Journal of Prime Research in Mathematics Vol. **3***(2007), 129-139*

QUALITY SURFACE CONSTRUCTION

CRIPPS R. J.

Abstract. Current surface construction methods in CADCAM use parametric polynomial equations in the form of a NURBS. This representation is ideal for computer-based implementations, allowing efficient interrogation. However, issues exist in constructing and manipulating such surfaces. When constructing a NURBS surface there are difficulties in determining constraints such as parameterisation, tangent magnitudes and twist vectors. Controlling the geometric features like curvature profiles of sectional/longitudinal curves on a NURBS surface is problematical as is joining several such surfaces together. A cause of these difficulties in control is that the control points do not lie on the surface itself. An alternative approach to surface construction is to specify the curvature and construct the surface so that it satisfies the curvature constraints. Since NURBS does not directly allow this, a fundamentally different approach is required. The key is to adopt a point-based approach where the surface is defined by a small number of points lying on the surface. Intermediate points are then constructed using a recursive approach which is defined to ensure that the curvature profile between adjacent points is of a very high quality. A case study is presented that illustrates the point-based approach.

Key words : Surface construction, Genralised Corn Spiral, NURBS, parameterisation.
AMS SUBJECT:16W60, 16W22, 16W20.

Generally, curves and surfaces can be represented mathematically either

- Explicitly, i.e. $y = f(x)$ or $z = f(x, y)$,
- Implicitly, i.e. $f(x, y) = 0$ or $f(x, y, z) = 0$,
- Parametrically, i.e. $< x = x(u), y = y(u), z = z(u) >$ or $< x = x(u, v), y = y(u, v), z = (z(u, v) >$.

Corresponding author. University of Birmingham, Edgbaston, B15 2TT, England. E-mail: r.cripps@bham.ac.uk.

Both the explicit and implicit forms are axis dependent making them unsuitable for use in computer graphics or CAGD. The parametric form is axis independent and can represent multi-valued curves and surfaces, are well defined for infinite derivatives and are relatively easy and quick to evaluate points and derivatives. Further more the parametric form is more flexible and therefore is ideally suited to computer-based applications. The use of the parametric representation for curves and surfaces is not without its problems. For example the derivative is now a function of the parameterisation which can leads to difficulties when trying to specify or control construction or modification of curves and surfaces. Further, some operations are more difficult in parametric form, for example finding the distance from a point to a surface. The aim is to construct high quality curves and surfaces that can be used in a wide range of applications including computer graphics and especially CADCAM applications where downstream activities are dependant on well-defined high quality surfaces. The term high quality relates to geometric quality, measured in terms of curvature. A high quality curve or surface has smooth curvature profiles with no unnecessary undulations and with continuous, to within acceptable tolerances, curvature across adjoining curves or surfaces.

A review of the basic ideas behind surface construction using parametric forms is given. Some of the inherent properties of the parametric form that affect the quality of geometry of the resulting surface are identified and current solutions are outlined. A fundamentally different approach to curve and surface construction is then given which is purely geometric in nature and which avoids the difficulties of the parametric form. Some of the basic algorithms are introduced followed by a discussion of how the point-based surface can be represented in parametric form thus making it available to existing CADCAM systems. A a case study illustrating the point-based approach to surface construction is then discussed and the article is concluded by summerising the advantages of the point-based approach and further work is identified.

Surface Construction in CAGD

Surface Representation. A Cartesian or tensor product parametric surface of degree (m, n) is given by:

$$\mathbf{S}(u,v) = \sum_{i=0}^{n} \sum_{j=0}^{m} B_{i,p}(u) B_{j,q}(v) \mathbf{P}_{i,j},$$

where $B_{i,p}(u)$ & $B_{j,q}(v)$ are the basis functions and $\mathbf{P}_{i,j} = \mathbf{V}_{i,j}/w_{i,j}, 0 \leq i \leq m, 0 \leq j \leq n$, are homogeneous representation of the Cartesian control points $\mathbf{V}_{i,j} =< x_{i,j}, y_{i,j}, z_{i,j} >$ and $w_{i,j} \geq 0$ are the weights associated with each vertex. The parameters u, v are usually normalised to $0 \leq u, v \leq 1$. The surface $\mathbf{S}(u_i, v_j)$ represents a mapping from u, v-space into object space $<$

$x, y, z >$. The nature and properties of the parametric surface are determined by the choice of the basis functions $B_{i,p}(u)$ & $B_{j,q}(v)$. The most common surface used in CAGD is the open clamped NUBS form. Here, the basis functions, $N_{i,p}(u)$ & $N_{j,q}(v)$ are defined recursively in terms of the knot vectors $[\mathbf{U}] = (u_0, \cdots, u_{m+p+1})$ & $[\mathbf{V}] = (v_0, \cdots, v_{n+q+1})$ with $u_p \leq u \leq u_{p+1}, v_q \leq v \leq v_{q+1}$ and knots at either end of $[\mathbf{U}]$ and $[\mathbf{V}]$ are repeated $p+1$ and $q+1$ times respectively. The weights are assigned a constant value, since in general values are difficult to specify. There use lies in reproducing known analytic curves and surfaces, for example conic sections, etc. The i^{th} basis function, $N_{i,p}(u)$, is givenby

$$N_{i,0}(u) = \begin{cases} 1 & \text{if} \quad u_i \leq u < u_{i+1} \\ 0 & \text{otherwise} \end{cases}$$

$$N_{i,p}(u) = \frac{u - u_i}{u_{i+1} - u_i} N_{i,p-1}(u) + \frac{u_{i+p+1} - u}{u_{i+p+1} - u_{i+1}} N_{i+1,p-1}(u) \quad 0 \leq i \leq m$$

noting that the quotient $0/0$ is defined as zero. The resulting NUBS surface consists of $(m+1-p) \times (n+1-q)$ sub-patches, which are C^{p-1} and C^{q-1} continuous in the u- and v-directions respectively, assuming there are no repeated interior knots. The open clamped knot vectors ensure that the surface passes through the four corner control points:$\mathbf{P}_{0,0}, \mathbf{P}_{m,0}, \mathbf{P}_{0,n}$ and $\mathbf{P}_{m,n}$, i.e. $\mathbf{S}(u_0, v_0) = \mathbf{P}_{0.0}$.

Since each of the basis functions are positive and sum to 1, the resulting surface point can be seen to be a convex linear combination of all the control points, the surface is guaranteed to lie inside the bounding box of the control points. This property is especially useful for computer graphics, visualisation and many interrogation algorithms. Evaluating the surface for a fixed value of u say and varying v results in a parametric curve lying on the surface. In particular the extremes of u, v result in four surface boundary curves, i.e. $\mathbf{S}(u, v_0) = \sum_{i=0}^{m} N_{i,p}(u)\mathbf{P}_{i,0}$ is a p^{th} degree NURBS curve. Thus, the surface can be thought of as a blend between four boundary curves (Fig. 1). It is noted that when the knot vectors consist of $0'$s and $1'$s then the basis functions are the Bernstein basis functions and the curve or surface is in Bézier form. This shows that the Bézier form is a true sub-set of NURBS.

Surface Interpolation. In order to construct a surface a mesh of surface points and possibly boundary derivatives is required. A surface can then be constructed that interpolates the given data. It is noted that this constrained interpolation is a linear system which is readily solved, whereas controlling curvature results in non-linear equations. Alternatively a more local approach can be taken where by boundary curves are constructed and then blended together to form the surface. This second approach is more problematic than the first since the resulting NUBS surface is not guaranteed to have any form

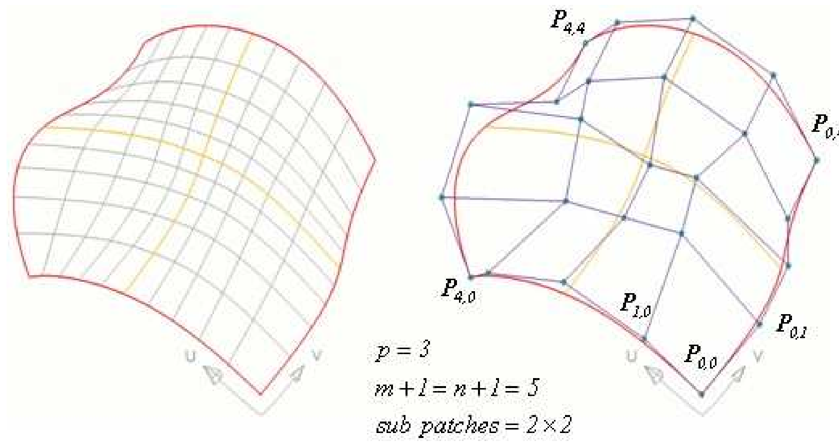


FIGURE 1. Open clamped B-spline surface with 2x2 sub-patches, and control polygon.

of continuity between sub-patches, and can result in multiple internal knots within the knot vector [1]. Adopting the global interpolation approach, there are several factors that affect the quality of the resulting surface.

Data specification: Number and spacing. Specifying geometric data is a realistic request. The number of data points directly corresponds to the number of control points required to represent the interpolated entity, so it is undesirable to use more points than necessary; doing so can lead to data proliferation and compromise downstream activities [2]. Also, sampling sparse data points from a poor quality object allows the interpolation process to 'smooth' over undesirable fluctuations in the geometry, whereas a dense point sampling will result in the blemishes being inherited. The aim is therefore to find the minimum number of data points that will satisfy the quality requirements. Closely related to the number of data points is the point spacing. Where curvature is constant (lines, planes, circles, cylinders), evenly spaced points yield optimal results [2]. This is not true of entities with varying curvature profiles. The method used to space points can have a significant effect on the quality of the interpolated entity; a good algorithm will characterise the geometry of the original shape and minimise positional/curvature errors.

Parameterisation. Parameterisation is the process that assigns a particular parameter value to each point on the surface. This can have a large effect on the quality of the resulting curve or surface as shown in Fig. 2. There are a number of parameterisation algorithms but the most common is chord length parameterisation.

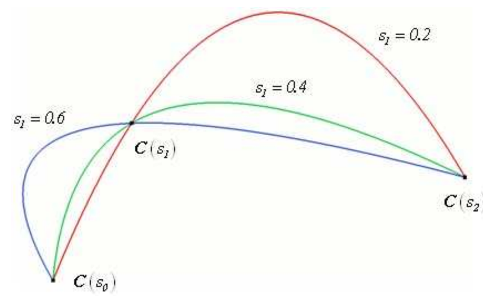


FIGURE 2. Effect of parameterisation on a quadratic curve.

Knot vector generation. The spacing of knots in the knot vector has a significant effect on the shape of a B-spline curve and [1] warn that a poor choice of knot vector can lead to a singular system of equations.

Derivative magnitude estimation. Greater control of the quality of the resulting surface can be gained by specifying boundary constraints. Boundary tangent directions are geometric and can be given. However, the interpolation process requires parametric derivatives which are more difficult to specify and therefore have to be estimated. Despite the magnitude being a function of the parameterisation, it is commonplace to approximate it using an estimate of the interpolated curves length, e.g. summing the chord lengths between data points.

Degree. The degree of a B-spline curve or surface dictates the degree of each span or sub-patch; the maximum possible degree is defined by the number of control points, i.e. for a curve $p \leq m$, and for a surface $p \leq m, q \leq n$. When $p = 1$, the curve is piecewise linear; it is clearly unsuitable for representing general freeform shapes. The quadratic curve, $p = 2$, has C^1 continuous spans. In general they do not provide sufficient flexibility for design: a single span cannot represent an inflection, nor can the derivative magnitudes at either end of the span be controlled independently. The cubic curve,$p = 3$, has C^2 continuous spans, and overcomes the problems encountered by the quadratic case. It is the most frequently used degree for interpolation [3]. The quartic curve,$p = 4$, has C^3 continuous spans; the additional continuity constraints of polynomials of degree $p > 4$ often force the curve to 'exaggerate' the features of the interpolated data, and undesirable oscillations may occur. When the flexibility of a higher degree curve is required, e.g. for refinement or manipulation of curvature without affecting tangents it is usual to interpolate with $p = 3$, and then raise to the desired degree (see [1]).

Reconciliation of parameter values. When interpolating surfaces, a single parameterisation has to be specified for each parametric direction. This requires a single parameterisation for both $\mathbf{S}(u,0)$ and $\mathbf{S}(u,1)$ boundaries. Except in the trivial cases, the parameterisation on these two boundaries is likely to be different. One method is simply to average the corresponding parameter values, however the parameterisation is a compromise for both boundaries, causing possible distortions and undulations. Unfortunately, some form of parameter reconciliation method is necessary, because global interpolation requires that parallel strips of points have identical parameterisations. The only way to avoid poor quality results is to ensure the reconciliation process causes minimal disruption to the ideal parameter values, i.e. by careful point distribution.

Twist vector estimation. Specifying data points and cross-boundary derivatives leaves four outstanding vector-valued pieces of information. These are the mixed second order partial derivatives, or twist vectors, from each corner of the interpolated surface,$\mathbf{S}_{uv}(a,b), a,b = (0,1)$. Since they are a function of the surfaces parameterisation, they have no geometric interpretation and must be estimated. In terms of their effect on the surface, they control the rate of change of the cross boundary derivatives, which must be identical in both parametric directions at the corners. Inappropriate control of the twist vectors can lead to poor surface geometry and/or poor parameterisation.

Point-based Surfaces

As indicated, when constructing a NURBS surface there are difficulties in determining constraints such as parameterisation, tangent magnitudes and twist vectors. Further, controlling geometric features like curvature is also problematical. An alternative approach is to specify the curvature and construct the surface to satisfy the curvature constraints. Since NURBS does not directly allow this, a fundamentally different approach is required. The key is to adopt a point-based approach where the surface is defined by a small number of points lying on the surface. Intermediate points are then constructed using a recursive approach which is defined to ensure that the curvature profile between adjacent points is of a very high quality.

The idea of using a discrete set of points to define the geometry of a surface was first introduced by [4]. in which the need for a closed-from mathematical model, NUBS, was questioned. The typical process in the CADCAM chain is to start with geometric data, usually in the form of discrete points. The next step is to fit a closed-form function, most likely NURBS, that approximates the data. This closed-form expression is then discretised in order to use it in follow-on processes for example interrogations, FE analysis, NC machining, etc. McLaughlin argued that the fitting stage was not always necessary. The fitting stage is not only time consuming but can actually lead to distortion of the original shape of the data. This phenomenon is unavoidable and is due to the nature of polynomial interpolation and approximation.

Rather than find a closed-form approximation, [4] suggested looking for a geometric approach to describing the shape by refining or filling-in data between original points. If this refinement can be controlled it should avoid the oscillatory problems associated with polynomials. Essentially the point-based approach requires the user to supply a grid of 3D design points which have been selected to characterise the intended design shape. It is assumed that the number of points and their physical location are sufficient for this purpose. This then defines the surface. Any intermediate point can be generated by using recursion.

Recursive Scheme. The recursive scheme is based on the GCS [5] which is a 2D curve that has a rational linear curvature profile, parameterised with respect to arc length, s, and is given by:

$$\kappa(s) = \frac{(\kappa_1 - \kappa_0 + r\kappa_1)\, s + \kappa_0 S}{rs + S} \quad , r > -1, \tag{1}$$

where κ_0 and κ_1 are the curvatures at the start and end of the curve segment, S is the total arc length and r is the shape factor which controls the fullness of the curve. It is easy to show that $\kappa(s)$ is a monotonic function wrt s since its first derivative is always greater than zero. Points on the GCS curve are given by:

$$x(S) = \int_0^S \cos\left(\frac{S}{r^2}\left[(1+r)(\kappa_0 - \kappa_1)\ln(1+\frac{rt}{S}) + r((1+r)\kappa_1 - \kappa_0)\frac{t}{S}\right]\right) dt$$

$$y(S) = \int_0^S \sin\left(\frac{S}{r^2}\left[(1+r)(\kappa_0 - \kappa_1)\ln(1+\frac{rt}{S}) + r((1+r)\kappa_1 - \kappa_0)\frac{t}{S}\right]\right) dt$$

These are Fresnel integrals and numerical evaluation has to be used. The end slope of the GCS segment is given directly by:

$$\theta_1 = \theta(S) = \begin{cases} \frac{S}{r^2}\left[(1+r)(\kappa_0 - \kappa_1)\ln(1+r) + r((1+r)\kappa_1 - \kappa_0)\right] & , r \neq 0 \\ \frac{S(\kappa_0+\kappa_1)}{2} & , r = 0 \end{cases}$$

where $\left(|\theta_1| \leq \frac{\pi}{2}\right)$.

It is noted that the GCS contains straight lines ($p = q = 0$), circular arcs ($q = r = 0$), logarithmic spirals ($q = 0; r \neq 0$) and Cornu spirals ($q \neq 0; r = 0$).

Thus, given a string of points the next stage is to construct GCS segments between adjacent pairs of points. To achieve this, arc lengths are estimated using circular arc interpolation from which accurate estimates of curvature vectors and hence tangent vectors can be found [6]. The GCS can match end points and tangents and approximate end curvatures. It is conjectured

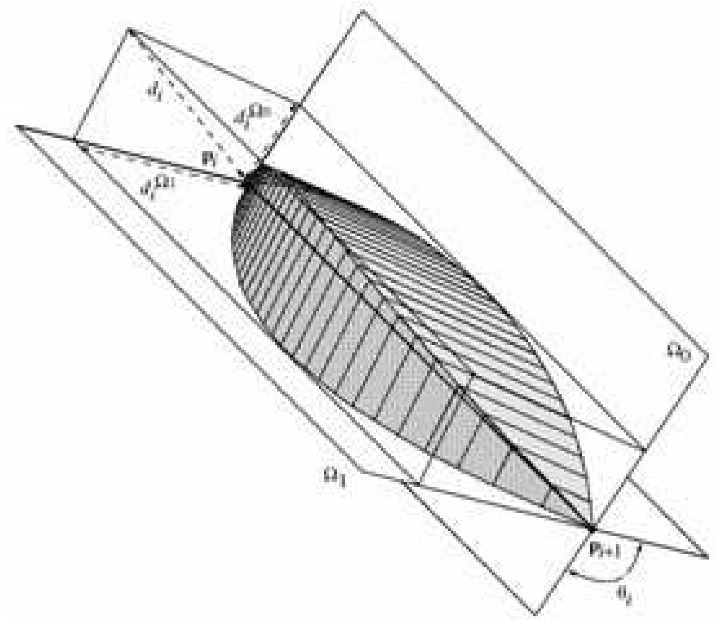


FIGURE 3. 3D GCS construction.

that the curvature discontinuities between adjacent GCS segments will be acceptable if the original string of data characterised the required geometry. Large discontinuities in curvature indicates that the data is either not dense enough or uncharacteristic of the intended shape. So far only planar GCSs have been constructed. Assuming both of the planar GCSs are single-valued with respect to the chord between the end points, $\mathbf{l}_i$, then the 3-D curve is well-defined and is the intersection curve between the two ruled surfaces constructed by extruding each of the planar GCSs parallel to their respective plane normals. Fig. 3 illustrates two planar GCSs, the corresponding ruled surfaces and the 3-D intersection curve on which the intermediate point, $\mathbf{q}_i$, lies.

The original design grid is now the 3D surface. Any intermediate points can be generated by the recursive scheme by treating the surface as strings of points in two directions. However, to remove any bias, refine using rows then columns followed by columns then rows and average the two refined grids.

NURBS Approximation

The high quality point-based surface, to be of any use to the CADCAM community, has to be able to used by existing software. This implies that it has to be converted to a NURBS representation. Clearly this cannot be

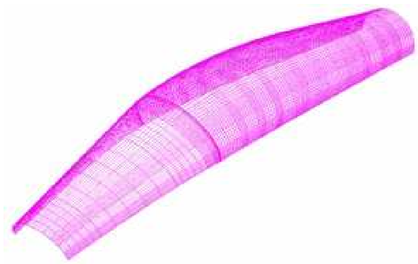


FIGURE 4. NURBS surface representation.

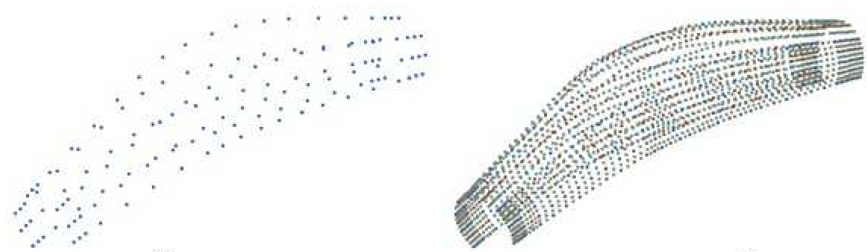
achieved exactly an dhas to be approximated. Since the geometry of the limit surface of the point-based definition is known, the selection of the parameters of the interpolation process can be guided by the information contained in the point-based data [7].

Since the initial mesh of points is small the number of sub-patches will also be small thus satisfying the assertion that fewer sub-patches result in fairer surfaces [8]. The tolerance to the point-based surface can readily checked by recursion of the data set and matching with a grid of parametrically generated points taken from the approximate NUBS surface. If the discrepancies are out of tolerance, the point-based surface can be further refined and a new approximation constructed. This clearly increases the number of sub-patches which is the usual trade-off between accuracy and data proliferation.

CASE STUDY

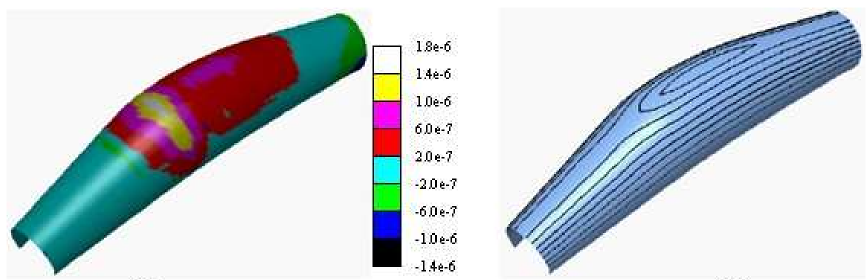
A canopy and windscreen was supplied by BAe Systems, UK with the aim of validating the point-based approach. The parametric definition, as shown in Fig. 4 was required to match the front optically defined windscreen to the free-form canopy whilst maintaining point and boundary tangent angle tolerances and controlling the blend between the two surfaces without compromising the optical surface. A minimal point set that characterised the underlying shape of the component was sampled from the parametric definition (Fig. 5(a)). The resulting point-based surface after two levels of refinement is shown in Fig. 5(b).

Gaussian curvature and isophotes were generated from the refined point-based data to assess the shape characteristics and are shown in Figs. 6(a) and (b) respectively. For the shaded Gaussian curvature, eight colours, spanning between $-1.4e^{-6}$ and $1.8e^{-6}$ with each colour spanning $4e^{-4}$ were used. For the isophote analysis a light direction of $[0, 0, 1]$ and isophote angles between 10^o and 90^o, spaced every 10^o were used. As can be seen, the Gaussian curvature shows no unnecessary undulations and has a smooth transition from the optic surface through the blend region into the free-form canopy. The isophotes



(a) (b)

FIGURE 5. Equivalent point-based surface(a) Intermediate points on the surface(b).



(a) (b)

FIGURE 6. Gaussian curvature map(a) Isophotes(b).

(Fig. 6(b)) gives further evidence of the high quality of the surface as the light lines show no unnatural behaviour indicating that the optical properties have been maintained. The NURBS approximation was constructed from the point-based surface and matched against the original parametric definition to ensure boundary tolerances were maintained.

CONCLUSIONS

A point-based surface construction method has been introduced that uses curves that possess a rational linear curvature profile. The resulting surfaces are of high quality in terms of the curvatures. The construction of a NUBS approximation to enable the point-based surface to be accessed by existing polynomial based software is guided by the knowledge of the geometry of the point-based definition. Outstanding research is to develop the methods to deal with irregular topologies and to consider an improved construction of the 3D GCS curve.

Acknowledgements

The authors are pleased to acknowledge the support from the Engineering and Physical Sciences Research Council (Grant reference number: GRK31268) and our collaborators Delcam plc and British Aerospace plc.

References

[1] Piegl, L. and Tiller, W. (1997), "The NURBS Book", Second Edition, Springer-Verlag, N.Y.
[2] Cripps, R.J. and Lockyer, P.S. (2005), Circle approximation for CADCAM using orthogonal C2 cubic B-splines. The International Journal of Machine Tools and Manufacture. 45,1222-1229.
[3] G. Farin. (2002) Curves and Surfaces for CADCAM. 2^{nd} ed. Academic Press, N.Y.
[4] H.W. McLaughlin. (1986), Describing the surface: algorithms for geometric modelling. *Computers in Mechanical Engineering*, November, 38–41.
[5] Ali, J.M., Tookey, R.M., Ball, J.V. and Ball, A.A. (1999), The generalised Cornu spiral and its application to span generation. Journal of Computational and Applied Mathematics, 102, 37-47.
[6] Tookey, R.M. and Ball, A.A. (1997), Estimation of curvatures from planar point data. In *The Mathematics of Surfaces VII*, eds. T.N.T. Goodman and R.R. Martin. Information Geometers, 131–144.
[7] Lockyer, P.S. (2006), Controlling the construction of NURBS curves and surfaces, Ph.D. thesis, University of Birmingham.
[8] Rogers, D.F. (2001) An introduction to NURBS with historical perspective. Morgan Kaufmann, N.Y.