Implicit Surfaces

The natural generalisation of the plane equation $f = [P - Q].N = 0$ is to move to more generic functions $f(P) = 0$. Whatever representations we use will define not only the zero-set surface, but also two **half-spaces**, $f(P) > 0$ and $f(P) < 0$.

Planes, Quadrics and Tori have relatively simple implicit equations, and there are also a few free-form formulations which have been used in animated movies.

An implicit surface cannot have edges, because it is the frontier between the positive and negative half spaces, and you cannot avoid crossing the frontier by going round the side.

It is possible, however, to have piecewise implicit surfaces defined as the frontiers of halfspaces built as boolean combinations of primitive halfspaces. The union and intersection of two halfspaces can be implemented numerically as the max and min of their functions. This gives composite functions whose boundary can have sharp corners. There are also composites called **R-functions** defined for example as

$$f = \frac{f_1 + f_2 + \sqrt{f_1^2 + f_2^2}}{2}$$

To define an implicit surface more complex than a torus requires care, as it is relatively easy to define equations whose zero-set has bubbles (unintended components not connected to the part you intended) and handles (pieces of surface bridging torus-like between other parts of the surface).

Parametric Surfaces

A parametric surface may be thought of as a distorted plane or piece of plane, because it is a map from $\mathbb{R}^2$ (or some subset of $\mathbb{R}^2$) into $E3$, the space of points.

Not all surfaces are definable in this way, but this formulation is very convenient for design.

The generic equation linear in the coefficients is

$$P(u,v) = \sum_i A_i f_i(u,v)$$

where again the A_i are the **coefficients** and the f_i are the **basis functions**.

The proof that all such surfaces are invariant under linear changes of coordinate system is exactly the same as in the curve case.

Construction of sets of basis functions with good predictable properties is less easy for this generality than for certain special cases. In CAGD we look at the following

- Tensor Products
- Bezier Triangles
- Box-Splines
- a few ad-hoc methods

Tensor Products

In this special case we structure the set of basis functions as a cartesian or tensor product of two sets of univariate functions.

The tensor product of two sets is defined at the set of pairs, one member of each pair being taken from the first set and one from the second.

In this case we have two sets of basis functions $\{f_i(u),\ i = 1 \ldots m\}$ and $\{g_j(v),\ j = 1 \ldots n\}$, and the bivariate functions so constructed are just

$$\{f_{ij}(u,v) = f_i(u)g_j(v),\ i = 1 \ldots m,\ j = 1 \ldots n\}$$

This implies a structuring of the coefficients into a grid, the elements of which we denote by A_{ij} and so the equation for the surface becomes

$$P(u,v) = \sum_{i=1}^{m} \sum_{j=1}^{n} A_{ij} f_i(u) g_j(v)$$

Theorem If $\sum_i f_i(u) = 1$ and $\sum_j g_j(v) = 1$ then $\sum_i \sum_j f_i(u) g_j(v) = 1$

The proof is achieved by merely taking $f_i(u)$ out of the inner summation and substituting the assumed properties.

■

Thus if we take univariate basis function sets which have all coefficients points, their tensor products will have the same property. The two sets do not need to be the same, and so it is possible to have a Bezier basis in one direction and a B-spline basis in the other. This is unusual, but a mixture of different degrees is frequently used. For example, a ruled surface can be defined by using first degree Bezier in one direction and some more complex basis in the other.

Theorem If $f_i(u) \geq 0 \ \forall i, \forall u$ and $g_j(v) \geq 0 \ \forall j, \forall v$ then $f_{ij}(u, v) = f_i(u) g_j(v) \geq 0$

The product of two non-negative values is always non-negative.

■

Theorem If both $f_i(u)$ and $g_j(v)$ are continuous in all derivatives up to the n^{th}, the product $f_i(u) g_j(v)$ is equally continuous.

This is set as an exercise. It is not quite as trivial as it looks, but almost.

■

The industry-standard surface definition uses the tensor product NonUniform Rational B-Spline basis (NURBS). NonUniform means that the pieces of polynomial are not all of the same length in parameter space: in particular they can degenerate to zero length with a nice controlled effect of reducing the continuity at a particular value of parameter by one level. Rational means that the definition is carried out in the Projective space in which points are represented by homogeneous coordinates.

Although CADCAM systems using this standard are more general than those using the equal interval non-rational formulation, the latter is a perfectly valid subset, and many specific surfaces are in fact designed with equal intervals and constant weights.

Bezier Triangles

If, instead of using the expansion of $(s + t)^n, \ s = 1 - t$ to give the univariate Bernstein basis functions over $s, t \geq 0$, we use that of $(u + v + w)^n, \ u + v + w = 1$, we find a logical triangle of functions over the triangle $u, v, w \geq 0$. Each of the functions has its maximum at a place within the triangle, and these maxima form a regular triangular subgrid.

The function with its maximum at $(i/n, j/n, k/n), \ i + j + k = n$ is

$$f_{ijk}(u, v, w) = \left(\frac{n!}{i!j!k!} \right) u^i v^j w^k$$

There are trivial proofs of the non-negativity of the basis functions within the triangle, and of the fact that they sum to 1.

We associate coefficients with the basis functions in the usual way, and regard them as being in the parameter plane at the maxima of the corresponding functions.

$$P(u,v,w) = \sum_{i,j,k\geq 0,\ i+j+k=1} A_{ijk} f_{ijk}(u,v,w)$$

The edges of the triangle, where u, v or w are equal to 0, turn out to be curves defined by the coefficients A_{0jk}, A_{i0k} and A_{ij0} on the corresponding edges of the triangle of coefficients, and, in particular, the vertices of the triangular patch lie at A_{00n}, A_{0n0} and A_{n00}.

The **directional derivative** of such a function is defined as the derivative taken along some specific direction in the u,v-plane.

Let a path be defined so that u,v and w are linear functions of t, say, and denote du/dt by u', etc. Note that $u'+v'+w'=0$ because $u+v+w=1$ identically.

Then the directional derivative df/dt of $f(u,v,w)$ is given by

$$u'\partial f/\partial u + v'\partial f/\partial v + w'\partial f/\partial w.$$

Each of the partial derivatives is itself a polynomial of degree one lower than the original triangle, and because u',v',w' are constants, so is the directional derivative. This polynomial can itself be expressed in terms of the trinomial Bernstein basis, and its coefficients are then given by the combinations $u'A_{(i+1)j(k-1)} + v'A_{i(j+1)(k-1)} + w'A_{ijk}$ taken over the triangles of the original set of coefficients.

This is effectively a hodograph property. We can determine from it that the derivative of any path crossing an edge depends only on the control points on that edge and those in the next row. For example, the derivative of a path crossing $u=0$ is not influenced by any coefficient A_{ijk} for which $i>1$. Similarly the second derivative is not influenced by any coefficient for which $i>2$.

We can use these derivative properties to determine the level of continuity, across the common edge, of two triangles adjacent in the parameter plane, and it is possible, though hard work, to build piecewise functions over regular triangular partitionings of the plane which have properties analogous to the univariate B-splines.

Box-Splines

There are constructions, called Box-Splines, which generate not only splines over regular triangular partitionings, but also over certain semi-regular grids, and which also contain the tensor product equal interval B-splines within the same theoretical framework.

These are not actually used in any commercial CADCAM systems. They are, however, extremely important theoretically, especially in the subdivision context, and they will be described in the subdivision notes.

There also exist constructions called Simplex-Splines, which are the most fundamental of all, and which can give splines over very irregular partitionings, but the partitioning has to come out of the construction: it cannot be pre-imposed. This has meant that these splines have proved unusable for any practical purposes, despite their incredible theoretical elegance and fundamentality. If we get through the material of the course early I will tell you about them at the end.

Some ad hoc methods

There are a number of ways in which slope continuous surfaces can be fitted to data not arranged in an exploitable pattern. These were mainly devised in the context of the contouring of scattered data, where points are defined in the plane with some data values attached, and where the positioning of the points in the plane has no exploitable structure. They are not used in CAGD, primarily because of the practical problem of deciding what positions in the parameter plane to give to points of which we only know the coordinates.

$z = f(x, y)$ over an irregular triangulation

An irregular triangulation can be fitted by putting piecewise polynomials into each of the given triangles. The three cases which have not been subsumed by Box-Spline theory are

- Clough-Tocher

Each triangle is divided into three subtriangles by joining its vertices to some interior point (typically the centroid, but this is arbitrary), and then a cubic piece is fitted into each sub-triangle, matching the value and first derivatives at each original corner and also derivatives across the middles of the edges.

- Powell-Sabin 6

Each triangle is divided into six subtriangles by joining a point in the interior of each triangle to the vertices of its own triangle and also to the interior points of the adjacent triangles. The incentre is usually chosen, because this choice ensures that the line joining two interior points actually cuts the edge where the triangles meet. Each subtriangle then has a quadratic piece fitted, matching the value and first derivatives at each original corner. In the regular grid case, this turns out to be equivalent to a certain Simplex Spline, but nobody has been able to extend this result in any useful or meaningful way.

- Powell-Sabin 12

Each triangle is divided into twelve subtriangles by lines joining the centroid to the vertices and to the three midpoints of the edges. These midpoints are also joined to each other. The subtriangles are then filled with quadratic pieces chosen to match the same conditions as the Clough-Tocher split described above.

Non-polynomial surfaces

There are two main classes of bivariate basis functions outside the polynomials and the piecewise polynomials. We can still express them in the form

$$z(x,y) = \sum_i z_i f_i(x,y)$$

and the $f_i(x,y)$ depend on the x,y positions of the data points, but not on the ordinates z_i.

- Weighted Means

$$f_i(x,y) = r_i^{-k} / \sum_i r_i^{-k}$$

where $r_i = (x-x_i)^2 + (y-y_i)^2$

Positivity and summation to unity are trivial to prove. This is in fact an interpolating basis because if $(x,y) \to (x_i,y_i)$ then $r_i \to 0$ and any negative power tends to infinity, dominating all other terms in the sum. We then have the ratio of two infinite quantities, but any limiting process leads to the same limit, =1.

The value of k is an ad hoc choice.

- Radial Basis Functions

The weighted means have a problem that the value cannot anywhere go outside the range of the data values, and that if k is taken high enough to ensure the definition of a first derivative at each data point, that first derivative turns out to be zero. There are thus flat spots.

A better scheme uses

$$z(x,y) = \sum_i A_i f(r_i)$$

where r_i is again the distance from the evaluation point to the i^{th} data point. The A_i are computed from the z_i by solving the equation system

$$z_j = z(x_j,y_j) = \sum_i A_i f(r_i(x_j,y_j))$$

There are many variants of this scheme, and there is a major industry going on at the moment within academic Approximation Theory, exploring them. For example, for any proposed choice of $f(.)$ it is necessary to prove that the equations for the A_i are non-singular, and if possible well-conditioned. An important category of schemes are those for which $f(r)$ has finite support (because the equation system then becomes sparse, with important resulting performance improvement), and it is then interesting to determine the continuity of the resulting surfaces.