CAD: Curves and Surfaces: Part I of III

We saw earlier how CSG or BREP can be used as schemes to represent solids in a computer. Using either a CSG or a BREP model, one can, potentially answer any question about the solid that is represented. Some of these questions could be:

• What is the weight of the solid ?
• If one is to view the solid from a given direction, what would that view be like (that is, which edges will be visible, and which ones will be hidden) ?
As an extension of this question, consider this: If we define the location of a set of sources of light, and their intensity, and the optical properties (such as reflectivity, color etc.) of the solid, what will the solid look like ? Such images are normally drawn by most off-the-shelf solid modeling packages, and are called **rendered** images. The method of figuring out what this image is like is called **rendering**.
• What are the coordinates of the center of mass of the solid ?
• What is its moment of inertia about a given axis ?

There are many other questions that arise when performing analysis of solids, depending upon the application, some of which we shall see later on.

Problem: In CSG, the solid is built using a set of primitives. There may be some shapes that can not be built using any combination of a given set of primitives. An example of this is shapes that are known as sculptured (or free-form) shapes. Free-form shapes, as far as we are concerned, represents shapes that are not linear/planar, or based on conic sections.

Many objects in life have shapes that can not be described by operations on simple primitives. This means that more complex mathematical representations are needed to describe such shapes. When a CSG representation is built, the user specifies the values of the parameters for the primitives. For instance, changing the length, width or height of a block; changing the diameter or height of a cylinder etc. In all these operations, the basic *shape* of the primitive remains the same, in some sense.. for instance, a block is scaled differently along the X-Y-Z axes, but it is still a rectangular parallelepiped.

However, in case of sculptured surfaces, each object we study has a different basic shape. How then can we specify a primitive shape for it in a CSG system ?

This problem was studied in depth in the early days of CAD due to its importance to the automobile (and later, aircraft) industry. Basically, we still need a "nice" mathematical representation of a surface in order to use it in a computerized system. In terms of "nice", one of the best behaved set of functions (and also most studied) are polynomials. (Why ?)

Review of vector algebra

A point in 3-D Euclidean space can be represented by the linear sum of its coordinates along three mutually independent vectors (called the basis). The most common basis is three mutually orthogonal unit vectors denoted (**i**, **j**, **k**).

Thus a point, P, is denoted by its position vector:
P = p_x**i** + p_y**j** + p_z**k**; p_x, p_y, p_z are the components of **P** along the basis vectors.

The norm of **P**, denoted |**P**|, is the Euclidean distance of **P** from the origin;

$|\mathbf{P}|^2 = p_x^2 + p_y^2 + p_z^2$.

The scalar product of two vectors is given by:

$\mathbf{P} \bullet \mathbf{Q} = p_x q_x + p_y q_y + p_z q_z = |\mathbf{P}|\,|\mathbf{Q}| \cos\theta$;   θ is the angle between **P** and **Q**.

What is the physical significance of the dot product ?

The cross product of two vectors is given by:

$$\mathbf{P} \times \mathbf{Q} = \begin{vmatrix} i & j & k \\ p_x & p_y & p_z \\ q_x & q_y & q_z \end{vmatrix}$$

= |**P**| |**Q**| sinθ **u**; where **u** is a unit vector perpendicular to ***P*** and ***Q***, and oriented by the right-hand rule, measuring θ from ***P*** to ***Q***.

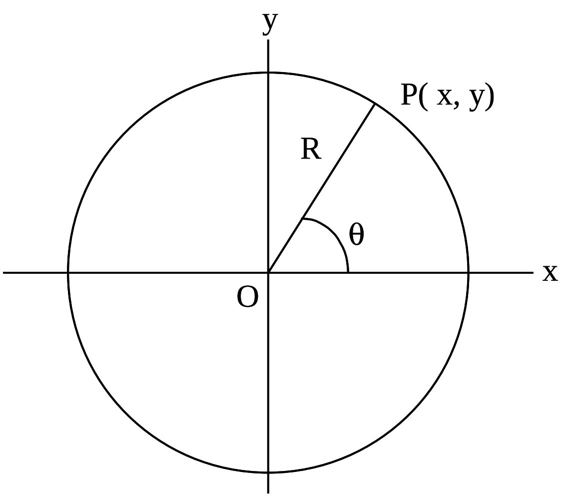
What is the physical significance of the cross product ?

Curve Geometry:

There are three ways of representing a curve:

- an implicit curve;
- an explicit (or non-parametric) curve;
- a parametric curve;

Consider a circle as shown below:



The circle can be represented as: $x^2 + y^2 = R^2$.

This is called the implicit equation of the circle, since the value of x is not directly represented as a function of y (or vice versa). The general form of an implicit equation is: $g(x, y, z) = 0$.

We can rewrite the circle equation as:

$$y = (R^2 - x^2)^{1/2}$$

Does the function $y(x)$ represent the circle ?

The form $y = y(x)$ is said to be in explicit form.

Equivalently, we can represent the circle in the following way:

$$P = (p_x, p_y) = (R\cos\theta, R\sin\theta).$$

In this case, θ is a parameter of the curve, and any point on the circle can be uniquely identified by the corresponding value of θ.

Consider the following re-paramtrization:

$\phi = \theta/2$

$p_x = R\cos\theta = R \cos(2 \phi) = R(\cos^2 \phi - \sin^2 \phi) = R \cos^2 \phi (1 - \tan^2 \phi) = R(1 - \tan^2 \phi) / \sec^2 \phi = R (1 - \tan^2 \phi) / (1 + \tan^2 \phi)$

and,

$p_y = R\sin\theta = R \sin(2 \phi) = 2R \sin \phi \cos \phi = 2R \tan \phi \cos^2 \phi = 2R \tan \phi / \sec^2 \phi = 2R \tan \phi /(1 + \tan^2 \phi)$

Substituting $t = \tan \phi$, we get:

$p_x = R (1 - t^2) / (1 + t^2)$;

$p_y = R 2t / (1 + t^2)$

The above is a parametric equation of the circle, and is said to be in ***rational polynomial form***, since each equation is a ratio of polynomial functions of the parameter t.

The best way to represent space curves (or surfaces) for use in computational geometry are parametric forms, using equations of the form:

$x = x(t), y = y(t), z = z(t)$.

Or, if we represent the curve as the locus of the position vector R, the parametric equation is:

$\mathbf{r}(t) = (x(t), y(t), z(t))$.

Remember that a curve may be planar, or 3-Dimensional. For example, what does the following curve represent:

$\mathbf{r}(t) = a \cos t\, \mathbf{i} + a \sin t\, \mathbf{j} + ct\, \mathbf{k}$?

Some properties of curves:

Arc Length:

We have seen that one parameter, *t*, varies in its value over real interval [a,b], with $\mathbf{r}(a)$ being the starting point of the curve, and $\mathbf{r}(b)$ the end point. It is therefore possible to describe a curve by a parameter, s, in the range [*0, L*] where *L* is the length of the curve, and the parameter value at any point equals the length of the curve from the start point to ***r****(s)*. For most curves, it is not easy (or convenient to find such a parametric form). We will distinguish the arc-length parameter (s) from all other types of parameters (t) that we shall use.

For curve ***r****(s)*, arc length between points ***r****(a)* and ***r****(b)* = | *b – a* |.

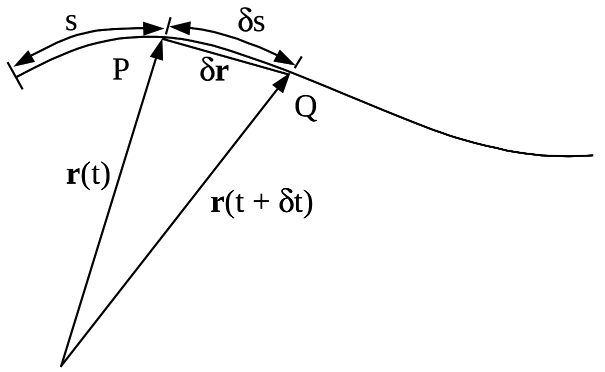
For a curve ***r****(t)*, the arc length between two points corresponding to t = a and t = b is given by:

$$s = \int_a^b \sqrt{\dot{r}.\dot{r}}\, dt \qquad where\ \dot{r} = dr/dt$$

If we express the arc length from t = a to an arbitrary point, t, then we get the following important form:

arc length = $s(t) = \int_a^t \sqrt{\dot{r}.\dot{r}}\, dt$

Tangent:



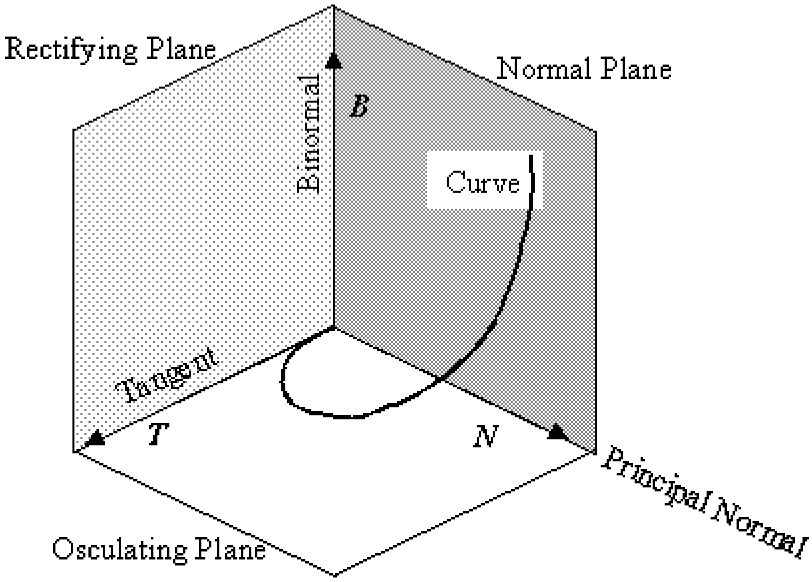
A curve, r(t) is shown in the figure above. As δt → 0, the vector *d**r**/dt* will become parallel to the tangent. Thus the unit tangent vector, ***T*** = *(d**r**/dt) /* | *d**r**/dt* |. In particular, since *s* is the arc length, as ds → 0, the chord *δr* will approach the length of the arc, δs, and | *dr/ds* | → 1. Thus, if the curve is represented in terms of the arc length as ***r****(s)*, the unit tangent vector is given by *d**r**/ds*.

The vector d**r**/dt is the tangent to the arc **r**(t).
The unit tangent vector, **T** = (d**r**/dt) / | d**r**/dt | is the unit tangent vector to the curve.

[**Notations**: We will use bold-face to represent vectors, normal font to represent scalars, a dot above a function to show its derivative with respect to *t*, and the prime, or apostrophe sign, to denote its derivative with respect to *s*.]

T = **r'** = d**r** / ds.

What is the parametric equation of the tangent to a curve ?



The figure above shows a general 3D curve. Notice that the curve is characterized at any point by two things: how much it is curving, and how much it is twisting. For such 3D curves, the plane perpendicular to T is a normal plane, and any vector in the normal plane is a normal vector. The vector *d**T**/dt* lies in the Normal plane (since it is perpendicular to ***T***) The unit vector parallel to *d**T**/dt* is called the **principal normal, *N***. In particular, if r = r(s), then |dT/ds | indicates the magnitude by which the tangent vector is "curving" – in other words:

$$d\mathbf{T}/ds = \kappa\mathbf{N}$$

Where κ is the curvature of the curve, and the ***radius of curvature*** of the curve is defined: $\rho = 1 / \kappa$.

It can be shown that $curvature = \kappa(s) = |u'(s)| = |r''(s)| = \frac{|\dot{r} x \ddot{r}|}{|\dot{r}|^3}$

A vector **B**, called the unit ***binormal***, perpendicular to ***T*** and ***N***, forms a right-handed system as follows: **B** = **T** x **N**.

Surfaces:

All physical objects are bound by surfaces. Surfaces can be described mathematically in implicit, explicit (non-parametric) or parametric form. These representations are extensions of definitions of curves.

Implicit forms:

Consider a sphere of radius R, centered around point (r_x, r_y, r_z).

What is the representation of all point on the surface of the sphere ?

$(x - r_x)^2 + (y - r_y)^2 + (z - r_z)^2 - R^2 = 0.$

What is the representation of all points on or inside this sphere ?

Non-parametric forms:

Rearranging the sphere equation from above, we get the corresponding explicit form:

$z = [R^2 - (x - r_x)^2 - (y - r_y)^2]^{1/2} + z_x.$

Parametric forms:

The parametric representation of a curve was a set of functions, *x(t), y(t)* and *z(t),* in terms of one parameter, *t.*

Surfaces are represented similarly, using two parameters, *u* and *v*. Any position vector on the surface, ***r****(u, v),* is represented:

r*(u, v) = (x(u, v), y(u, v), z(u, v)).*

For instance, a unit sphere, centered on the origin, can be represented in terms of parameters u and v as:

r(u, v) = (cos v cos u, cos v sin u, sin v), with $0 \le u \le 2\pi$, $0 \le v \le \pi/2$.

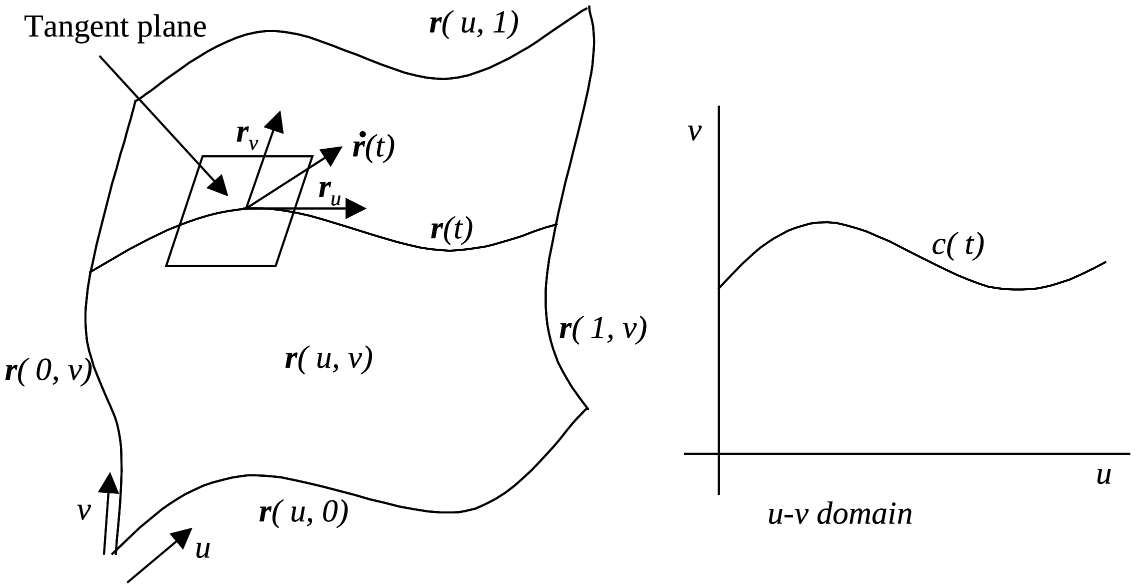
A rational parametric form of the unit sphere can be derived to be:

$x(u, v) = (1 - u^2)(1 - v^2) / [(1 + u^2)(1 + v^2)]$,

$y(u, v) = 2u(1 - v^2) / [(1 + u^2)(1 + v^2)]$,

$z(u, v) = 2v(1 + u^2) / [(1 + u^2)(1 + v^2)]$.

Tangent and Normal of Surfaces:



For a surface defined by ***r***(u, v), consider a curve, ***c***(t), lying on the surface. Thus, ***c***(t) = (u(t), v(t)). Since ***c***(t) lies on the surface ***r***(u, v), therefore ***c***(t) forms a curve, ***r***(t), on the surface.

r(t) = ***r***(u(t), v(t)) = (x(u(t), v(t)), y(u(t), v(t)), z(u(t), v(t))).

Of course, a tangent to the curve **r** will be a tangent to the surface.

Differentiating the above expression,

$$\dot{r} = \frac{dr}{dt} = \frac{\partial r}{\partial u}\frac{du}{dt} + \frac{\partial r}{\partial v}\frac{dv}{dt} = r_u \dot{u} + r_v \dot{v}$$

Where d**r**/dt is the tangent vector of the curve ***r(t)***; $\mathbf{r}_u$ and $\mathbf{r}_v$ lie in the tangent plane, and are the tangents to ***isoparametric curves*** (curves with one of the parameters, *u* or *v*, remaining constant). The normal vector lies along the cross product of $\mathbf{r}_u$ and $\mathbf{r}_v$.

Unit normal vector = $\boldsymbol{n} = (\mathbf{r}_u \times \mathbf{r}_v) / | \mathbf{r}_u \times \mathbf{r}_v |$

The normal vector is extremely important in machining, since it is used in computing the ***offset surface*** for **r**(u, v). For instance, an offset surface with offset *d*, is given by:

$$\mathbf{r}^{\circ} (u, v) = \mathbf{r} (u, v) + d \ \mathbf{n} (u, v).$$

Remember from your engineering maths that the gradient of an implicit equation of a surface defines the normal to the surface *f(x, y, z) = 0*:

$$grad\ f = \frac{\partial f}{\partial x} i + \frac{\partial f}{\partial y} j + \frac{\partial f}{\partial z} k$$

The gradient has an important property that it is invariant with respect to the coordinate frame. It also has the property that it lies along the normal to the surface. This can be proved as follows. Consider a scalar representation of a surface, *f(x , y, z) = 0*, and a curve, ***r***(t) = x(t) ***i*** + y(t) ***j*** + z(t) ***k***, that lies on this surface. Since all points on this curve must satisfy the equation of the surface, we have: *f(x(t), y(t), z(t)) = 0*. differentiating with respect to the curve paramater *t*, and using the chain rule, we get:

$$\frac{\partial f}{\partial x}\dot{x} + \frac{\partial f}{\partial y}\dot{y} + \frac{\partial f}{\partial z}\dot{z} = (grad\ f).\dot{r} = 0$$

This indicates that (grad f) is perpendicular to the direction of the tangent – that is, it lies along the normal to the surface. This is an easy way to compute the normal of surfaces described by an implicit equation. However, most of the time, we will deal with surfaces that are defined in a parametric form.

Curvature of the surface:

Another property of a surface we often need to compute is its curvature. For example, if we need to machine a surface, and our tool, say a ball-end mill, touches the surface at some point. If the radius of the surface (along any direction) is less than the radius of the tool, then the tool will cut into the part. This situation, called gouging, is undesirable. Thus, if we have a surface that needs to be machined, we must first compute, for each point on it, the radius of curvature. The minimum value for the radius (at the point on the surface that is ‘most curved’) indicates the *largest tool* that can machine the entire surface without gouging. There are two problems:

(1) The surface has an infinite number of points. This problem is handled by discretization of the surface, and only considering a mesh of points that approximate the surface closely.

(2) At each point, there are an infinite number of directions, and a (possibly) different curvature along each direction. Fortunately, it can be shown that at each point, we can compute the principle directions, and the curvatures along the two orthogonal principal directions are the local extrema – one principal curvature is the minimum, and the other the maximum. Hence, one only needs to compute the values of the principal curvatures at each point.

Consider again, a surface ***r****(u,v)* with a curve ***r****(t)* lying on it – denoted ***r****(u(t), v(t)).* We shall use a 2x1 matrix, ***u****(t)* = ***u*** = *[u(t), v(t)]*^{T} to denote the functions of the two parameters.

Let us look at the curve r(t) lying on our surface. Using the chain rule for differentiation, the tangent vector of this curve is obtained as:

$$\dot{r} = \frac{\partial r}{\partial u}\dot{u} + \frac{\partial r}{\partial v}\dot{v} \quad = \quad A\dot{u}, \quad where \quad A = \begin{bmatrix} \partial x/\partial u & \partial x/\partial v \\ \partial y/\partial u & \partial y/\partial v \\ \partial z/\partial u & \partial z/\partial v \end{bmatrix}$$

The length of the tangent is calculated from:

$$|\dot{r}|^2 = \dot{r}^T\dot{r} = \dot{u}^T A^T A\dot{u} \qquad we\,denote\ A^T A \quad = G = \begin{bmatrix} \frac{\partial r}{\partial u}\cdot\frac{\partial r}{\partial u} & \frac{\partial r}{\partial u}\cdot\frac{\partial r}{\partial v} \\ \frac{\partial r}{\partial v}\cdot\frac{\partial r}{\partial u} & \frac{\partial r}{\partial v}\cdot\frac{\partial r}{\partial v} \end{bmatrix}$$

G is called the ***first fundamental matrix*** of the surface, and the unit tangent vector along the curve ***u****(t)* on the surface is given by:

$$T = \frac{\dot{r}}{|\dot{r}|} = \frac{A\dot{u}}{(\dot{u}^T G\dot{u})^{1/2}}$$

For a general curve, u(t) lying on a surface, using the definitions in the previous section and the chain rule, we see that:

$$\frac{du}{dt} = \frac{du}{ds}\frac{ds}{dt} = \dot{s}T; \quad and \quad \ddot{r} = \frac{d(\dot{s}T)}{dt} = \ddot{s}T + \dot{s}\frac{dT}{dt} = \ddot{s}T + \dot{s}\frac{dT}{ds}\frac{ds}{dt} = \ddot{s}T + \dot{s}^2\frac{dT}{ds} = \ddot{s}T + \dot{s}^2\kappa N$$

And using the equation for tangent derived earlier, and the chain rule, we get the following:

$$\frac{d\dot{r}}{dt} = \frac{\partial r}{\partial u}\ddot{u} + \dot{u}\frac{d(\partial r/\partial u)}{dt} + \frac{\partial r}{\partial v}\ddot{v} + \dot{v}\frac{d(\partial r/\partial v)}{dt} = \frac{\partial^2 r}{\partial u^2}\dot{u}^2 + 2\frac{\partial^2 r}{\partial u \partial v}\dot{u}\dot{v} + \frac{\partial^2 r}{\partial v^2}\dot{v}^2 + \frac{\partial r}{\partial u}\ddot{u} + \frac{\partial r}{\partial v}\ddot{v}$$

From the above two expressions, taking the dot product with a unit vector, ***n***, normal to the surface, we get the expression for the curvature (note: ***n*** is perpendicular to ***T*** and to the partial derivatives of ***r*** with respect to *u* and *v*):

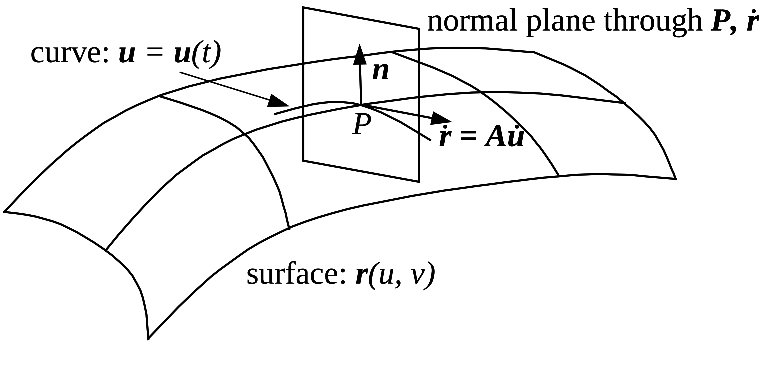
$$\dot{s}^2 \kappa N \cdot n = n \cdot \frac{\partial^2 r}{\partial u^2}\dot{u}^2 + 2n \cdot \frac{\partial^2 r}{\partial u \partial v}\dot{u}\dot{v} + n \cdot \frac{\partial^2 r}{\partial v^2}\dot{v}^2$$

Using the matrix notation, we can write this as:

$$\dot{s}^2 \kappa N \cdot n = \dot{u}^T D \dot{u}, \quad where \quad D = \begin{bmatrix} n \cdot \frac{\partial^2 r}{\partial u^2} & n \cdot \frac{\partial^2 r}{\partial u \partial v} \\ n \cdot \frac{\partial^2 r}{\partial u \partial v} & n \cdot \frac{\partial^2 r}{\partial v^2} \end{bmatrix}$$

D is called the ***second fundamental matrix*** of the surface. Note that ***N*** is a unit normal to the curve (on the surface), while ***n*** is a unit normal to the surface. In general, these two vectors may not be parallel. You can imagine this by thinking of a point ***P*** on a surface, and imagining several curves passing through this point. The normal vectors to these curves at ***P*** may not all be parallel. Imagine a curve such that its normal at ***P*** is parallel to the surface normal. The curvature of the surface along this curve, at point ***P***, is called the normal curvature and denoted κ_n. In this case, since $N.n = 1$, we get:

$$\kappa_n = \frac{\dot{u}^T D \dot{u}}{\dot{s}^2} = \frac{\dot{u}^T D \dot{u}}{\dot{u}^T G \dot{u}}$$



As we rotate the normal plane in the above figure through the point ***P***, the value of κ_n changes. The minimum and maximum values of the curvature can therefore be derived by differentiating the expression of κ_n with respect to the tangent direction [du/dt, dv/dt].

These directions are called the ***principal directions***, and the corresponding values of curvature are called the ***principal curvatures***. The values of the principle curvatures are given by the following expression:

$$\kappa_n = \frac{(g_{11}d_{22} + d_{11}g_{22} - 2g_{12}d_{12}) \pm \sqrt{(g_{11}d_{22} + d_{11}g_{22} - 2g_{12}d_{12})^2 - 4|G||D|}}{2|G|}$$

where g_{ij} and d_{ij} are the *(i,j)* elements of the *G* and *D* matrices, and |*G*|, |*D*| are the determinants of the *G* and *D* matrices respectively.

We need surfaces for representing the geometry of useful parts. These surfaces are complex by nature. What must we be able to do with our surface equations ?
(a) We must be able to generate the Cartesian coordinates of the points on the surface.
(b) We must be able to generate the equations of tangent planes, and normal vectors at any point on the surface
(c) We must be able to find the curvature (if it is defined) at any point on the surface.
(d) We must be able to approximate the surface by a grid of small (planar) patches. In particular, we must be able to approximate the entire surface by a set of small, well-behaved triangles. Likewise, we must be able to convert a solid contained by one or more surfaces by a set of small, well-behaved tetrahedrons.

Normals and curvatures have practical applications in generating the machining plans for surfaces. Triangulation of surfaces is useful in creating a planar representation called STL, which is commonly used for layered manufacturing. Tetrahedral tessellation is useful for finite element analysis.

In the next section, we study the following problem: among the different methods of representing curves and surfaces (implicit, explicit, parametric), which format is best for what we need to do ?