrt{d_1} \\
dr_1 &:= \sqrt{d_2} \\
dl_n &:= \sqrt{d_{n-1}} \\
dr_n &:= \sqrt{d_n} \\
\forall j \in 1 \ldots n : \big\{Vnew_j &:= [dr_j * V_{j-1} + dl_j * V_j]/(dr_j + dl_j) \\
R_j &:= [P_j - P_{j-1}] \times Vnew_j \\
c_j &:= \sqrt{(1 - R_j.R_j)} \\
Q_j &:= [P_j + P_{j-1} + |P_j - P_{j-1}|R_j]/2c_j \\
h_j &:= c_j(1 - c_j)/R_j.R_j \\
Pnew_j &:= \frac{P_{j-1}dr_j^2 + 2h_jQ_jdr_jdl_j + P_jdl_j^2}{dr_j^2 + 2h_jdr_jdl_j + dl_j^2}\}.
\end{aligned}$$

Erratum:
Q_j := (P_{j-1} + P_j) / 2
+ |P_j - P_{j-1}| * R_j / (2 * c_j)
Possible Erratum:
h_j := c_j
This has not been verified as an
error but is a plausible alternative.

The only aspect which is not trivial is the computation of the new point as the intersection of a sphere and a circular arc. It turns out that the parameter value of the standard rational Bézier quadratic representation of the arc is directly related to the ratio of the distances of the point from the end-points:

$$t = \frac{dl}{dl + dr},$$

and so all that is necessary is to determine the mid-control point and its homogeneous weight and thence evaluate the required point in the usual way. The relationship for t is not obvious and is proved in Sect. 8.

§6. Example

The equivalent of Figures 2 and 3 is, by design, utterly featureless. The limit curve is exactly a circle, and the curvature plot a straight line. We

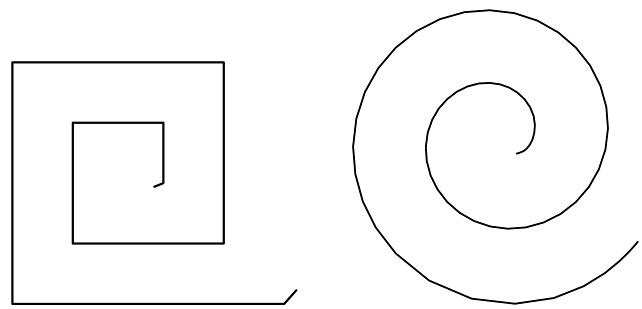

Fig. 6. (left) original control polygon (right) interpolated curve.

Fig. 7. corresponding curvature plot.

therefore show here instead a more typical example, where the curvature varies. The control points are spaced along a "rectangular spiral" with additional points added close to the end points to control the slopes near the ends. This illustrates the facts that

(i) under circumstances of smoothly varying curvature the new method gives a smooth curvature plot.

(ii) the method is tolerant of unevenly spaced data.

This figure makes it look as if the new scheme is C2, but as far as we can determine it is not. According to [3], the continuity level is the same as that of the stationary scheme to which this scheme converges. That scheme is the four-point scheme and so we claim $C^{2-\epsilon}$, like that scheme.

There may be a more sophisticated analysis based on the fact that the eigenvectors involved in the Jordan block become equal only in the limit, but it is unlikely that this will improve the continuity level. Our experience with the ternary 4-point scheme [6] throws light on this. That scheme has close eigenvalues and the curvature behaviour, while technically C2, is, in practice, actually no better than the standard 4-point scheme, having gross curvature spikes near the original control points.

Indeed, in the standard 4-point scheme the infinities of curvature are triggered by the fourth difference in the polygon. We therefore expect them to be triggered in this variant by the second differences of curvature. Figure 6 was deliberately designed to have definite but small first

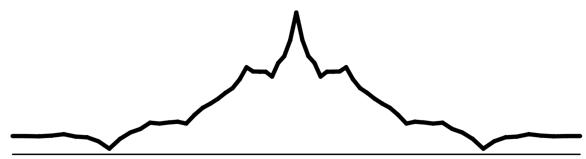

Fig. 8. curvature plot with 2nd difference of curvature vectors.

differences of curvature and close-to-zero second differences.

If we try points equally spaced in x on the parabola $y = x^2$, which is within the precision set of the standard scheme (see Figures 4 and 5) the curvature varies, and near the centre there are significant second differences of curvature. The new scheme gives the curvature plot of Figure 8, which shows irregularities similar to those of Figure 3.

§7. Conclusions

A geometry sensitive subdivision scheme has been described which gives very low longitudinal artifacts in situations where the curvature is varying smoothly. Important ideas exploited here which are relevant to the development of other subdivision schemes (both curves and surfaces) are:

- the use of averaging of second differences to give the coefficients,
- the use of curvature measures in place of second differences,
- separation of the positioning of new points within the abscissa space from positioning perpendicular to the image of that space.

§8. Proof of Ratio Property

If $V = 0$, the arc between P_{j-1} and P_j is a straight line, and the point at the required ratio can be constructed directly.

If $V \neq 0$ it defines a plane in which the arc lies. Figure 9 shows the configuration in this plane. A and B are local names for P_{j-1} and P_j. M is the midpoint of the other arc of the circle, Q is the middle Bézier control point of the rational quadratic representation of the circle. T is the point on AB satisfying $|AT|/|TB| = t/(1-t)$ being the required ratio of distances.

Claim: $|AP|/|PB| = |AT|/|TB|$.

Proof:

$$\text{In the circle } APBM,\ |AM| = |MB|$$
$$\text{and hence angle } APM(=APT) = MPB(=TPB)$$
$$\text{and so } \sin(APT) = \sin(TPB).$$

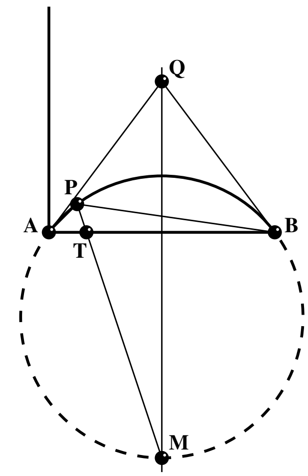

Fig. 9. Computation of P.

$$\begin{aligned}
\text{Because } & ATP + TPB = \pi, \\
\text{we also have } & \sin(ATP) = \sin(TPB). \\
\text{In triangle } APT,\ & |AT|/\sin(APT) = |AP|/\sin(ATP). \\
\text{In triangle } PTB,\ & |TB|/\sin(TPB) = |PB|/\sin(PTB). \\
\text{Hence } & |AT|/|AP| = \sin(APT)/\sin(ATP) \\
& = \sin(TPB)/\sin(TPB) = |TB|/|PB| \\
\text{and so } & |AT|/|TB| = |AP|/|PB|. \quad \square
\end{aligned}$$

§9. Longitudinal Artifact Measurement

We measure the amplitude of the longitudinal artifacts by placing points equally around a circle (with spatial frequency ω measured in terms of its reciprocal, points per cycle), refining just once, and then measuring the Fourier components, at frequencies of $\omega + 1/2$ and $\omega - 1/2$, of the new polygon. This measurement is carried out over a cyclic structure, so that end-conditions do not affect it. This gives four components: in- and out of-phase, and upper and lower sidebands. The two sidebands are combined, and the phase with larger effect is taken. In fact the in-phase component gives the visible artifact: the out of phase component gives the uneven spacing along the curve.

We also take the fundamental Fourier component as a measure of approximation order.

The reason why only one stage of refinement is made is that the bulk of the damage is done at the first step. Later stages introduce components at still higher frequencies, which alter the shape of the curve between the original data points, but do not significantly alter the magnitude of the artifact effect.

Quadratic B-spline

frequency	artifact	fundamental
8	0.056042	0.788580
16	0.007425	0.943456
32	0.000941	0.985623
64	0.000118	0.996390

Approximation error $= 14.784 * n^{-2}$
Artifact magnitude $= 30.969 * n^{-3}$

Cubic B-spline

frequency	artifact	fundamental
8	0.021446	0.728553
16	0.001448	0.925328
32	0.000092	0.980877
64	0.000005	0.995190

Approximation error $= 19.7 * n^{-2}$
Artifact magnitude $= 97.253 * n^{-4}$

Four-point

frequency	artifact	fundamental
8	0.058058	0.941941
16	0.004235	0.995764
32	0.000275	0.999724
64	0.000017	0.999982

Approximation error $= 291.29 * n^{-4}$
Artifact magnitude $= 291.29 * n^{-4}$

Acknowledgments. This work was partially funded by the Engineering and Physical Sciences Research Council under grant no. GR/S67173/01. We are also grateful to the referee for several useful suggestions.

References

1. Dubuc, S., Interpolation through an iterative scheme, J. Math. Anal. Appl. **114**, (1986), 185–204
2. Dyn, N., Interpolatory Subdivision Schemes, Tutorials on Multiresolution in Geometric Modelling, A.Iske, E.Quak and M.S.Floater (eds), Springer–Verlag, 2002, ISBN 3–540–43639–1, 25–50

3. Dyn, N., D. Levin, and J. A. Gregory, A 4-point interpolatory scheme for curve design, Comput. Aided Geom. Design **4** (1987), 257–268
4. Dyn, N. and D. Levin, Analysis of asymptotically equivalent binary subdivision schemes. Math. Anal. Appl. **193** (1995), 594–621
5. Floater, M. S and C. A. Micchelli, Nonlinear Stationary Subdivision, in *Approximation Theory*, N. K. Govil, R. N. Mohapatra, Z. Nashed, A. Sharma and J. Szabados (eds.) Marcel Dekker, 1998, 209–224
6. Hassan, M. F., I. P. Ivrissimtzis, N. A. Dodgson and M. A. Sabin, An interpolating 4-point C^2 ternary stationary subdivision scheme, Comput. Aided Geom. Design **19** (2002), 1–18
7. Kuijt, F. and R. van Damme, Shape Preserving Subdivision Schemes, J. Approx. Theory **114** (2002), 1–32
8. Le Mehauté, A. and F. I. Utreras, Convexity-preserving interpolatory subdivision, Comput. Aided Geom. Design **11** (1994), 17–37
9. Marinov, M., N. Dyn, and D. Levin, Geometrically controlled 4-Point Interpolatory Schemes, Advances in Multiresolution for Geometric Modelling, N.A.Dodgson, M.S.Floater and M.A.Sabin (eds), Springer-Verlag, 2005, ISBN 3–540–21462–3, 301–315

Malcolm A. Sabin and Neil A. Dodgson
Computer Laboratory
University of Cambridge,
Cambridge, CB3 0FD
malcolm@geometry.demon.co.uk
http://www.damtp.cam.ac.uk/mas33
nad@cl.cam.ac.uk
http://www.cl.cam.ac.uk/users/nad