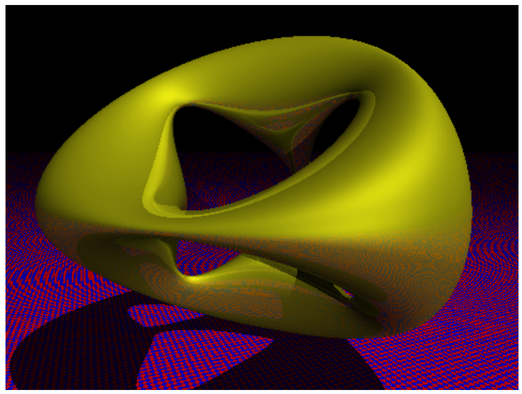
4 Surfaces in $\mathbf{R}^3$

4.1 Definitions

At this point we return to surfaces embedded in Euclidean space, and consider the differential geometry of these:



We shall not forget the idea of an abstract surface though, and as we meet objects which we call *intrinsic* we shall show how to define them on a surface which is not sitting in $\mathbf{R}^3$. These remarks are printed in a smaller typeface.

Definition 11 *A smooth surface in $\mathbf{R}^3$ is a subset $X \subset \mathbf{R}^3$ such that each point has a neighbourhood $U \subset X$ and a map $\mathbf{r} : V \to \mathbf{R}^3$ from an open set $V \subseteq \mathbf{R}^2$ such that*

- $\mathbf{r} : V \to U$ *is a homeomorphism*
- $\mathbf{r}(u, v) = (x(u, v), y(u, v), z(u, v))$ *has derivatives of all orders*
- *at each point $\mathbf{r}_u = \partial \mathbf{r}/\partial u$ and $\mathbf{r}_v = \partial \mathbf{r}/\partial v$ are linearly independent.*

Already in the definition we see that X is a topological surface as in Definition 2, since $\mathbf{r}$ defines a homeomorphism $\varphi_U : U \to V$. The last two conditions make sense if we use the *implicit function theorem* (see Appendix 1). This tells us that a local invertible change of variables in $\mathbf{R}^3$ "straightens out" the surface: it can be locally defined by $x_3 = 0$ where (x_1, x_2, x_3) are (nonlinear) local coordinates on $\mathbf{R}^3$. For any two open sets U, U', we get a smooth invertible map from an open set of $\mathbf{R}^3$ to another which takes $x_3 = 0$ to $x'_3 = 0$. This means that each map $\varphi_{U'}\varphi_U^{-1}$ is a smooth invertible homeomorphism. This motivates the definition of an abstract smooth surface:

Definition 12 *A smooth surface is a surface with a class of homeomorphisms φ_U such that each map $\varphi_{U'}\varphi_U^{-1}$ is a smoothly invertible homeomorphism.*

Clearly, since a holomorphic function has partial derivatives of all orders in x, y, a Riemann surface is an example of an abstract smooth surface. Similarly, we have

Definition 13 *A smooth map between smooth surfaces X and Y is a continuous map $f : X \to Y$ such that for each smooth coordinate system φ_U on U containing x on X and ψ_W defined in a neighbourhood of $f(x)$ on Y, the composition*

$$\psi_W \circ f \circ \varphi_U^{-1}$$

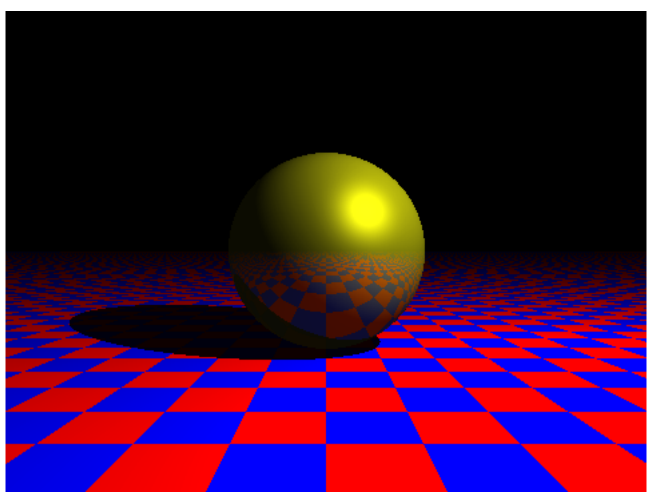
is smooth.

We now return to surfaces in $\mathbf{R}^3$:

Examples:

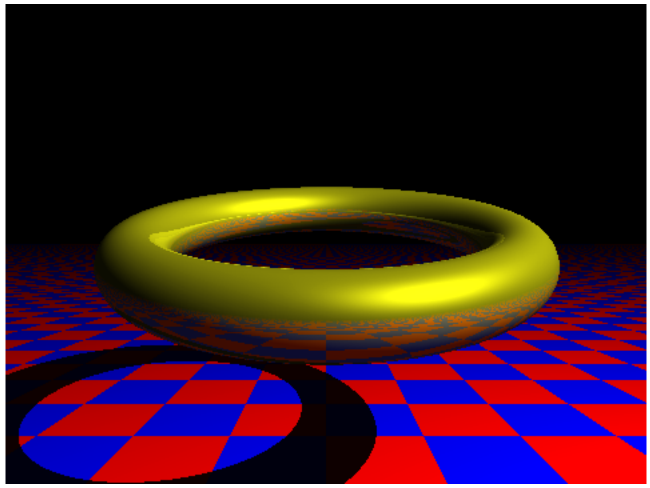
1) A sphere:

$$\mathbf{r}(u, v) = a \sin u \sin v \,\mathbf{i} + a \cos u \sin v \,\mathbf{j} + a \cos v \,\mathbf{k}$$



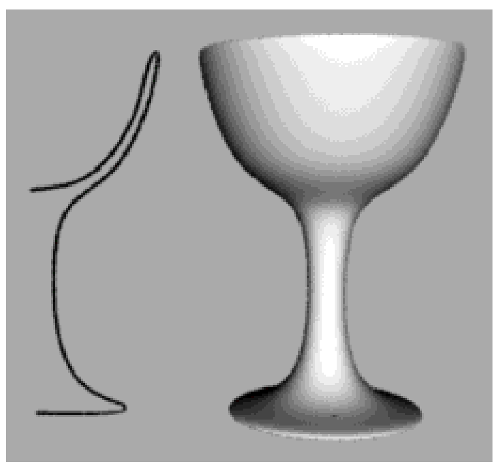
2) A torus:

$$\mathbf{r}(u, v) = (a + b \cos u)(\cos v \,\mathbf{i} + \sin v \,\mathbf{j}) + b \sin u \,\mathbf{k}$$



3) A surface of revolution:

$$\mathbf{r}(u,v) = f(u)(\cos v\,\mathbf{i} + \sin v\,\mathbf{j}) + u\mathbf{k}$$



These are the only compact surfaces it is easy to write down, but the following non-compact ones are good for local discussions:

Examples:

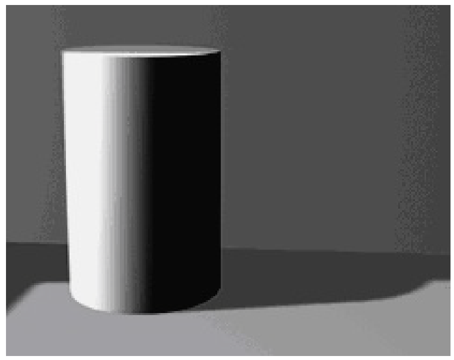
1) A plane:

$$\mathbf{r}(u,v) = \mathbf{a} + u\mathbf{b} + v\mathbf{c}$$

for constant vectors $\mathbf{a}, \mathbf{b}, \mathbf{c}$ where $\mathbf{b}, \mathbf{c}$ are linearly independent.

2) A cylinder:

$$\mathbf{r}(u,v) = a(\cos v\,\mathbf{i} + \sin v\,\mathbf{j}) + u\mathbf{k}$$

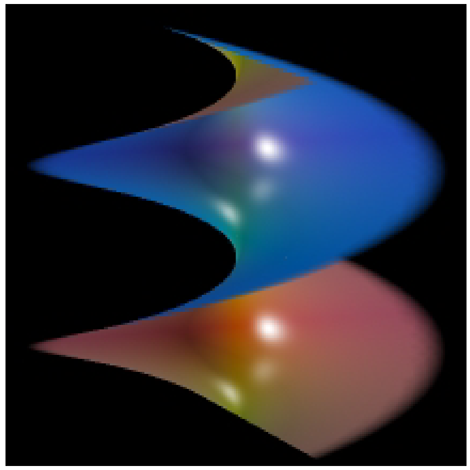


3) A cone:

$$\mathbf{r}(u,v) = au\cos v\,\mathbf{i} + au\sin v\,\mathbf{j} + u\mathbf{k}$$

4) A helicoid:

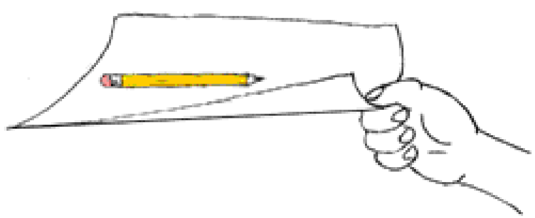
$$\mathbf{r}(u,v) = au\cos v\,\mathbf{i} + au\sin v\,\mathbf{j} + v\mathbf{k}$$



5) A developable surface: take a curve $\boldsymbol{\gamma}(u)$ parametrized by arc length and set

$$\mathbf{r}(u,v) = \boldsymbol{\gamma}(u) + v\boldsymbol{\gamma}'(u)$$

This is the surface formed by bending a piece of paper:



A change of parametrization of a surface is the composition

$$\mathbf{r} \circ f : V' \to \mathbf{R}^3$$

where $f : V' \to V$ is a *diffeomorphism* – an invertible map such that f and f^{-1} have derivatives of all orders. Note that if

$$f(x,y) = (u(x,y), v(x,y))$$

then by the chain rule

$$\begin{aligned} (\mathbf{r} \circ f)_x &= \mathbf{r}_u u_x + \mathbf{r}_v v_x \\ (\mathbf{r} \circ f)_y &= \mathbf{r}_u u_y + \mathbf{r}_v v_y \end{aligned}$$

so
$$\begin{pmatrix} (\mathbf{r}\circ f)_x \\ (\mathbf{r}\circ f)_y \end{pmatrix} = \begin{pmatrix} u_x & v_x \\ u_y & v_y \end{pmatrix} \begin{pmatrix} \mathbf{r}_u \\ \mathbf{r}_v \end{pmatrix}.$$
Since f has a differentiable inverse, the Jacobian matrix is invertible, so $(\mathbf{r}\circ f)_x$ and $(\mathbf{r}\circ f)_y$ are linearly independent if $\mathbf{r}_u, \mathbf{r}_v$ are.

Example: The (x, y) plane
$$\mathbf{r}(x, y) = x\mathbf{i} + y\mathbf{j}$$
has a different parametrization in polar coordinates
$$\mathbf{r}\circ f(r, \theta) = r\cos\theta\,\mathbf{i} + r\sin\theta\,\mathbf{j}.$$

We have to consider changes of parametrizations when we pass from one open set V to a neighbouring one V'.

Definition 14 *The tangent plane (or* tangent space*) of a surface at the point a is the vector space spanned by $\mathbf{r}_u(a), \mathbf{r}_v(a)$.*

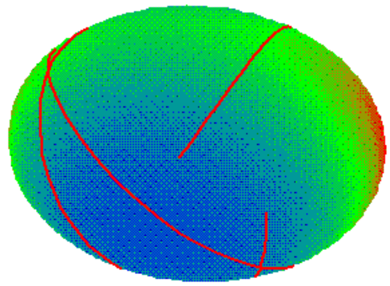
Note that this space is independent of parametrization. One should think of the origin of the vector space as the point a.

Definition 15 *The vectors*
$$\pm\frac{\mathbf{r}_u\wedge\mathbf{r}_v}{|\mathbf{r}_u\wedge\mathbf{r}_v|}$$
are the two unit normals ("inward and outward") to the surface at (u, v).

4.2 The first fundamental form

Definition 16 *A smooth curve lying in the surface is a map $t \mapsto (u(t), v(t))$ with derivatives of all orders such that $\boldsymbol{\gamma}(t) = \mathbf{r}(u(t), v(t))$ is a parametrized curve in $\mathbf{R}^3$.*

A parametrized curve means that $u(t), v(t)$ have derivatives of all orders and $\boldsymbol{\gamma}' = \mathbf{r}_u u' + \mathbf{r}_v v' \neq 0$. The definition of a surface implies that $\mathbf{r}_u, \mathbf{r}_v$ are linearly independent, so this condition is equivalent to $(u', v') \neq 0$.



The arc length of such a curve from $t = a$ to $t = b$ is:

$$\begin{aligned}\int_a^b |\boldsymbol{\gamma}'(t)| dt &= \int_a^b \sqrt{\boldsymbol{\gamma}' \cdot \boldsymbol{\gamma}'} dt \\ &= \int_a^b \sqrt{(\mathbf{r}_u u' + \mathbf{r}_v v') \cdot (\mathbf{r}_u u' + \mathbf{r}_v v')} dt \\ &= \int_a^b \sqrt{E u'^2 + 2F u'v' + G v'^2} dt\end{aligned}$$

where

$$E = \mathbf{r}_u \cdot \mathbf{r}_u, \quad F = \mathbf{r}_u \cdot \mathbf{r}_v, \quad F = \mathbf{r}_v \cdot \mathbf{r}_v.$$

Definition 17 *The first fundamental form of a surface in $\mathbf{R}^3$ is the expression*

$$E du^2 + 2F du dv + G dv^2$$

where $E = \mathbf{r}_u \cdot \mathbf{r}_u, F = \mathbf{r}_u \cdot \mathbf{r}_v, G = \mathbf{r}_v \cdot \mathbf{r}_v$.

The first fundamental form is just the quadratic form

$$Q(\mathbf{v}, \mathbf{v}) = \mathbf{v} \cdot \mathbf{v}$$

on the tangent space written in terms of the basis $\mathbf{r}_u, \mathbf{r}_v$. It is represented in this basis by the symmetric matrix

$$\begin{pmatrix} E & F \\ F & G \end{pmatrix}.$$

So why do we write it as $E du^2 + 2F du dv + G dv^2$? At this stage it is not worth worrying about what exactly du^2 is, instead let's see how the terminology helps to manipulate the formulas.

For example, to find the length of a curve $u(t), v(t)$ on the surface, we calculate

$$\int \sqrt{E\left(\frac{du}{dt}\right)^2 + 2F\frac{du}{dt}\frac{dv}{dt} + G\left(\frac{dv}{dt}\right)^2}\,dt$$

– divide the first fundamental form by dt^2 and multiply its square root by dt.

Furthermore if we change the parametrization of the surface via $u(x, y), v(x, y)$ and try to find the length of the curve $(x(t), y(t))$ then from first principles we would calculate

$$u' = u_x x' + u_y y' \qquad v' = v_x x' + v_y y'$$

by the chain rule and then

$$\begin{aligned} Eu'^2 + 2Fu'v' + Gv'^2 &= E(u_x x' + u_y y')^2 + 2F(u_x x' + u_y y')(v_x x' + \ldots \\ &= (Eu_x^2 + 2Fu_x v_x + Gv_x^2)x'^2 + \ldots \end{aligned}$$

which is heavy going. Instead, using du, dv etc. we just write

$$\begin{aligned} du &= u_x dx + u_y dy \\ dv &= v_x dx + v_y dy \end{aligned}$$

and substitute in $Edu^2 + 2Fdudv + Gdv^2$ to get $E'dx^2 + 2F'dxdy + G'dy^2$. Using matrices, we can write this transformation as

$$\begin{pmatrix} u_x & u_y \\ v_x & v_y \end{pmatrix}\begin{pmatrix} E & F \\ F & G \end{pmatrix}\begin{pmatrix} u_x & v_x \\ u_y & v_y \end{pmatrix} = \begin{pmatrix} E' & F' \\ F' & G' \end{pmatrix}$$

Example: For the plane

$$\mathbf{r}(x, y) = x\mathbf{i} + y\mathbf{j}$$

we have $\mathbf{r}_x = \mathbf{i}, \mathbf{r}_y = \mathbf{j}$ and so the first fundamental form is

$$dx^2 + dy^2.$$

Now change to polar coordinates $x = r\cos\theta, y = r\sin\theta$. We have

$$\begin{aligned} dx &= dr\cos\theta - r\sin\theta d\theta \\ dy &= dr\sin\theta + r\cos\theta d\theta \end{aligned}$$

so that

$$dx^2 + dy^2 = (dr\cos\theta - r\sin\theta d\theta)^2 + (dr\sin\theta + r\cos\theta)^2 = dr^2 + r^2 d\theta^2$$

Here are some examples of first fundamental forms:

Examples:

1. The cylinder

$$\mathbf{r}(u,v) = a(\cos v\,\mathbf{i} + \sin v\,\mathbf{j}) + u\mathbf{k}.$$

We get

$$\mathbf{r}_u = \mathbf{k}, \quad \mathbf{r}_v = a(-\sin v\,\mathbf{i} + \cos v\,\mathbf{j})$$

so

$$E = \mathbf{r}_u \cdot \mathbf{r}_u = 1, \quad F = \mathbf{r}_u \cdot \mathbf{r}_v = 0, \quad G = \mathbf{r}_v \cdot \mathbf{r}_v = a^2$$

giving

$$\boxed{du^2 + a^2 dv^2}$$

2. The cone

$$\mathbf{r}(u,v) = a(u\cos v\,\mathbf{i} + u\sin v\,\mathbf{j}) + u\mathbf{k}.$$

Here

$$\mathbf{r}_u = a(\cos v\,\mathbf{i} + \sin v\,\mathbf{j}) + \mathbf{k}, \quad \mathbf{r}_v = a(-u\sin v\,\mathbf{i} + u\cos v)$$

so

$$E = \mathbf{r}_u \cdot \mathbf{r}_u = 1 + a^2, \quad F = \mathbf{r}_u \cdot \mathbf{r}_v = 0, \quad G = \mathbf{r}_v \cdot \mathbf{r}_v = a^2u^2$$

giving

$$\boxed{(1+a^2)du^2 + a^2u^2dv^2}$$

3. The sphere

$$\mathbf{r}(u,v) = a\sin u\sin v\,\mathbf{i} + a\cos u\sin v\,\mathbf{j} + a\cos v\,\mathbf{k}$$

gives

$$\mathbf{r}_u = a\cos u\sin v\,\mathbf{i} - a\sin u\sin v\,\mathbf{j}, \quad \mathbf{r}_v = a\sin u\cos v\,\mathbf{i} + a\cos u\cos v\,\mathbf{j} - a\sin v\,\mathbf{k}$$

so that

$$E = \mathbf{r}_u \cdot \mathbf{r}_u = a^2\sin^2 v, \quad F = \mathbf{r}_u \cdot \mathbf{r}_v = 0, \quad G = \mathbf{r}_v \cdot \mathbf{r}_v = a^2$$

and so we get the first fundamental form

$$\boxed{a^2dv^2 + a^2\sin^2 v du^2}$$

4. A surface of revolution

$$\mathbf{r}(u,v) = f(u)(\cos v\,\mathbf{i} + \sin v\,\mathbf{j}) + u\mathbf{k}$$

has

$$\mathbf{r}_u = f'(u)(\cos v\,\mathbf{i} + \sin v\,\mathbf{j}) + \mathbf{k}, \quad \mathbf{r}_v = f(u)(-\sin v\,\mathbf{i} + \cos v\,\mathbf{j})$$

so that

$$E = \mathbf{r}_u \cdot \mathbf{r}_u = 1 + f'(u)^2, \quad F = \mathbf{r}_u \cdot \mathbf{r}_v = 0, \quad G = \mathbf{r}_v \cdot \mathbf{r}_v = f(u)^2$$

gives

$$\boxed{(1 + f(u)'^2)du^2 + f(u)^2 dv^2}$$

5. A developable surface

$$\mathbf{r}(u,v) = \boldsymbol{\gamma}(u) + v\mathbf{t}(u).$$

here the curve is parametrized by arc length $u = s$ so that

$$\mathbf{r}_u = \mathbf{t}(u) + v\mathbf{t}'(u) = \mathbf{t} + v\kappa\mathbf{n}, \quad \mathbf{r}_v = \mathbf{t}$$

where $\mathbf{n}$ is the normal to the curve and κ its curvature. This gives

$$\boxed{(1 + v^2\kappa^2)du^2 + 2dudv + dv^2}$$

The analogue of the first fundamental form on an abstract smooth surface X is called a *Riemannian metric*. On each open set U with coordinates (u,v) we ask for smooth functions E, F, G with $E > 0, G > 0, EG - F^2 > 0$ and on an overlapping neighbourhood with coordinates (x,y) smooth functions E', F', G' with the same properties and the transformation law:

$$\begin{pmatrix} u_x & u_y \\ v_x & v_y \end{pmatrix} \begin{pmatrix} E & F \\ F & G \end{pmatrix} \begin{pmatrix} u_x & v_x \\ u_y & v_y \end{pmatrix} = \begin{pmatrix} E' & F' \\ F' & G' \end{pmatrix}$$

A smooth curve on X is defined to be a map $\gamma : [a,b] \to X$ such that $\varphi_U \gamma$ is smooth for each coordinate neighbourhood U on the image. The length of such a curve is well-defined by a Riemannian metric.

Examples:

1. The torus as a Riemann surface has the metric

$$dzd\bar{z} = dx^2 + dy^2$$

as the local holomorphic coordinates are z and $z + m\omega_1 + n\omega_2$ so that the Jacobian matric is the identity. We could also multiply this by any positive smooth doubly-periodic function.

2. The hyperelliptic Riemann surface $w^2 = p(z)$ where $p(z)$ is of degree $2m$ has Riemannian metrics given by

$$\frac{1}{|w|^2}(a_0 + a_1|z|^2 + \ldots + a_{m-2}|z|^{2(m-2)})dzd\bar{z}$$

where the a_i are positive constants.

3. The upper half-space $\{x + iy \in \mathbf{C} : y > 0\}$ has the metric

$$\frac{dx^2 + dy^2}{y^2}.$$

None of these have anything to do with the first fundamental form of the surface embedded in $\mathbf{R}^3$.

We introduced the first fundamental form to measure lengths of curves on a surface but it does more besides. Firstly if two curves $\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2$ on the surface intersect, the angle θ between them is given by

$$\cos\theta = \frac{\boldsymbol{\gamma}_1' \cdot \boldsymbol{\gamma}_2'}{|\boldsymbol{\gamma}_1'||\boldsymbol{\gamma}_2'|} \tag{1}$$

But $\boldsymbol{\gamma}_i' = \mathbf{r}_u u_i' + \mathbf{r}_v v_i'$ so

$$\begin{aligned}
\boldsymbol{\gamma}_i' \cdot \boldsymbol{\gamma}_j' &= (\mathbf{r}_u u_i' + \mathbf{r}_v v_i') \cdot (\mathbf{r}_u u_j' + \mathbf{r}_v v_j') \\
&= E u_i' u_j' + F(u_i' v_j' + u_j' v_i') + G v_i' v_j'
\end{aligned}$$

and each term in (1) can be expressed in terms of the curves and the coefficients of the first fundamental form.

We can also define *area* using the first fundamental form:

Definition 18 *The area of the domain $\mathbf{r}(U) \subset \mathbf{R}^3$ in a surface is defined by*

$$\int_U |\mathbf{r}_u \wedge \mathbf{r}_v| du dv = \int_U \sqrt{EG - F^2} du dv.$$

The second form of the formula comes from the identity

$$|\mathbf{r}_u \wedge \mathbf{r}_v|^2 = (\mathbf{r}_u \cdot \mathbf{r}_u)(\mathbf{r}_v \cdot \mathbf{r}_v) - (\mathbf{r}_u \cdot \mathbf{r}_v)^2 = EG - F^2.$$

Note that the definition of area is independent of parametrization for if

$$\mathbf{r}_x = \mathbf{r}_u u_x + \mathbf{r}_v v_x, \quad \mathbf{r}_y = \mathbf{r}_u u_y + \mathbf{r}_v v_y$$

then

$$\mathbf{r}_x \wedge \mathbf{r}_y = (u_x v_y - v_x u_y)\mathbf{r}_u \wedge \mathbf{r}_v$$

so that

$$\int_U |\mathbf{r}_x \wedge \mathbf{r}_y| dxdy = \int_U |\mathbf{r}_u \wedge \mathbf{r}_v||u_x v_y - v_x u_y| dxdy = \int_U |\mathbf{r}_u \wedge \mathbf{r}_v| dudv$$

using the formula for change of variables in multiple integration.

Example: Consider a surface of revolution

$$(1 + f'(u)^2)du^2 + f(u)^2 dv^2$$

and the area between $u = a, u = b$. We have

$$EG - F^2 = f(u)^2(1 + f'(u)^2)$$

so the area is

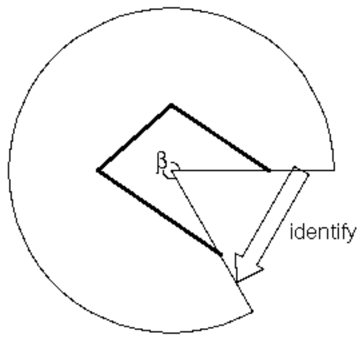
$$\int_a^b f(u)\sqrt{1 + f'(u)^2} dudv = 2\pi \int_a^b f(u)\sqrt{1 + f'(u)^2} du.$$

If a closed surface X is triangulated so that each face lies in a coordinate neighbourhood, then we can define the area of X as the sum of the areas of the faces by the formula above. It is independent of the choice of triangulation.

4.3 Isometric surfaces

Definition 19 *Two surfaces X, X' are isometric if there is a smooth homeomorphism $f : X \to X'$ which maps curves in X to curves in X' of the same length.*

A practical example of this is to take a piece of paper and bend it: the lengths of curves in the paper do not change. The cone and a subset of the plane are isometric this way:



Analytically this is how to tell if two surfaces are isometric:

Theorem 4.1 *The coordinate patches of surfaces U and U' are isometric if and only if there exist parametrizations $\mathbf{r} : V \to \mathbf{R}^3$ and $\mathbf{r}' : V \to \mathbf{R}^3$ with the same first fundamental form.*

Proof: Suppose such a parametrization exists, then the identity map is an isometry since the first fundamental form determines the length of curves.

Conversely, suppose X, X' are isometric using the function $f : V \to V'$. Then

$$\mathbf{r}' \circ f : V \to \mathbf{R}^3, \qquad \mathbf{r} : V \to \mathbf{R}^3$$

are parametrizations using the same open set V, so the first fundamental forms are

$$\tilde{E}du^2 + 2\tilde{F}dudv + \tilde{G}dv^2, \quad Edu^2 + 2Fdudv + Gdv^2$$

and since f is an isometry

$$\int_I \sqrt{\tilde{E}u'^2 + 2\tilde{F}u'v' + \tilde{G}v'^2}dt = \int_I \sqrt{Eu'^2 + 2Fu'v' + Gv'^2}dt$$

for all curves $t \mapsto (u(t), v(t))$ and *all intervals*. Since

$$\frac{d}{dt}\int_a^{a+t} h(u)du = h(t)$$

this means that

$$\sqrt{\tilde{E}u'^2 + 2\tilde{F}u'v' + \tilde{G}v'^2} = \sqrt{Eu'^2 + 2Fu'v' + Gv'^2}$$

for all $u(t), v(t)$. So, choosing u, v appropriately:

$$\begin{aligned} u = t, v = a &\Rightarrow \tilde{E} = E \\ u = a, v = t &\Rightarrow \tilde{G} = G \\ u = t, v = t &\Rightarrow \tilde{F} = F \end{aligned}$$

and we have the same first fundamental form as required. □

Example:

The cone has first fundamental form

$$(1 + a^2)du^2 + a^2u^2dv^2.$$

Put

$$r = \sqrt{1 + a^2}u$$

then we get

$$dr^2 + \left(\frac{a^2}{1 + a^2}\right) r^2 dv^2$$

and now put

$$\theta = \sqrt{\frac{a^2}{1 + a^2}} v$$

to get the plane in polar coordinates

$$dr^2 + r^2 d\theta^2.$$

Note that as $0 \leq v \leq 2\pi$, $0 \leq \theta \leq \beta$ where

$$\beta = \sqrt{\frac{a^2}{1 + a^2}} 2\pi < 2\pi$$

as in the picture.

Example: Consider the unit disc $D = \{x + iy \in \mathbf{C} | x^2 + y^2 < 1\}$ with first fundamental form

$$\frac{4(dx^2 + dy^2)}{(1 - x^2 - y^2)^2}$$

and the upper half plane $H = \{u + iv \in \mathbf{C} | v > 0\}$ with the first fundamental form
$$\frac{du^2 + dv^2}{v^2}.$$
We shall show that there is an isometry from H to D given by
$$w \mapsto z = \frac{w - i}{w + i}$$
where $w = u + iv \in H$ and $z = x + iy \in D$.

We write $|dz|^2 = dx^2 + dy^2$ and $|dw|^2 = du^2 + dv^2$. If $w = f(z)$ where $f : D \to H$ is holomorphic then
$$f'(z) = u_x + iv_x = v_y - iu_y$$
and so
$$|f'(z)|^2 |dz|^2 = (u_x^2 + v_x^2)(dx^2 + dy^2) = (u_x dx + u_y dy)^2 + (v_x dx + v_y dy)^2 = du^2 + dv^2 = |dw|^2.$$
Thus we can substitute
$$|dw|^2 = \left|\frac{dw}{dz}\right|^2 |dz|^2 \tag{2}$$
to calculate how the first fundamental form is transformed by such a map.

The Möbius transformation
$$w \mapsto z = \frac{w - i}{w + i} \tag{3}$$
restricts to a smooth bijection from H to D because $w \in H$ if and only if $|w - i| < |w + i|$, and its inverse is also a Möbius transformation and hence is also smooth. Substituting (3) and (2) with
$$\frac{dw}{dz} = \frac{1}{w + i} - \frac{(w - i)}{(w + i)^2} = \frac{2i}{(w + i)^2}$$
into $v^{-2}|dw|^2$ gives $4(1 - |z|^2)^{-2}|dz|^2$, so this Möbius transformation gives us an isometry from H to D as required.

4.4 The second fundamental form

The first fundamental form describes the intrinsic geometry of a surface – the experience of an insect crawling around it. It is this that we can generalize to abstract surfaces. The second fundamental form relates to the way the surface sits in $\mathbf{R}^3$, though as we shall see, it is not independent of the first fundamental form.

First take a surface $\mathbf{r}(u, v)$ and push it inwards a distance t along its normal to get a one-parameter family of surfaces:

$$\mathbf{R}(u, v, t) = \mathbf{r}(u, v) - t\mathbf{n}(u, v)$$

with

$$\mathbf{R}_u = \mathbf{r}_u - t\mathbf{n}_u, \quad \mathbf{R}_v = \mathbf{r}_v - t\mathbf{n}_v.$$

We now have a first fundamental form $Edu^2 + 2Fdudv + Gdv^2$ depending on t and we calculate

$$\frac{1}{2}\frac{\partial}{\partial t}(Edu^2 + 2Fdudv + Gdv^2)|_{t=0} = -(\mathbf{r}_u \cdot \mathbf{n}_u du^2 + (\mathbf{r}_u \cdot \mathbf{n}_v + \mathbf{r}_v \cdot \mathbf{n}_u)dudv + \mathbf{r}_v \cdot \mathbf{n}_v dv^2).$$

The right hand side is the second fundamental form. From this point of view it is clearly the same type of object as the first fundamental form — a quadratic form on the tangent space.

In fact it is useful to give a slightly different expression. Since $\mathbf{n}$ is orthogonal to $\mathbf{r}_u$ and $\mathbf{r}_v$,

$$0 = (\mathbf{r}_u \cdot \mathbf{n})_u = \mathbf{r}_{uu} \cdot \mathbf{n} + \mathbf{r}_u \cdot \mathbf{n}_u$$

and similarly

$$\mathbf{r}_{uv} \cdot \mathbf{n} + \mathbf{r}_u \cdot \mathbf{n}_v = 0, \quad \mathbf{r}_{vu} \cdot \mathbf{n} + \mathbf{r}_v \cdot \mathbf{n}_u = 0$$

and since $\mathbf{r}_{uv} = \mathbf{r}_{vu}$ we have $\mathbf{r}_u \cdot \mathbf{n}_v = \mathbf{r}_v \cdot \mathbf{n}_u$. We then define:

Definition 20 *The second fundamental form of a surface is the expression*

$$Ldu^2 + 2Mdudv + Ndv^2$$

where $L = \mathbf{r}_{uu} \cdot \mathbf{n}, M = \mathbf{r}_{uv} \cdot \mathbf{n}, N = \mathbf{r}_{vv} \cdot \mathbf{n}$.

Examples:

1) The plane

$$\mathbf{r}(u, v) = \mathbf{a} + u\mathbf{b} + v\mathbf{c}$$

has $\mathbf{r}_{uu} = \mathbf{r}_{uv} = \mathbf{r}_{vv} = 0$ so the second fundamental form vanishes.

2) The sphere of radius a: here with the origin at the centre, $\mathbf{r} = a\mathbf{n}$ so

$$\mathbf{r}_u \cdot \mathbf{n}_u = a^{-1}\mathbf{r}_u \cdot \mathbf{r}_u, \quad \mathbf{r}_u \cdot \mathbf{n}_v = a^{-1}\mathbf{r}_u \cdot \mathbf{r}_v, \quad \mathbf{r}_v \cdot \mathbf{n}_v = a^{-1}\mathbf{r}_v \cdot \mathbf{r}_v$$

and

$$Ldu^2 + 2Mdudv + Ndv^2 = a^{-1}(Edu^2 + 2Fdudv + Gdv^2).$$

The plane is characterised by the vanishing of the second fundamental form:

Proposition 4.2 *If the second fundamental form of a surface vanishes, it is part of a plane.*

Proof: If the second fundamental form vanishes,

$$\mathbf{r}_u \cdot \mathbf{n}_u = 0 = \mathbf{r}_v \cdot \mathbf{n}_u = \mathbf{r}_u \cdot \mathbf{n}_v = \mathbf{r}_v \cdot \mathbf{n}_v$$

so that

$$\mathbf{n}_u = \mathbf{n}_v = 0$$

since $\mathbf{n}_u, \mathbf{n}_v$ are orthogonal to $\mathbf{n}$ and hence linear combinations of $\mathbf{r}_u, \mathbf{r}_v$. Thus $\mathbf{n}$ is constant. This means

$$(\mathbf{r} \cdot \mathbf{n})_u = \mathbf{r}_u \cdot \mathbf{n} = 0, \quad (\mathbf{r} \cdot \mathbf{n})_v = \mathbf{r}_v \cdot \mathbf{n} = 0$$

and so

$$\mathbf{r} \cdot \mathbf{n} = const$$

which is the equation of a plane. □

Consider now a surface given as the graph of a function $z = f(x, y)$:

$$\mathbf{r}(x, y) = x\mathbf{i} + y\mathbf{j} + f(x, y)\mathbf{k}.$$

Here

$$\mathbf{r}_x = \mathbf{i} + f_x\mathbf{k}, \qquad \mathbf{r}_y = \mathbf{j} + f_y\mathbf{k}$$

and so

$$\mathbf{r}_{xx} = f_{xx}\mathbf{k}, \quad \mathbf{r}_{xy} = f_{xy}\mathbf{k}, \quad \mathbf{r}_{yy} = f_{yy}\mathbf{k}.$$

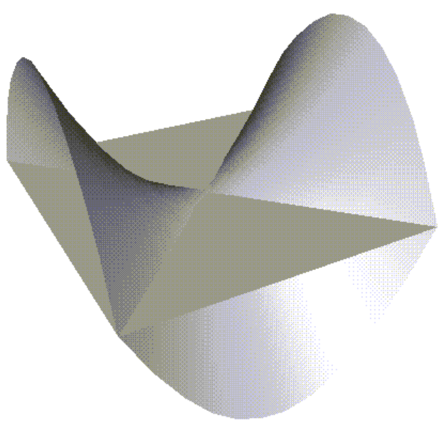
At a critical point of f, $f_x = f_y = 0$ and so the normal is $\mathbf{k}$. The second fundamental form is then the *Hessian* of the function at this point:

$$\begin{pmatrix} L & M \\ M & N \end{pmatrix} = \begin{pmatrix} f_{xx} & f_{xy} \\ f_{xy} & f_{yy} \end{pmatrix}.$$

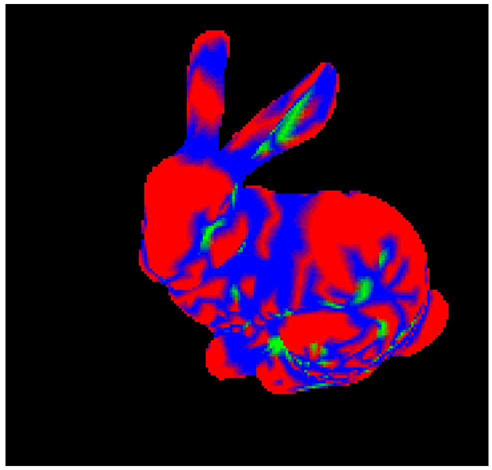
We can use this to qualitatively describe the behaviour of the second fundamental form at different points on the surface. For any point P parametrize the surface by its projection on the tangent plane and then $f(x, y)$ is the height above the plane. Now use the theory of critical points of functions of two variables.

If $f_{xx}f_{yy} - f_{xy}^2 > 0$ then the critical point is a local maximum if the matrix is negative definite and a local minimum if it is positive definite. For the surface the difference is only in the choice of normal so the local picture of the surface is like the sphere – it lies on one side of the tangent plane at the point P.

If on the other hand $f_{xx}f_{yy} - f_{xy}^2 < 0$ we have a saddle point and the surface lies on both sides of the tangent plane:



A general surface has points of both types, like this rabbit:



In fact any closed suface X in $\mathbf{R}^3$, not just rabbit-shaped ones, have both types of points.

Proposition 4.3 *Any closed surface X in $\mathbf{R}^3$ has points at which the second fundamental form is positive definite.*

Proof: Since X is compact, it is bounded and so can be surrounded by a large sphere. Gradually deflate the sphere until at radius R it touches X at a point. With X described locally as the graph of a function f we then have

$$f - (R - \sqrt{R^2 - x^2 - y^2}) \geq 0$$

and the first nonzero term in the Taylor series of this is

$$\frac{1}{2}(f_{xx}x^2 + 2f_{xy}xy + f_{yy}y^2) - \frac{1}{2R}(x^2 + y^2)$$

so

$$Lx^2 + 2Mxy + Ny^2 \geq \frac{1}{R}(x^2 + y^2) > 0.$$

$\square$

It is easy to understand qualitatively the behaviour of a surface from whether $LN-M^2$ is positive or not. In fact there is a closely related function called the Gaussian curvature which we shall study next.

4.5 The Gaussian curvature

Definition 21 *The Gaussian curvature of a surface in* $\mathbf{R}^3$ *is the function*

$$K = \frac{LN - M^2}{EG - F^2}$$

Note that under a coordinate change

$$\begin{pmatrix} u_x & u_y \\ v_x & v_y \end{pmatrix} \begin{pmatrix} E & F \\ F & G \end{pmatrix} \begin{pmatrix} u_x & v_x \\ u_y & v_y \end{pmatrix} = \begin{pmatrix} E' & F' \\ F' & G' \end{pmatrix}$$

so taking determinants

$$(u_x v_y - u_y v_x)^2 (EG - F^2) = (E'G' - F'^2).$$

Since the second fundamental form is a quadratic form on the tangent space just like the first, it undergoes the same transformation, so the ratio $(LN - M^2)/(EG - F^2)$ is independent of the choice of coordinates.

Examples:

1. For a plane, $L = M = N = 0$ so $K = 0$

2. For a sphere of radius a, the second fundamental form is a^{-1} times the first so that $K = a^{-2}$.

We defined K in terms of the second fundamental form which we said describes the extrinsic geometry of the surface. In fact it only depends on E, F, G and its derivatives, and so is intrinsic – our insect crawling on the surface could in principle work it out. It was Gauss who showed this in 1828, a result he was particularly pleased with.

What it means is that if two surfaces are locally isometric, then the isometry maps the Gaussian curvature of one to the Gaussian curvature of the other – for example the Gaussian curvature of a bent piece of paper is zero because it is isometric to the plane. Also, we can define Gaussian curvature for an abstract Riemannian surface.

We prove Gauss's "egregious theorem", as he proudly called it, by a calculation. We consider locally a smooth family of tangent vectors

$$\mathbf{a} = f\mathbf{r}_u + g\mathbf{r}_v$$

where f and g are functions of u, v. If we differentiate with respect to u or v this is no longer necessarily tangential, but we can remove its normal component to make it so, and call this the *tangential derivative*:

$$\begin{aligned} \nabla_u \mathbf{a} &= \mathbf{a}_u - (\mathbf{n}\cdot\mathbf{a}_u)\mathbf{n} \\ &= \mathbf{a}_u + (\mathbf{n}_u\cdot\mathbf{a})\mathbf{n} \end{aligned}$$

since $\mathbf{a}$ and $\mathbf{n}$ are orthogonal.

The important thing to note is that this tangential derivative only depends on E, F, G and their derivatives, because we are taking a tangent vector like $\mathbf{r}_u$, differentiating it to get $\mathbf{r}_{uu}$ and $\mathbf{r}_{uv}$ and then projecting back onto the tangent plane which involves taking dot products like $\mathbf{r}_{uu}\cdot\mathbf{r}_u = (\mathbf{r}_u\cdot\mathbf{r}_u)_u/2 = E_u/2$ etc.

Now differentiate $\nabla_u\mathbf{a}$ tangentially with respect to v:

$$\nabla_v\nabla_u\mathbf{a} = \mathbf{a}_{vu} - (\mathbf{n}\cdot\mathbf{a}_{vu})\mathbf{n} + \nabla_v((\mathbf{n}_u\cdot\mathbf{a})\mathbf{n}).$$

But since we are taking the tangential component, we can forget about differentiating the coefficient of $\mathbf{n}$. Moreover, since $\mathbf{n}$ is a unit vector, $\mathbf{n}_v$ is already tangential, so we get:

$$\nabla_v\nabla_u\mathbf{a} = \mathbf{a}_{vu} - (\mathbf{n}\cdot\mathbf{a}_{vu})\mathbf{n} + (\mathbf{n}_u\cdot\mathbf{a})\mathbf{n}_v$$

Interchanging the roles of u and v and using the symmetry of the second derivative $\mathbf{a}_{uv} = \mathbf{a}_{vu}$ we get

$$\nabla_v\nabla_u\mathbf{a} - \nabla_u\nabla_v\mathbf{a} = (\mathbf{n}_u\cdot\mathbf{a})\mathbf{n}_v - (\mathbf{n}_v\cdot\mathbf{a})\mathbf{n}_u = (\mathbf{n}_u\wedge\mathbf{n}_v)\wedge\mathbf{a}.$$

Now

$$\mathbf{n}_u\wedge\mathbf{n}_v = \lambda\mathbf{n} \tag{4}$$

so we see that $\nabla_v\nabla_u - \nabla_u\nabla_v$ acting on $\mathbf{a}$ rotates it in the tangent plane by $90°$ and multiplies by λ, where λ is intrinsic. Now from (4),

$$\lambda\mathbf{n}\cdot\mathbf{r}_u\wedge\mathbf{r}_v = (\mathbf{n}_u\wedge\mathbf{n}_v)\cdot(\mathbf{r}_u\wedge\mathbf{r}_v) = (\mathbf{n}_u\cdot\mathbf{r}_u)(\mathbf{n}_v\cdot\mathbf{r}_v) - (\mathbf{n}_u\cdot\mathbf{r}_v)(\mathbf{n}_v\cdot\mathbf{r}_u) = LN - M^2$$

but also

$$\mathbf{n} \cdot \mathbf{r}_u \wedge \mathbf{r}_v = \sqrt{EG - F^2}$$

which gives

$$\lambda = (LN - M^2)/\sqrt{EG - F^2}. \tag{5}$$

It follows that $LN - M^2$ and hence K depends only on the first fundamental form.

4.6 The Gauss-Bonnet theorem

One of the beautiful features of the Gaussian curvature is that it can be used to determine the topology of a closed orientable surface – more precisely we can determine the Euler characteristic by integrating K over the surface. We shall do this by using a triangulation and summing the integrals over the triangles, but the boundary terms involve another intrinsic invariant of a curve in a surface:

Definition 22 *The geodesic curvature κ_g of a smooth curve in X is defined by*

$$\kappa_g = \mathbf{t}' \cdot (\mathbf{n} \wedge \mathbf{t})$$

where $\mathbf{t}$ is the unit tangent vector of the curve, which is parametrized by arc length.

This is the tangential derivative of the unit tangent vector $\mathbf{t}$ and so is intrinsic.

The first version of Gauss-Bonnet is:

Theorem 4.4 *Let γ be a smooth simple closed curve on a coordinate neighbourhood of a surface X enclosing a region R, then*

$$\int_\gamma \kappa_g ds = 2\pi - \int_R K dA$$

where κ_g is the geodesic curvature of γ, ds is the element of arc-length of γ, K is the Gaussian curvature of X and dA the element of area of X.

Proof: Recall Stokes' theorem in $\mathbf{R}^3$:

$$\int_C \mathbf{a} \cdot d\mathbf{s} = \int_S \operatorname{curl} \mathbf{a} \cdot d\mathbf{S}$$

for a curve C spanning a surface S. In the xy plane with $\mathbf{a} = (P, Q, 0)$ this becomes Green's formula

$$\int_\gamma (Pu' + Qv')dt = \int_R (Q_u - P_v) du dv \tag{6}$$

Now choose a unit length tangent vector field, for example $\mathbf{e} = \mathbf{r}_u/\sqrt{E}$. Then $\mathbf{e}, \mathbf{n}\wedge\mathbf{e}$ is an orthonormal basis for each tangent space. Since $\mathbf{e}$ has unit length, $\nabla_u\mathbf{e}$ is tangential and orthogonal to $\mathbf{e}$ so there are functions P, Q such that

$$\nabla_u\mathbf{e} = P\,\mathbf{n}\wedge\mathbf{e}, \qquad \nabla_v\mathbf{e} = Q\,\mathbf{n}\wedge\mathbf{e}.$$

In Green's formula, take $\mathbf{a} = (P, Q, 0)$ then the left hand side of (6) is

$$\int_\gamma (u'\nabla_u\mathbf{e} + v'\nabla_v\mathbf{e})\cdot(\mathbf{n}\wedge\mathbf{e}) = \int_\gamma \mathbf{e}'\cdot(\mathbf{n}\wedge\mathbf{e}) \tag{7}$$

Let $\mathbf{t}$ be the unit tangent to γ, and write it relative to the orthonormal basis

$$\mathbf{t} = \cos\theta\,\mathbf{e} + \sin\theta\,\mathbf{n}\wedge\mathbf{e}.$$

So

$$\mathbf{t}'\cdot(\mathbf{n}\wedge\mathbf{e}) = \cos\theta\,\mathbf{e}'\cdot(\mathbf{n}\wedge\mathbf{e}) + \cos\theta\,\theta'.$$

The geodesic curvature of γ is defined by $\kappa_g = \mathbf{t}'\cdot(\mathbf{n}\wedge\mathbf{t})$ so

$$\mathbf{t}' = \alpha\mathbf{n} + \kappa_g\mathbf{n}\wedge\mathbf{t} = \alpha\mathbf{n} + \kappa_g(\cos\theta\,\mathbf{n}\wedge\mathbf{e} - \sin\theta\,\mathbf{e})$$

and so

$$\kappa_g = \mathbf{e}'\cdot(\mathbf{n}\wedge\mathbf{e}) + \theta'.$$

We can therefore write (7) as

$$\int_\gamma (\kappa_g - \theta')ds = \int_\gamma \kappa_g ds - 2\pi.$$

To compute the right hand side of (6), note that

$$\nabla_v\nabla_u\mathbf{e} = \nabla_v(P\,\mathbf{n}\wedge\mathbf{e}) = P_v\,\mathbf{n}\wedge\mathbf{e} + P\,\mathbf{n}\wedge\nabla_v\mathbf{e} = P_v\,\mathbf{n}\wedge\mathbf{e} + PQ\,\mathbf{n}\wedge(\mathbf{n}\wedge\mathbf{e})$$

since $\mathbf{n}_v\wedge\mathbf{e}$ is normal. Interchanging the roles of u and v and subtracting we obtain

$$(\nabla_v\nabla_u - \nabla_u\nabla_v)\mathbf{e} = (P_v - Q_u)\mathbf{n}\wedge\mathbf{e}$$

and from (5) this is equal to $K\sqrt{EG - F^2}$.
Applying Green's theorem and using $dA = \sqrt{EG - F^2}dudv$ gives the result. $\square$

Note that the extrinsic normal was only used to define $\mathbf{n}\wedge\mathbf{e}$ which is one of the two unit tangent vectors to X orthogonal to $\mathbf{e}$. If the surface is orientable we can systematically make a choice and then the proof is intrinsic.

If the curve γ is piecewise smooth – a curvilinear polygon – then θ jumps by the external angle δ_i at each vertex, so the integral of θ' which is 2π in the theorem is replaced by

$$\int_\gamma \theta' ds = 2\pi - \sum_i \delta_i = \sum_i \alpha_i - (n-2)\pi$$

where α_i are the internal angles. The Gauss-Bonnet theorem gives in particular:

Theorem 4.5 *The sum of the angles of a curvilinear triangle is*

$$\pi + \int_R K dA + \int_\gamma \kappa_g ds.$$

Examples:

1. In the plane, a line has constant unit tangent vector and so $\kappa_g = 0$. Since the Gaussian curvature is zero too this says that the sum of the angles of a triangle is π.

2. A great circle on the unit sphere also has κ_g zero, for example if $\gamma(s) = (\cos s, \sin s, 0)$, then $\mathbf{t} = (-\sin s, \cos s, 0)$ and $\mathbf{t}' = -(\cos s, \sin s, 0)$ which is normal to the sphere. Since here $K = 1$, we have, for the triangle Δ with angles A, B, C

$$\alpha + \beta + \gamma = \pi + \text{Area}(ABC).$$

Here is the most interesting version of Gauss-Bonnet:

Theorem 4.6 *If X is a smooth orientable closed surface with a Riemannian metric, then*

$$\int_X K dA = 2\pi\chi(X)$$

Proof: Take a smooth triangulation so that each triangle is inside a coordinate neighbourhood and apply Theorem 4.5 and add. The integrals of κ_g on the edges cancel because the orientation on the edge from adjacent triangles is opposite (this is for Green's theorem – we use the anticlockwise orientation on γ). The theorem gives the total sum of internal angles as

$$\pi F + \int_X K dA.$$

But around each vertex the internal angles add to 2π so we have

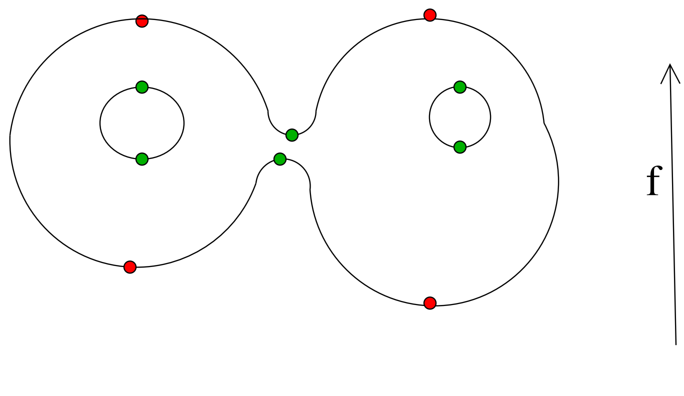
$$2\pi V = \pi F + \int_X K dA$$

and as our faces are triangles whose sides meet in pairs there are $3F/2$ edges. Hence

$$2\pi\chi(X) = 2\pi(V - E + F) = \pi F + \int_X K dA - 3\pi F + 2\pi F = \int_X K dA.$$

$\square$

The Gauss-Bonnet theorem and its method of proof give another formula for the Euler characteristic, involving smooth real-valued functions $f : X \to \mathbf{R}$ on a closed surface X. Since X is compact, f certainly has a maximum and a minimum, but may have other critical points too. Think of a surface in $\mathbf{R}^3$ and the function f given by its height above a plane:



This has 2 maxima, 2 minima and 6 saddle points. We shall be able to calculate the Euler characteristic from these numbers.

First recall that a smooth function $f(u, v)$ has a critical point at a if

$$f_u(a) = f_v(a) = 0.$$

Because of the chain rule, this condition is independent of coordinates: if $u = u(x, y), v = v(x, y)$ then

$$f_x = f_u u_x + f_v v_x, \qquad f_y = f_u u_y + f_v v_y$$

so f_u and f_v vanish if and only if f_x and f_y vanish. This means we can unambiguously talk about the critical points of a smooth function on a surface X.

The Hessian matrix

$$\begin{pmatrix} f_{uu} & f_{uv} \\ f_{uv} & f_{vv} \end{pmatrix}$$

at a critical point transforms like

$$\begin{pmatrix} u_x & u_y \\ v_x & v_y \end{pmatrix} \begin{pmatrix} f_{xx} & f_{xy} \\ f_{xy} & f_{yy} \end{pmatrix} \begin{pmatrix} u_x & v_x \\ u_y & v_y \end{pmatrix} = \begin{pmatrix} f_{uu} & f_{uv} \\ f_{uv} & f_{vv} \end{pmatrix}$$

and so

$$(f_{uu}f_{vv} - f_{uv}^2) = (u_x v_y - u_y v_x)^2 (f_{uu}f_{vv} - f_{uv}^2)$$

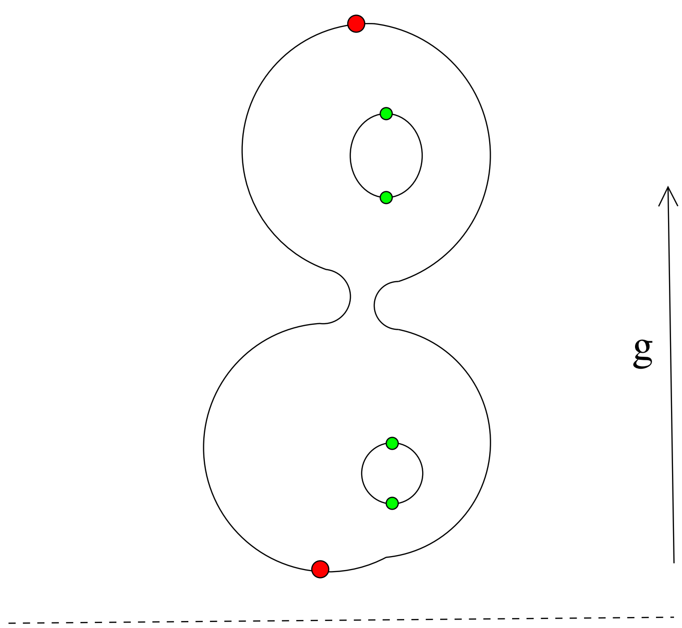
therefore to say that the determinant of the Hessian is non-zero, or positive or negative, is again independent of the choice of coordinate.

Definition 23 *A function f on a surface X has a nondegenerate critical point at $a \in X$ if its Hessian at a is invertible.*

We know from calculus that if $f_{uu}f_{vv} - f_{uv}^2 > 0$ and $f_{uu} > 0$ we have a local minimum, if $f_{uu} < 0$ a local maximum and if $f_{uu}f_{vv} - f_{uv}^2 < 0$ a saddle point. The theorem is the following:

Theorem 4.7 *Let f be a smooth function on a closed surface X with nondegenerate critical points, then the Euler characteristic $\chi(X)$ is the number of local maxima and minima minus the number of saddle points.*

In the picture, we have $\chi(X) = 4 - 6 = -2$ which is correct for the connected sum of two tori. If we turn it on its side we get one maximum, one minimum and 4 saddle points again giving the same value: $2 - 4 = -2$.



Proof: Given a function f on X we can define its gradient vector field:

$$\mathbf{a} = \frac{1}{EG - F^2}[(Gf_u - Ff_v)\mathbf{r}_u + (Ef_v - Ff_u)\mathbf{r}_v]$$

which is normal to the contour lines of f. Away from the critical points we can normalize it to get a unit vector field $\mathbf{e}$. Surround each critical point by a small closed curve γ_i enclosing a disc R_i. Let Y be the complement of the discs, then from the argument of Theorem 4.4

$$\int_Y K dA = -\sum_i \int_{\gamma_i} \mathbf{e}' \cdot (\mathbf{n} \wedge \mathbf{e}) ds$$

using the negative sign because Y is outside R_i.

Inside R_i we choose a unit vector field $\mathbf{f}$ and then we get

$$\int_{R_i} K dA = \int_{\gamma_i} \mathbf{f}' \cdot (\mathbf{n} \wedge \mathbf{f}) ds$$

so adding gives

$$\int_X K dA = \sum_i \int_{\gamma_i} [\mathbf{f}' \cdot (\mathbf{n} \wedge \mathbf{f}) - \mathbf{e}' \cdot (\mathbf{n} \wedge \mathbf{e})] ds.$$

From the proof of the theorem we had

$$\kappa_g = \mathbf{e}' \cdot (\mathbf{n} \wedge \mathbf{e}) + \theta' = \mathbf{f}' \cdot (\mathbf{n} \wedge \mathbf{f}) + \phi'$$

where θ is the angle between γ' and $\mathbf{e}$ and ϕ between γ' and $\mathbf{f}$. So the contribution is just the change in angle between the vector field $\mathbf{e}$ and a fixed one $\mathbf{f}$ which extends. This is an integer multiple of 2π so we can evaluate it by deforming to the standard Euclidean case. A local minimum is $f = x^2 + y^2$ which gives

$$\mathbf{e} = (\cos\theta, \sin\theta)$$

and contributes $+1$, as does the local minimum $-(\cos\theta, \sin\theta)$. For a saddle point $f = x^2 - y^2$ which gives

$$\mathbf{e} = (\cos\theta, -\sin\theta) = (\cos(-\theta), \sin(-\theta)$$

and contributes -1. □

4.7 Geodesics

Geodesics on a surface are curves which are the analogues of straight lines in the plane. Lines can be thought of in two ways:

- shortest curves
- straightest curves

The first point of view says that a straight line minimizes the distance between any two of its points. Conceptually this leads to the idea of stretching a string between two points on a surface until it tightens, and this certainly is one approach to geodesics. The second approach is however generally easier. A line is straightest because its tangent vector doesn't change – it is constant along the line. We generalize this to a curve on a surface by insisting that the component of $\mathbf{t}'$ tangential to the surface should vanish. Or....

Definition 24 *A geodesic on a surface X is a curve $\boldsymbol{\gamma}(s)$ on X such that $\mathbf{t}'$ is normal to the surface.*

From Definition 22 this is the same as saying that the geodesic curvature vanishes.

The general problem of finding geodesics on a surface is very complicated. The case of the ellipsoid is a famous example, needing hyperelliptic functions to solve it – integrals of $dz/\sqrt{p(z)}$ where $p(z)$ is a polynomial of degree 6. But there are cheap ways to find some of them, as in these examples:

Examples:

1) The normal to a curve in the plane is parallel to the plane, so the condition that $\mathbf{t}'$ is normal to the plane means $\mathbf{t}' = 0$ which integrates to $\mathbf{r} = s\mathbf{a} + \mathbf{c}$, the equation of a straight line. Geodesics in the plane really are straight lines, then.

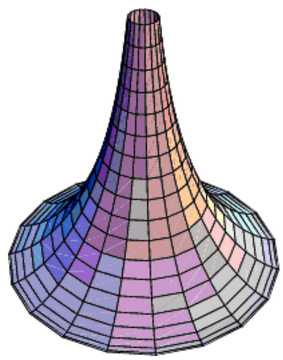
2) Take the unit sphere and a plane section through the origin. We saw earlier that $\kappa_g = 0$ here.

3) Similarly, any plane of symmetry intersects a surface in a geodesic, because the normal to the surface at such a point must be invariant under reflection in the plane of symmetry and hence lie in that plane. It is orthogonal to the tangent vector of the curve of intersection and so $\mathbf{t}'$ points normally.

A useful class of examples is provided by a surface of revolution

$$\mathbf{r}(u, v) = f(u)(\cos v\,\mathbf{i} + \sin v\,\mathbf{j}) + u\mathbf{k}$$

The reflection $(x, y, z) \mapsto (x, -y, z)$ maps the surface to itself, as, by symmetry, does any reflection in a plane containing the z-axis. So the *meridians* $v = const.$ are geodesics:



To find the geodesics in general we need to solve a nonlinear system of ordinary differential equations:

Proposition 4.8 *A curve $\boldsymbol{\gamma}(s) = (u(s), v(s))$ on a surface parametrized by arc length is a geodesic if and only if*

$$\begin{aligned}\frac{d}{ds}(Eu' + Fv') &= \frac{1}{2}(E_u u'^2 + 2F_u u'v' + G_u v'^2) \\ \frac{d}{ds}(Fu' + Gv') &= \frac{1}{2}(E_v u'^2 + 2F_v u'v' + G_v v'^2)\end{aligned}$$

Proof: We have for the curve $\boldsymbol{\gamma}$

$$\mathbf{t} = \mathbf{r}_u u' + \mathbf{r}_v v'$$

and it is a geodesic if and only if $\mathbf{t}'$ is normal i.e.

$$\mathbf{t}' \cdot \mathbf{r}_u = \mathbf{t}' \cdot \mathbf{r}_v = 0.$$

Now

$$\mathbf{t}' \cdot \mathbf{r}_u = (\mathbf{t} \cdot \mathbf{r}_u)' - \mathbf{t} \cdot \mathbf{r}_u'$$

so the first equation is

$$(\mathbf{t} \cdot \mathbf{r}_u)' = \mathbf{t} \cdot \mathbf{r}_u'.$$

The left hand side is

$$\frac{d}{ds}((\mathbf{r}_u u' + \mathbf{r}_v v') \cdot \mathbf{r}_u) = \frac{d}{ds}(Eu' + Fv')$$

an the right hand side is

$$\begin{aligned}\mathbf{t}\cdot(\mathbf{r}_{uu}u'+\mathbf{r}_{uv}v') &= \mathbf{r}_u\cdot\mathbf{r}_{uu}u'^2+(\mathbf{r}_v\cdot\mathbf{r}_{uu}+\mathbf{r}_u\cdot\mathbf{r}_{uv})u'v'+\mathbf{r}_v\cdot\mathbf{r}_{uv}v'^2\\ &= \frac{1}{2}E_u u'^2+(\mathbf{r}_v\cdot\mathbf{r}_u)_u u'v'+\frac{1}{2}G_u v'^2\\ &= \frac{1}{2}(E_u u'^2+2F_u u'v'+G_u v'^2)\end{aligned}$$

The other equation follows similarly. □

It is clear from 4.8 that geodesics only depend on the first fundamental form, so that geodesics can be defined for abstract surfaces and moreover an isometry takes geodesics to geodesics.

Examples:

1) The plane: $E=1, F=0, G=1$ in Cartesian coordinates, so the geodesic equations are

$$x''=0=y''$$

which gives straight lines

$$x=\alpha_1 s+\beta_1,\quad y=\alpha_2 s+\beta_2.$$

2) The cylinder

$$\mathbf{r}(u,v)=a(\cos v\,\mathbf{i}+\sin v\,\mathbf{j})+u\mathbf{k}$$

has first fundamental form

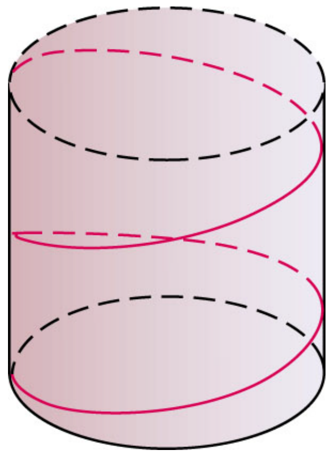
$$du^2+a^2dv^2=du^2+d(av)^2.$$

This is isometric to the plane so the geodesics are of the form

$$u=\alpha_1 s+\beta_1,\quad v=\alpha_2 s+\beta_2$$

which gives a helix

$$\boldsymbol{\gamma}=a(\cos(\alpha_2 s+\beta_2)\,\mathbf{i}+\sin(\alpha_2 s+\beta_2)\,\mathbf{j})+(\alpha_1 s+\beta_1)\mathbf{k}$$



The differential equation for geodesics gives us the following general fact:

Proposition 4.9 *Through each point P on a surface and in each direction at P there passes a unique geodesic.*

Proof: We are solving a differential equation of the form

$$u'' = a(u, v, u', v'), \quad v'' = b(u, v, u', v')$$

or equivalently a first order system

$$\begin{aligned} u' &= p \\ v' &= q \\ p' &= a(u, v, p, q) \\ q' &= b(u, v, p, q) \end{aligned}$$

and the Cauchy existence theorem (see Appendix B) gives a unique solution with initial conditions (u, v, p, q), namely the point of origin and the direction. $\square$

Example: Given a point $\mathbf{a}$ on the unit sphere and a tangential direction $\mathbf{b}$ the span of $\mathbf{a}, \mathbf{b}$ is a plane through the origin which meets the sphere in a great circle through $\mathbf{a}$ with tangent $\mathbf{b}$. Thus *every* geodesic is a great circle.

There is one case – a surface of revolution – where the geodesic equations can be "solved", or anyway, reduced to a single integration. We have

$$E = 1 + f'(u)^2, \quad F = 0, \quad G = f(u)^2$$

and the equations become

$$\begin{aligned}\frac{d}{ds}((1+f'^2)u') &= f'(f''u'^2+fv'^2)\\ \frac{d}{ds}(f^2v') &= 0\end{aligned}$$

We ignore the first equation – it is equivalent to a more obvious fact below. The second says that

$$f^2v' = c \tag{8}$$

where c is a constant. Now use the fact that the curve is parametrized by arc length (this is an "integral" of the equations), and we get

$$(1+f'^2)u'^2 + f^2v'^2 = 1 \tag{9}$$

Substitute for v' from (8) in (9) to get

$$(1+f'^2)u'^2 + \frac{c^2}{f^2} = 1$$

and then

$$s = \int f\sqrt{\frac{1+f'^2}{f^2-c^2}}du$$

which is "only" an integration. Having solved this by $u = h(s)$, v can be determined by a further integration from (8):

$$v(s) = \int \frac{c}{f(h(t))^2}dt.$$

If we are only interested in the curve and not its arclength parametrization, then (8) and (9) give

$$(1+f'(u)^2)\left(\frac{du}{dv}\right)^2 + f(u)^2 = \frac{f(u)^4}{c^2}$$

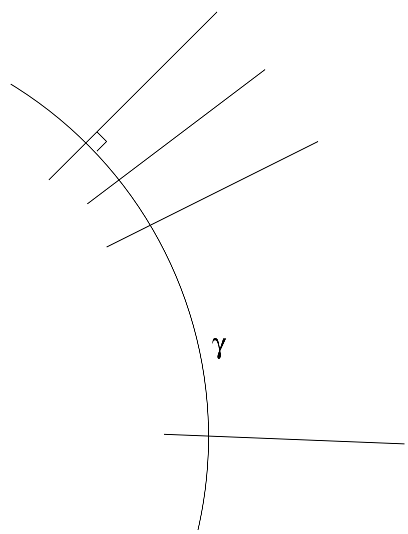
which reduces to the single integration

$$v = \int \frac{c}{f(u)}\sqrt{\frac{1+f'(u)^2}{f(u)^2-c^2}}du.$$

4.8 Gaussian curvature revisited

We may not be able to solve the geodesic equations explicitly, but existence of geodesics through a given point and in a given direction give rise to various natural coordinate systems, modelled on Cartesian coordinates. Here is one: choose a geodesic γ parametrized by arc length. Through the point $\gamma(v)$ take the geodesic $\gamma_v(s)$ which intersects γ orthogonally, and define

$$\mathbf{r}(u, v) = \gamma_v(u).$$



Since $\mathbf{r}_u$ and $\mathbf{r}_v$ are orthogonal at $u = 0$ they are linearly independent in a neighbourhood and so are good coordinates.

Now the curves $v = const.$ are parametrized by arc length, so $E = 1$. These curves are also geodesics and u is arc length so in the second geodesic equation

$$\frac{d}{ds}(Fu' + Gv') = \frac{1}{2}(E_v u'^2 + 2F_v u'v' + G_v v'^2)$$

we put $v = const.$ and $u = s$ which, with $E = 1$, gives $F_u = 0$. But F vanishes at $u = 0$ because the two geodesics are orthogonal there, hence $F = 0$ and the first fundamental form is

$$du^2 + G(u, v)dv^2.$$

In this form the Gaussian curvature is simple:

Proposition 4.10 *The Gaussian curvature of the metric $du^2 + G(u, v)dv^2$ is*

$$K = -G^{-1/2}(G^{1/2})_{uu}$$

Examples:

1. For the plane $dx^2 + dy^2$, $G = 1$ and $K = 0$.

2. For the unit sphere with first fundamental form $du^2 + \sin^2 u dv^2$, $G = \sin^2 u$ so

$$K = -\frac{1}{\sin u}(\sin u)_{uu} = \frac{1}{\sin u}\sin u = 1.$$

3. For the upper half-space with metric $(dx^2 + dy^2)/y^2$ put $u = \log y$ and $v = x$ and then we have $du^2 + e^{-2u}dv^2$, so that

$$K = -e^u(e^{-u})_{uu} = -e^u e^{-u} = -1.$$

Proof: Recall the tangential derivative ∇: the tangential component of the ordinary derivative. Then since by construction $\mathbf{r}_u$ is the unit tangent vector of a geodesic, by the definition of a geodesic its u-derivative is normal so $\nabla_u \mathbf{r}_u = 0$.

Consider now $\nabla_v \mathbf{r}_u = A\mathbf{r}_u + B\mathbf{r}_v$. The dot product with $\mathbf{r}_u$ gives

$$E_v/2 = \mathbf{r}_{vu} \cdot \mathbf{r}_u = A$$

but $E = 1$ so $A = 0$.

Using $E = 1$ and $F = 0$ the product with $\mathbf{r}_v$ gives

$$G_u/2 = \mathbf{r}_v \cdot \mathbf{r}_{vu} = BG.$$

Now from (5)

$$(\nabla_v\nabla_u - \nabla_u\nabla_v)\mathbf{r}_u = K\sqrt{EG - F^2}\mathbf{n}\wedge\mathbf{r}_u = KG^{1/2}(\mathbf{r}_v G^{-1/2}) = K\mathbf{r}_v$$

But the left hand side (using $\nabla_u \mathbf{r}_v = \nabla_v \mathbf{r}_u$ which follows from $\mathbf{r}_{uv} = \mathbf{r}_{vu}$) is

$$-\nabla_u(G_u/2G)\mathbf{r}_v = -((G_u/2G)_u + (G_u/2G)^2)\mathbf{r}_v$$

which gives the result. $\square$

With this coordinate system we can characterize surfaces with *constant* Gaussian curvature:

Theorem 4.11 *A surface with $K = 0$ is locally isometric to the plane, with $K = 1$ locally isometric to the unit sphere and with $K = -1$ locally isometric to the upper half space with metric $(dx^2 + dy^2)/y^2$.*

Proof: Use the form $du^2 + Gdv^2$.

i) If $K = 0$ then $(G^{1/2})_{uu} = 0$ so $G = A(v)u + B(v)$. But at $u = 0$, $\mathbf{r}_u$ and $\mathbf{r}_v$ are unit so $B(v) = 1$. Also, the curve $u = 0$ is a geodesic – the initial curve γ – with v arc length. So the geodesic equation

$$\frac{d}{ds}(Eu' + Fv') = \frac{1}{2}(E_u u'^2 + 2F_u u'v' + G_u v'^2)$$

gives $0 = G_u(0, v)/2$ and this means in our case $A(v) = 0$. The first fundamental form is therefore $du^2 + dv^2$ and by 4.1 this is isometric to the plane.

ii) If $K = 1$, the equation for $G^{1/2}$ is

$$(G^{1/2})_{uu} + G^{1/2} = 1$$

which is solved by $G^{1/2} = A(v)\sin u + B(v)\cos u$. The boundary conditions give $G = \cos^2 u$ and the metric $du^2 + \cos^2 u\, dv^2$ – the sphere.

iii) If $K = -1$ we have $du^2 + \cosh^2 u\, dv^2$. The substitution $x = v\tanh u, y = v\,\mathrm{sech}\, u$ takes this into $(dx^2 + dy^2)/y^2$. $\square$