Geometry of curves and surfaces

B. H. Bowditch
University of Warwick
[October 2012]

0. Introduction.

These are informal notes intended to accompany the course MA3D9. These focus on the theory. We leave pictures and many of the examples to lectures and support classes.

These notes might not coincide exactly with what is done in lectures. Where there is any difference, it is the material presented in lectures that will determine the examinable content of the course.

Sources :

I have made much use of the lecture notes by Nigel Hitchin [H], combined with material from McCleary's book [M], as well as some examples from do Carmo's Book [C]. I would probably also have made use of [T] if I'd been aware of its existence earlier.

Lecture notes :

[H] Nigel Hitchin, "The geometry of surfaces", University of Oxford, available at:
http://people.maths.ox.ac.uk/∼hitchin/hitchinnotes/hitchinnotes.html.

[T] George Terizakis, "MA3D9 Geometry of curves and surfaces", University of Warwick, available via MathStuff.

Books :

[M] John McCleary, "Geometry from a differentiable viewpont", CUP 1994.

[S] Dirk J. Struik, "Lectures on classical differential geometry" : Addison-Wesley 1950

[C] Manfredo P. do Carmo, "Differential geometry of curves and surfaces" : Prentice-Hall 1976

[O] Barrett O'Neill, "Elementary differential geometry" : Academic Press 1966

[MR] Sebastian Montiel, Antonio Ros, "Curves and surfaces" : American Mathematical Society 1998

[G] Alfred Gray, "Modern differential geometry of curves and surfaces" : CRC Press 1993

Web page :

A web page for this course will be maintained at:
http://www.warwick.ac.uk/∼masgak/cas/course.html

Contents :

History :

Most of the theory described here was developed in the first half of the 19th century, though of course, we adopt more modern terminology and notation.

The notions of curvature and torsion of space curves were developed in independent work of Frenet and Serret around 1850. The differential geometry of surfaces was explored by Euler in the 18th century and later by Monge. Much of what we describe here originates in work of Gauß around 1825.

Around 1854, Riemann developed a much more general theory of manifolds. This allows one to describe surfaces abstractly, without describing any embedding in euclidean space. If there is time, we will say something about this at the end of the course.

Background and notation.

(a) Linear algebra.

$\mathbf{R}^n$ will denote n-dimensional euclidean space. An element $(x_1, x_2, \ldots, x_n) \in \mathbf{R}^n$ can be thought of either as locating a point is space, or as a vector indicating direction. In its latter role, it will frequently be denoted as a column vector:

$$\mathbf{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}.$$

Formally however, there is no distinction between these — a point is the same as its position vector from the origin.

Given $\mathbf{x}, \mathbf{y} \in \mathbf{R}^n$ we write $\mathbf{x}.\mathbf{y}$, $||\mathbf{x}|| = \sqrt{\mathbf{x}.\mathbf{x}}$ and $\mathbf{x} \wedge \mathbf{y}$ respectively for the dot product, norm, and wedge (or cross) product. The angle between $\mathbf{x}$ and $\mathbf{y}$ is $\cos^{-1}(\mathbf{x}.\mathbf{y}/||\mathbf{x}|| \, ||\mathbf{y}||)$. The distance between $\mathbf{x}$ and $\mathbf{y}$ is $||\mathbf{x} - \mathbf{y}||$.

A linear map $\Phi : \mathbf{R}^n \longrightarrow \mathbf{R}^m$ has the form $[\mathbf{x} \mapsto A\mathbf{x}]$ for an $m \times n$ matrix A. If $m = n$, Φ or A is "orthogonal" if $A^T A = AA^T = I$ (where A^T denotes transpose).

Exercise : Φ is orthogonal $\Leftrightarrow (\forall \mathbf{x}, \mathbf{y})(\Phi(\mathbf{x}).\Phi(\mathbf{y}) = \mathbf{x}.\mathbf{y}) \Leftrightarrow (\forall \mathbf{x})(||\Phi(\mathbf{x})|| = ||\mathbf{x}||)$.

A "dilation" is a linear map of the the form $\mathbf{x} \mapsto \lambda\mathbf{x}$, for some $\lambda > 0$, i.e. given by the matrix λI.

If Φ is invertible, Φ is "orientation preserving" (respectively "reversing") if $\det A > 0$ (respectively $\det A < 0$).

Exercise : If Φ is orthogonal, then $\Phi(\mathbf{x} \wedge \mathbf{y}) = \pm\Phi(\mathbf{x}) \wedge \Phi(\mathbf{y})$, where the sign depends on whether it is orientation preserving or reversing.

Definition : An *affine map* from $\mathbf{R}^n$ to $\mathbf{R}^m$ is a map of the form $\mathbf{x} \mapsto A\mathbf{x} + \mathbf{b}$ where A is an $m \times n$ matrix, and $\mathbf{b} \in \mathbf{R}^m$.

Definition : A *rigid motion* is an affine map where A is orthogonal.

(A rigid motion is commonly called an "isometry", though we shall reserve that term here for isometric maps between surfaces, as described in Section 4.)

Exercises :

(1) A rigid motion preserves distances.

(2) (Harder) A map from $\mathbf{R}^n$ to $\mathbf{R}^n$ that preserves distances is a rigid motion. (The "hard" part is showing that such a map must be affine.)

We make some use of bilinear and quadratic forms. The relevant background will be discussed in Section 6.

(b) Topology.

We will use some basic notions from topology, namely open and closed sets, continuity etc. In practice these will only be applied to subsets of $\mathbf{R}^n$.

(c) Differentiation.

Definition : We say that a map $f : \mathbf{R}^n \longrightarrow \mathbf{R}^m$ is *smooth* if it has derivatives of all orders.

In practice, we will only be taking derivatives up to third order here.

If $p \in \mathbf{R}^n$, the derivative of f at p is a linear map $f_{*,p} : \mathbf{R}^n \longrightarrow \mathbf{R}^m$. We will abbreviate this to f_* when the point p is implicitly understood. It is represented by the $m \times n$ "jacobian" matrix $(\partial f_i/\partial x_j)_{ij}$. We frequently abbreviate $\partial f_i/\partial x_j$ to $(f_i)_{x_j}$. Note that f_* is injective if and only if it has rank n (which implies $n \leq m$).

Theorem 0.1 : *If f_* has rank n, then f is locally injective at p. (That is, there is an open set, $U \subseteq \mathbf{R}^n$ with $p \in U$ such that $f|U$ is injective.)*

Theorem 0.2 : (Inverse function theorem) *If $f_* : \mathbf{R}^n \longrightarrow \mathbf{R}^n$ has rank n (i.e. is invertible) then there are open subsets $U, V \subseteq \mathbf{R}^n$ with $p \in U$ such that $f|U$ is bijective, and $(f|U)^{-1} : V \longrightarrow U$ is also smooth.*

We refer to $f|U$ as a diffeomorphism from U to V.
A proof of Theorem 0.2 can be found in [MA225 ???]. Theorem 0.1 can be deduced from Theorem 0.2 (exercise).

We will also be using a result on the existence and uniqueness of solutions of systems of first order linear differential equations (Theorem 3.4)).

1. Curves in $\mathbf{R}^n$.

Let $I \subseteq \mathbf{R}$ be an interval, i.e. I has the form $[a,b]$, $(a,b]$, $[a,b)$ or (a,b) with $a \in [-\infty, \infty)$, $b \in (-\infty, \infty]$.
Let $\beta : I \longrightarrow \mathbf{R}^n$ be smooth. We write

$$\beta'(t) = \frac{\partial \beta}{\partial t}$$

Definition : $\beta'(t)$ is the *tangent* to β at t.

Definition : $||\beta'(t)||$ is the *speed*.

Example : The curve $\beta :\longrightarrow \mathbf{R}^2$ given by $\beta(t) = (t^2, t^3)$ has $\beta'(t) = (2t, 3t^2)$. In particular, $\beta'(0) = (0,0)$.

Definition : β is *regular* at t if $||\beta'(t)|| \neq 0$.
β is *regular* if it is regular at t for all $t \in I$.

Examples :
(1) Straight line : $t \mapsto \mathbf{a}t + b$.
(2) Circle : $t \mapsto (a \cos \omega t, a \sin \omega t)$
(3) Helix : $t \mapsto (a \cos \omega t, a \sin \omega t, b\omega t)$
Straight line, circle, helix.

Note that the curve $[t \mapsto (t^2, t^3)$ described above not regular.

Exercise : In fact there is no regular curve with the same image.

Definition : The *length* of β is defined as

$$\text{length}(\beta) = \int_I ||\beta'(t)||dt.$$

(This is finite if $I = [a, b]$ and with $a, b \in \mathbf{R}$.)

Exercise : $d(\beta(a), \beta(b)) \leq \text{length}\, \beta)$.

Reparameterisation.

Definition : Suppose I, J are intervals. A *diffeomorphism* from I to J is a map $\sigma : I \longrightarrow J$ which is bijective and such that σ and σ^{-1} are smooth.

Note : : For σ^{-1} to be smooth, it's enough that σ be smooth and $\sigma'(t) \neq 0$ for all t.

Definition : If $\beta : I \longrightarrow \mathbf{R}^n$ and $\gamma : J \longrightarrow \mathbf{R}^n$ are curves, then γ is a *reparameterisation* of β of there is a diffeomorphism $\sigma : J \longrightarrow I$ such that $\gamma = \beta \circ \sigma$.

(Note that this is reflexive symmetric and transitive.)

The chain rule tells us that

$$\gamma'(u) = \beta'(\sigma(u))\sigma'(u).$$

Lemma 1.1 : *Length is invariant under reparameterisation.*

Proof : Let $\gamma = \beta \circ \sigma$. Then

$$\begin{aligned} \text{length}(\gamma) &= \int_J ||\gamma'(u)||\, du \\ &= \int_I ||\beta'(\sigma(u))||\, |\sigma'(u)|\, dt \\ &= \int_I ||\beta'(t)||\, dt \\ &= \text{length}(\beta). \end{aligned}$$

$\diamondsuit$

Suppose $\beta : I \longrightarrow \mathbf{R}^n$ is smooth where $I = [a, b]$. Given $t \in [a, b]$, let

$$s(t) = \int_a^t ||\beta'(u)||\, du = \text{length}(\beta|[a, t]).$$

Then $s'(t) = ||\beta'(t)||$. We have $s : I \longrightarrow J$, where $J = [0, \text{length}(\beta)]$. If β is regular, then s is a diffeomorphism. Let $\sigma : J \longrightarrow I$ be the inverse function, and let $\gamma = \beta \circ \sigma$. Thus, $\gamma : J \longrightarrow \mathbf{R}^n$ is a reparameterisation of γ. Moreover, $||\gamma'(t)|| = ||\beta'(\sigma(t))|| \, |\sigma'(t)| = 1$, i.e. γ has unit speed.

Definition : We refer to s as *arc length* and to γ as the arc-length reparameterisation of β.

Example : Circle

Piecewise smooth paths.

Suppose $a = t_0 < t_1 < \cdots < t_n = b$. A curve $\beta : [a, b] \longrightarrow \mathbf{R}^n$ is piecewise smooth if β is continuous and $\beta|[t_i, t_{i+1}]$ is smooth for all i. Let

$$\text{length}(\beta) = \sum_{i=1}^{n} \text{length}(\beta|[t_{i-1}, t_i]).$$

Definition : β is *polygonal* if each segment $\beta|[t_i, t_{i+1}]$ is a straight line. We refer to the points $\beta(t_i)$ as its *vertices*.

Clearly β is determined (up to reparameterisation) by its vertices.

Exercise : If β is polygonal, then

$$\text{length}(\beta) = \sum_{i=1}^{n} ||\beta(t_i) - \beta(t_{i-1})||.$$

Fact : If β is a smooth curve, given $\epsilon > 0$, there is some $\delta > 0$ such that if $a = t_0 < t_1 < \cdots < t_n = b$ and $|t_{i+1} - t_i| \leq \delta$ for all i, then if γ is the polygonal curve with vertices $\beta(t_0), \beta(t_1), \ldots, \beta(t_n)$, then $\text{length}(\beta) - \text{length}(\gamma) \leq \epsilon$. (Note that by the from the earlier exercise, we always have $\text{length}\, \beta \leq \text{length}\, \gamma$.)

2. Planar curves.

Let $\gamma : I \longrightarrow \mathbf{R}^2$ be a curve parameterised by arc-length. Given $s \in I$, let $\mathbf{T}(s) = \gamma'(s)$ be the unit tangent.

Let $\mathbf{N}(s)$ be the unit normal — obtained by rotating $\mathbf{T}(s)$ through $\pi/2$ anticlockwise.

Now $\mathbf{T}.\mathbf{T} = 1$ and so $\mathbf{T}.\mathbf{T}' = 0$. Thus we can write

$$\mathbf{T}'(s) = \kappa(s)\mathbf{N}(s),$$

where $\kappa(s) \in \mathbf{R}$. Here κ is a smooth function of s. (Question: why is it smooth?)

Definition : κ is the *(signed) curvature* of γ.
(We refer to $|\kappa|$ as the unsigned curvature.)

Note that

$$|\kappa(s)| = ||\mathbf{T}'(s)|| = ||\gamma''(s)||.$$

Example : Circle of radius r. This has constant curvature $1/r$.

Definition : If $\kappa \neq 0$, then $\rho(s) = 1/\kappa(s)$ is the *(signed) radius of curvature.*
The point $c(s) = \gamma(s) + \rho(s)\mathbf{N}(s)$ is the *centre of curvature.*
The circle of centre $c(s)$ and radius $\rho(s)$ is the *osculating circle.*

Note that for a circle, the osculating circle is constant and equal to the original curve.

General formula.

Suppose $\gamma(t) = (x(t), y(t))$ is regular (not necessarily parameterised by arc-length).

Proposition 2.1 : *The unsigned curvature of γ is given by:*

$$|\kappa(t)| = \left| \frac{x''y' - y''x'}{((x')^2 + (y')^2)^{3/2}} \right|.$$

Proof : Exercise. $\diamondsuit$

Example : The ellipse: $\gamma(t) = (a\cos t, b\sin t)$:

$$\kappa(t) = \frac{ab}{\sqrt{(a^2 + (b^2 - a^2)\cos^2 t)^3)}}.$$

What are the maximum and minimum curvatures?

Remarks :
(1) Unsigned curvature is invariant under (postcompositon with) rigid motion of $\mathbf{R}^2$. The sign remains the same if and only if the rigid motion is orientation preserving. What are the maximum and minimum curvatures?
(2) Under dilation by a factor of $\lambda > 0$, it gets multiplied by $1/\lambda$.

The tractrix.

The tractrix is defined as a curve $\gamma : (0, 1] \longrightarrow \mathbf{R}^2$ with $\gamma(0) = (1, 0)$, such that the tangent at any point $\gamma(t)$ intercepts the y-axis at a point a unit distance away from $\gamma(t)$. (As usual we write $\gamma(t) = (x(t), y(t))$.) In other words:

$$\frac{dy}{dx} = -\frac{\sqrt{1-x^2}}{x}.$$

Using the formula for curvature we get:

$$|\kappa| = \frac{x}{\sqrt{1-x^2}}.$$

Let $s \mapsto \gamma(s)$ $s \geq 0$ be the arc length paramerisation, with $\gamma(0) = (1, 0)$. Since $(x')^2 + (y')^2 = 1$, we get $x' = -x$ (for example, note that $\frac{dy}{dx} = y'/x'$ and express $1 + \left(\frac{dy}{dx}\right)^2$ in terms of x and in terms of x'). Solving this differential equation with $x(0) = 0$, we get $x = e^{-s}$.

Also

$$\frac{dy}{ds} = -\frac{\sqrt{1-x^2}}{x}\frac{dx}{ds} = \sqrt{1-x^2} = \sqrt{1-e^{-2s}}.$$

Integrating:

$$y = \cosh^{-1}(e^s) - \sqrt{1-e^{-2s}}$$

so

$$\gamma(s) = (e^{-s}, \cosh^{-1}(e^s) - \sqrt{1-e^{-2s}})$$
$$\kappa(s) = e^{-s}/\sqrt{1-e^{-2s}}.$$

Another parametrisation, used in Section 8, is derived by taking $x = \sin(\theta)$. Integrating for y (as in lectures) we get:

$$\gamma(\theta) = (\sin(\theta), -\log(\tan(\theta/2)) - \cos(\theta)).$$

Note that $\kappa(\theta) = \tan(\theta)$.

Other examples : Ellipse, cycloid etc.

Closed curves.

Definition : A curve $\gamma : \mathbf{R} \longrightarrow \mathbf{R}^2$ is *periodic* with *period* $t_0 > 0$ if $\gamma(t + t_0) = \gamma(t)$ for all $t \in \mathbf{R}$.

Example : The circle $\gamma(t) = (r\cos t, r\sin t)$ has period 2π.

Definition : A *fundamental domain* is any interval of the form $[a, b] \subseteq \mathbf{R}$, where $b = a + t_0$.

Thus $\gamma(a) = \gamma(b)$, $\gamma'(a) = \gamma'(b)$, $\gamma''(a) = \gamma''(b)$ etc.

Remark : We can more intuitively think of γ as a map of the circle into $\mathbf{R}^2$ — gluing the endpoints a and b together to form circle. In these terms, we can refer to γ as a *closed curve*. To be more specific we can identify the circle with the unit circle in the plane, with the parameter t corresponding to the point $(\cos(2\pi t/t_0), \sin(2\pi t/t_0))$ in the circle.

If $f(t)$ is any continuous periodic function in t, we can define $\int_\gamma f(t)dt = \int_a^b f(t)dt$, where $[a, b]$ is any fundamental domain. It is independent of the choice of fundamental domain. Again, we can think of it more informally as an integral over the circle.

Total turning

Suppose that γ is a regular closed curve. We write $\mathbf{T}(t)$ for the unit tangent. Thus $\mathbf{T}(t) = (\cos\theta(t), \sin\theta(t))$ for some $\theta(t) \in \mathbf{R}$ well defined up to some integer multiple of 2π.

Definition : The *total turning* of γ is equal to $\int_\gamma \theta'(t)dt$.

Since $\theta(a) = \theta(b)$ up to some multiple of 2π, we have that the total turning of γ is equal to $2\pi n$ for some $n \in \mathbf{Z}$. Here n is called the *turning number* of γ.

Examples : $n = 1, -1, 0, 2$ etc.

Remark : We can define total turning for any regular closed curve: set $\gamma(t) = r(t)(\cos\theta(t), \sin\theta(t))$. Note that this is invariant under reparameterisation.

Jordan curves.

We state some facts which are intuitively clear, though they are more difficult to prove formally. They more properly belong in a course on topology.

Definition : A (smooth) Jordan curve is a closed curve that is injective (as a map from the circle into $\mathbf{R}^2$).

Examples : Circle, ellipse.

The following, though intuitively reasonable, will not be proven in this course.

Theorem 2.2 : (Smooth Schoenflies) *If γ is a smooth Jordan curve, then there is a diffeomorphism $f : \mathbf{R}^2 \longrightarrow \mathbf{R}^2$ sending the unit circle to γ.*

Here a "diffeomorphism" f is a bijective map $f : \mathbf{R}^2 \longrightarrow \mathbf{R}^2$ such that f and f^{-1} are both smooth.

If we think of γ as a map of the unit circle into $\mathbf{R}^2$ then this means that f is simply an extention of $\gamma : S^1 \longrightarrow \mathbf{R}^2$.

If we think of γ as a periodic map of period t_0, then we get

$$\gamma(t) = f((\cos(2\pi t/t_0), \sin(2\pi t/t_0)).$$

Let $D = \{x \in \mathbf{R}^2 \mid ||x|| \leq 1\}$ be the unit disc. Its boundary is the unit circle, and so mapping under f we see that the image of γ is the boundary of Ω, where $\Omega = f(D)$.

Definition : A *(smooth) Jordan domain* is the image of a smooth disc under some diffeomorphism of $\mathbf{R}^2$.

Thus, the image any smooth Jordan curve, γ, is the boundary of a smooth Jordan domain, Ω. We can think of Ω as the "inside" of γ. We shall sometimes write informally: $\gamma = \partial\Omega$. (In fact, Ω is uniquely determined by γ.)

Note that γ may be positively or negatively oriented depending on whether the unit normal (as defined above) points in or out of Ω. In the former case, it has total turing 2π and in the latter case, it has total turning -2π (though we won't prove that here).

We can generalise this to a piecewise smooth closed curve. This has a sequence of exterior angles, $\theta_1, \ldots, \theta_n$, in $[-\pi, \pi]$ at the "vertices" where f is not differentiable. In this case, the total turning is defined as $\int_\gamma \theta' dt + \sum_i \theta_i$.

For example, for a polygonal curve, we have $\theta' = 0$ on all the straight segments, and so the total turning is just $\sum_i \theta_i$. Thus, for a positively oriented polygonal Jordan curve (i.e. a "polygon" in the traditional sense) we get the familiar formula: $\sum_i \theta_i = 2\pi$.

Area.

(We will assume the notion of integration of smooth functions over a "nice" subset, Ω of $\mathbf{R}^2$. Thus, for example, the area of Ω can be defined by intergating the constant function 1 over Ω. The formal definitions of these concepts requires some measure theory, beyond the scope of this course. See [Measure theory course ???].)

Suppose that Ω is a Jordan domain with boundary the smooth Jordan curve, γ. We write $\gamma(t) = (x(t), y(t))$. Let A be the area of Ω.

Lemma 2.3 :
$$A = -\int_\gamma y(t)x'(t)dt = \int_\gamma x(t)y'(t)dt = \frac{1}{2}\int_\gamma xy' - yx'dt.$$

Proof : See lectures. $\diamondsuit$

Example : Circle.

Remark : We have the following:

Theorem : *Let Ω be a Jordan domain, with boundary $\gamma = \partial\Omega$. Let $l = \text{length}(\gamma)$ and $A = \text{area}(\Omega)$. Then $A \leq l^2/4\pi$, with equality if and only if γ is a circle.*

A relatively simple proof can be found for example in [C, p33]. $\diamondsuit$

Green's Theorem.

Suppose that Ω is a Jordan domain, and $\gamma = \partial\Omega$ is positively oriented. Suppose that $P, Q : \mathbf{R}^2 \longrightarrow \mathbf{R}$ are smooth functions. (We only really need them defined in a neighbournood of Ω.) Then:

Theorem 2.4 : (Green):
$$\int_\gamma Px' + Qy'\,dt = \int_\Omega \frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\,dA.$$

Here $\gamma(t) = (x(t), y(t)) \in \mathbf{R}^2$, and $dA = dx\,dy$ is the area element.

Proof : See vector analysis course ???. $\diamondsuit$

Remark : This is sometimes expressed as follows. We can view $\mathbf{w} = (P, Q)$ as a vector field on $\mathbf{R}^2$, that is $\mathbf{w}(x, y) = (P(x, y), Q(x, y))$ is the vector based at (x, y). Embedding $\mathbf{R}^2 \hookrightarrow \mathbf{R}^3$ by $(x, y) \mapsto (x, y, 0)$, its curl is:
$$\operatorname{curl}\mathbf{w} = \left(0, 0, \frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right).$$
Thus, $\operatorname{curl}\mathbf{w}.d\mathbf{A} = \operatorname{curl}\mathbf{w}.\mathbf{k}\,dA = \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right)dt$. Also $\gamma'(t) = (x'(t), y'(t))$ so $Px' + Qy' = \mathbf{w}(\gamma(t)).\gamma'(t)$. Thus, Green's Theorem reduces to:
$$\int_\gamma \mathbf{w}(\gamma(t)).\gamma'(t)dt = \int_\Omega \operatorname{curl}\mathbf{w}.d\mathbf{A}$$
for any planar vector field $\mathbf{w}$. (This is Stokes's Theorem in the plane.)

Area again.

Set $P = -y$ and $Q = x$. Then $\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} = 2$. So

$$LHS = \int_\gamma -yx' + xy'\, dt$$

$$RHS = \int_\Omega 2\, dA = 2A.$$

Thus

$$A = \frac{1}{2}\int_\gamma xy' - yx' dt$$

as before.

3. Curves in $\mathbf{R}^3$.

Frenet frames.

Let $\gamma : I \longrightarrow \mathbf{R}^3$ be a smooth curve parameterised by arc length. Let $\mathbf{T}(s) = \gamma'(s)$ be the unit tangent vector. As in Section 2, $\mathbf{T}.\mathbf{T}' = 0$, so we can write $\mathbf{T}' = \kappa\mathbf{N}$, where $\mathbf{N}$ is a unit normal to γ and $\kappa(s) = \kappa_\gamma(s) = |\mathbf{T}'(s)| = |\gamma''(s)| \geq 0$. Here $\kappa \geq 0$ — there is no preferred sign. We refer to κ as the *curvature* of γ.

If $\kappa \neq 0$, then $\mathbf{N}$ is well defined. Let

$$\mathbf{B} = \mathbf{T} \wedge \mathbf{N}.$$

Thus, $\mathbf{T}, \mathbf{N}, \mathbf{B}$ is an orthonormal frame (based at $\gamma(s)$).

Definition : We refer to $\mathbf{T}$, $\mathbf{N}$ and $\mathbf{B}$ as the unit *tangent*, *normal* and *binormal* respectively, to γ.
We refer to $(\mathbf{T}, \mathbf{N}, \mathbf{B})$ as the *Frenet frame*.

(If there may be any confusion, we refer to $\mathbf{N}$ as the "principal normal" to γ — though this terminology is not standard.)

Note that, by definition of $\mathbf{B}$, the Frenet frame is positively oriented (so that there is an orientation preserving orthogonal linear map taking $\mathbf{T}, \mathbf{N}, \mathbf{B}$ to the standard basis, $(1, 0, 0), (0, 1, 0), (0, 0, 1)$ of $\mathbf{R}^3$).

Note that

$$\mathbf{N}'.\mathbf{N} = \mathbf{B}'.\mathbf{B} = 0.$$

Also:

$$\mathbf{B}' = (\mathbf{T} \wedge \mathbf{N})' = \mathbf{T}' \wedge \mathbf{N} + \mathbf{T} \wedge \mathbf{N}' = \mathbf{T} \wedge \mathbf{N}'.$$

Thus $\mathbf{B}'.\mathbf{T} = 0$. Since also $\mathbf{B}'.\mathbf{B} = 0$, we have

$$\mathbf{B}' = -\tau\mathbf{N}$$

for some $\tau \in \mathbf{R}$.

Definition : $\tau(s) = \tau_\gamma(s)$ is the *torsion* of γ at $\gamma(s)$.

This is defined when $\kappa(s) \neq 0$.

Remark : In some texts, $-\tau$ is used instead of τ.

The Frenet matrix.

Sometimes the following notation is convenient.
Let $\mathbf{T} = (T_1, T_2, T_3)$, $\mathbf{N} = (N_1, N_2, N_3)$, $\mathbf{B} = (B_1, B_2, B_3)$, and let P be the 3×3 matrix:

$$P = \begin{pmatrix} T_1 & T_2 & T_3 \\ N_1 & N_2 & N_3 \\ B_1 & B_2 & B_3 \end{pmatrix}.$$

Since $\mathbf{T}, \mathbf{N}, \mathbf{B}$ is orthonormal, $PP^T = I$ and so $P^T P = I$. Note that $P = P(s)$ depends on the parameter s. (Formally we can think of P as a curve $I \longrightarrow \mathbf{R}^9$, where the co-ordinates are the entries of the matrix.)

Now $\mathbf{T}'$, $\mathbf{N}'$ and $\mathbf{B}'$ can each be expressed as a linear combination of $\mathbf{T}$, $\mathbf{N}$ and $\mathbf{B}$. In other words, we can find a matrix $A(s)$ such that $\frac{dP}{ds}(s) = A(s)P(s)$, which we simply write as

$$P' = AP.$$

Now $PP^T = I$, so

$$\begin{aligned} 0 &= (PP^T)' \\ &= P'P^T + P(P^T)' \\ &= (AP)P^T + P(P')^T \\ &= APP^T + P(AP)^T \\ &= A + A^T. \end{aligned}$$

(We have used the general matrix formula: $(XY)' = X'Y + XY'$.) Thus, $A^T = -A$, i.e. A is antisymmetric.

By the definition of κ and τ, we know that A has the form:

$$\begin{pmatrix} 0 & \kappa & 0 \\ * & * & * \\ 0 & -\tau & 0 \end{pmatrix}.$$

Since $A + A^T = 0$, we must have:

$$A = \begin{pmatrix} 0 & \kappa & 0 \\ -\kappa & 0 & \tau \\ 0 & -\tau & 0 \end{pmatrix}.$$

From the middle row, we deduce the “Frenet-Serret (or Serret-Frenet) formula”:

Theorem 3.1 :

$$\frac{d\mathbf{N}}{ds} = -\kappa\mathbf{T} + \tau\mathbf{B}.$$

$\diamondsuit$

We also note that κ and τ are invariant under rigid motion. More precisely:

Proposition 3.2 : *If $\Phi : \mathbf{R}^3 \longrightarrow \mathbf{R}^3$ is a rigid motion, and $\gamma : I \longrightarrow \mathbf{R}^3$ is a curve, then for all s, $\kappa_\gamma(s) = \kappa_\beta(s)$ and $\tau_\gamma(s) = \pm\tau_\beta(s)$, where $\beta = \Phi \circ \gamma$, and the $\pm$ sign depends on whether Φ is orientation preserving or reversing.*

Proof : Exercise. $\diamondsuit$

Examples :

(1) Straight line: $\kappa \equiv 0$. The normal, binormal and torsion are undefined.

(2) Planar curve: That is there is a fixed vector, $\mathbf{a}$, with $\gamma(t).a$ constant. Then $\mathbf{T}.\mathbf{a} = 0$. We get $\mathbf{T}' = \kappa\mathbf{N}$ with $\mathbf{N}$ parallel to the plane and so $\mathbf{B} = \mathbf{T} \wedge \mathbf{N}$ is constant and ortogonal to the plane. We get $\tau \equiv 0$.

(3) Helix: Let

$$\gamma(s) = (a\cos\omega s, a\sin\omega s, b\omega s).$$

With $a > 0$ and b constant. We verify that $||\gamma'(s)|| = \omega\sqrt{a^2+b^2}$, so we set $\omega = 1/\sqrt{a^2+b^2}$ to get arc length.

Exercise:

$$\kappa = a\omega^2 \qquad \tau = -b\omega^2.$$

In particular, the curvature and torsion are constant.

Exercise: We can realise any constant positive curvature and any torsion with some helix.

Proposition 3.3 : *Let γ be a curve in $\mathbf{R}^3$.*

(1) If $\kappa \equiv 0$, then γ is a straight line.

(2) If $\kappa(s) \neq 0$ for all s, and $\tau \equiv 0$, then γ is planar.

Proof : (1) $\mathbf{T} = \gamma'$ is constant.

(2) $\mathbf{B}' = -\tau\mathbf{N}$, so $\mathbf{B}$ is constant. Thus, $\frac{d}{ds}(\gamma.\mathbf{B}) = \gamma'.\mathbf{B} = \mathbf{T}.\mathbf{B} = 0$, so $\gamma.\mathbf{B}$ is constant. $\diamondsuit$

We shall see later that if $\kappa \neq 0$ and τ are constant, then γ is a helix (up to rigid motion).

Linear systems of differential equations.

Before proceeding, we quote the following result regarding solutions of linear differential equations.

In what follows, $W : I \longrightarrow \mathbf{R}^{n^2}$ is formally a smooth curve in $\mathbf{R}^{n^2}$, though we think of it as a smooth family of matrices. That is, for any t, we think of $W(t)$ as the $n \times n$ matrix whose entries, $(w_{ij}(t))_{ij}$, are the coordinates in $\mathbf{R}^{n^2}$.

Theorem 3.4 : *Suppose $I \subseteq \mathbf{R}$ is an interval, and $t_0 \in I$ is any element of I. Suppose that $W : I \longrightarrow \mathbf{R}^{n^2}$ is a smooth family of $n \times n$ matrices. Suppose $\mathbf{p} \in \mathbf{R}^n$ is fixed. Then there is unique smooth path, $\mathbf{x} : I \longrightarrow \mathbf{R}^n$, such that $\mathbf{x}(t_0) = \mathbf{p}$, and such that for all t we have*

$$\mathbf{x}'(t) = W(t)\mathbf{x}(t).$$

$\diamondsuit$

Here we are writing $\mathbf{x}(t)$ as a collumn vector:

$$\begin{pmatrix} x_1(t) \\ \vdots \\ x_n(t) \end{pmatrix}.$$

In other words we have a system of n linear first order differential equations, and the theorem tells us that it has a unique solution.

For a proof, see ???.

We shall apply this to matrices in the following form:

Corollary 3.5 : *Suppose that $t \mapsto A(t)$ is a smooth family of $n \times n$ matrices, for $t \in I$. Suppose that $t_0 \in I$ and that P_0 is a fixed $n \times n$ matrix. Then there is a unique family of $n \times n$ matrices, $t \mapsto P(t)$, with $P(t_0) = P_0$ and with*

$$P'(t) = A(t)P(t)$$

for all t.

Proof : This follows by simply observing that $P'(t) = A(t)P(t)$ is a system of n^2 linear differential equations in the matrix entries. More formally, to relate this to the above, write $P(t)$ as a column vector $\mathbf{p}(t)$ with entries

$$p_{11}(t), \ldots, p_{1n}(t), p_{21}(t), \ldots, p_{nn}(t)$$

where $p_{ij}(t)$ are the entries in $P(t)$. Let $W(t)$ be the $n^2 \times n^2$ matrix, with sequence of n copies of $A(t)$ down the diagonal, and 0 everywhere else. The equation $P'(t) = A(t)P(t)$ is then the same as $\mathbf{p}'(t) = W(t)\mathbf{p}(t)$. $\diamondsuit$

Another variation on this will be used in the proof of Theorem 3.6 below.

Existence and uniqueness of curves with prescribed curvature and torsion.

Up to rigid motion in $\mathbf{R}^3$ there always exists a unique curve with specified (positive) curvature and torsion. More precisely:

Theorem 3.6 : *Let $I \subseteq \mathbf{R}$ be any interval, and let $\kappa : I \longrightarrow (0, \infty)$ and $\tau : I \longrightarrow \mathbf{R}$ be smooth functions. Then there is a smooth curve $\gamma : I \longrightarrow \mathbf{R}^3$ parameterised by arc length, such that for all $s \in I$, $\kappa_\gamma(s) = \kappa(s)$ and $\tau_\gamma(s) = \tau(s)$. Moreover, if $\beta : I \longrightarrow \mathbf{R}^3$ is another such curve, then there is an (orientation preserving) rigid motion, $\Phi : \mathbf{R}^3 \longrightarrow \mathbf{R}^3$ such that $\beta = \Phi \circ \gamma$.*

Proof : Let $s_0 \in I$ be any point. Recall that curvature and torsion are invariant under orientation preserving rigid motion. Moreover, up to rigid motion, we can assume that $\gamma(s_0) = (0, 0, 0)$ and the Frenet frame $\mathbf{T}(s_0), \mathbf{N}(s_0), \mathbf{B}(s_0)$ is just $(1, 0, 0), (0, 1, 0), (0, 0, 1)$. Thus we claim there is a unique curve γ is with the specifed curvature and torsion and with this initial condition.

Given $s \in I$ we set

$$A(s) = \begin{pmatrix} 0 & \kappa(s) & 0 \\ -\kappa(s) & 0 & \tau(s) \\ 0 & -\tau(s) & 0 \end{pmatrix}.$$

Thus $A : I \longrightarrow A(s)$ is smooth and $A(s) + A(s)^T = 0$. Now by Corollary 3.5, there is a unique family of matrices, $s \mapsto P(s)$ such that $P(s_0) = I$ and such that $P'(s) = A(s)P(s)$ for all s.

We claim that, for all s, $P(s)$ is an orthogonal matrix, i.e. $P(s)P(s)^T = I$. To see this, set $Q(s) = P(s)P(s)^T$. Thus, $Q(s_0) = I$. Moreover,

$$\begin{aligned} Q(s)' &= P(s)'P(s)^T + P(s)(P(s)^T)' \\ &= A(s)P(s)P(s)^T + P(s)(A(s)P(s))^T. \\ &= A(s)Q(s) + Q(s)A(s)^T \end{aligned}$$

I.e.

$$Q(s)' = A(s)Q(s) + Q(s)A(s)^T \qquad (*).$$

Note that $Q(s) = I$ for all s is one solution to $(*)$ (since $A(s) + A(s)^T = 0$). But $(*)$ is a system of n^2 first order linear differential equations in the matrix entries. Thus $(*)$ has a unique solution (cf. the proof of Corollary 3.5). It follows that we must indeed have $Q(s) = I$ for all s. That is $P(s)P(s)^T = I$, and so also $P(s)^T P(s) = I$.

Now, let $\mathbf{T}(s)$, $\mathbf{N}(s)$ and $\mathbf{B}(s)$ be respectively the first second and third rows of $P(s)$, thought of as vectors in $\mathbf{R}^3$. Since $P(s)$ is orthogonal, $\{\mathbf{T}(s), \mathbf{N}(s), \mathbf{B}(s)\}$ is an orthogonal frame.

Now, let $\gamma(s) = \int_{s_0}^s \mathbf{T}(u)du$. Thus $\gamma : I \longrightarrow \mathbf{R}^3$ is a smooth curve, with $\gamma(s_0) = 0$, and tangent $\gamma'(s) = \mathbf{T}(s)$. In particular, s is an arc length parameter.

Using $P(s)' = A(s)P(s)$, we get $\mathbf{T}'(s) = \kappa(s)\mathbf{N}(s)$ and $\mathbf{B}'(s) = -\tau(s)\mathbf{N}(s)$. Directly from their definitions, we see that $\mathbf{N}(s)$ and $\mathbf{B}(s)$ are the normal and binormal respectively, and that $\kappa(s)$ and $\tau(s)$ are its curvature and torsion respectively.

This proves the existence of γ.

For uniqueness, suppose that β is another curve with the same curvature and torsion. As observed above, we can assume that $\beta(s_0) = 0$ and that γ and β have the same Frenet frame at s_0. Let $\mathbf{T}_\beta(s)$, $\mathbf{N}_\beta(s)$, $\mathbf{B}_\beta(s)$ be the Frenet frame to β at s. Let $P_\beta(s)$ be the 3×3 matrix formed by arranging $\mathbf{T}_\beta(s)$, $\mathbf{N}_\beta(s)$, $\mathbf{B}_\beta(s)$ in three rows (as in the definition of Frenet matrix). Thus, $P_\beta(s)' = A(s)P_\beta(s)$ for all s. By the uniqueness of such a solution (Corollary 3.5), we get $P_\beta(s) = P(s)$ for all s. In particular, the first row is $\mathbf{T}_\beta(s) = \mathbf{T}(s)$. Thus,

$$\beta(s) = \int_{s_0}^{s} \mathbf{T}_\beta(u)du = \int_{s_0}^{s} \mathbf{T}(u)du = \gamma(s).$$

Thus $\beta = \gamma$ as required. $\diamondsuit$

One immediate consequence is:

Proposition 3.7 : *Suppose $\gamma : \mathbf{R} \longrightarrow \mathbf{R}^3$ is a unit speed curve with constant non-zero curvature and torsion. Then γ is a helix (up to rigid motion of $\mathbf{R}^3$).*

Proof : Let $\kappa_0 > 0$ and $\tau_0 \neq 0$ be the curvature and torsion. When discussing helices (lectures) we saw that we can always find a helix β with $\kappa_\beta \equiv \kappa_0$ and $\tau_\beta = \tau_0$. Thus, by Theorem 3.6, there is a rigid motion, Φ, of $\mathbf{R}^3$ with $\gamma = \Phi \circ \beta$. $\diamondsuit$

What if γ is not of unit speed?

Proposition 3.8 : *Suppose that $\gamma : I \longrightarrow \mathbf{R}^3$ is a regular curve. Then the curvature and torsion are given respectively by*

$$\kappa = \frac{|\gamma' \wedge \gamma''|}{|\gamma'|^3}$$

and

$$\tau = \frac{(\gamma' \wedge \gamma'').\gamma'''}{|\gamma' \wedge \gamma''|^2}.$$

Proof : Exercise. Or see [M p.87]. $\diamondsuit$

4. Parameterised surfaces in $\mathbf{R}^3$.

We now begin the study of 2-dimensional surfaces in 3-space.

Let $U \subseteq \mathbf{R}^2$ be open, and let $\mathbf{r} : U \longrightarrow \mathbf{R}^3$ be smooth, i.e. it has derivatives of all orders. We write $\mathbf{r}_u = \partial\mathbf{r}/\partial u$ and $\mathbf{r}_v = \partial\mathbf{r}/\partial v$, where u, v are the coordinates of $\mathbf{R}^2$. Thus, $\mathbf{r}_u : U \longrightarrow \mathbf{R}^3$ and $\mathbf{r}_v : U \longrightarrow \mathbf{R}^3$ are smooth.

Definition : We say that $\mathbf{r} : U \longrightarrow \mathbf{R}^3$ is a *parameterised surface* if, for all $(u, v) \in U$, $\mathbf{r}_u(u, v)$ and $\mathbf{r}_v(u, v)$ are linearly independent.

Recall that the linear map $[(\lambda, \mu) \mapsto \lambda\mathbf{r}_u + \mu\mathbf{r}_v] : \mathbf{R}^2 \longrightarrow \mathbf{R}^3$ is the *derivative* of $\mathbf{r}$. We denote it by $\mathbf{r}_*$. That is, $\mathbf{r}_*(\lambda, \mu) = \lambda\mathbf{r}_u + \mu\mathbf{r}_v$. Thus, the definition says that the derivative has rank 2 everywhere. With respect to the standard co-ordinates (u, v) on $\mathbf{R}^2$ and (x, y, z) on $\mathbf{R}^3$, the derivative has matrix

$$\begin{pmatrix} x_u & x_v \\ y_u & y_v \\ z_u & z_v \end{pmatrix}$$

where $x_u = \partial x/\partial u$ etc. This is the *jacobian* matrix.

Suppose, for the moment, that $\mathbf{r}$ is injective. We write $S = \mathbf{r}(U) \subseteq \mathbf{R}^3$. Given $(u, v) \in U$, write $T_p(S) \subseteq \mathbf{R}^3$ for the vector space spanned by $\mathbf{r}_u(u, v)$ and $\mathbf{r}_v(u, v)$, where $p = \mathbf{r}(u, v)$. This is the same as the image of the linear map $\mathbf{r}_*(u, v)$. By assumption it is 2-dimensional.

Definition : $T_p(S)$ is the *tangent space* to S at p.
The plane $p + T_p(S) \subseteq \mathbf{R}^3$ is the *tangent plane* to S at p.
The unit vector

$$\mathbf{n} = \mathbf{n}(u, v) = \frac{\mathbf{r}_u \wedge \mathbf{r}_v}{||\mathbf{r}_u \wedge \mathbf{r}_v||}$$

is the *normal* to S at p.

Less formally we can think of the tangent plane as "touching" S at the point p. Note that the parameterisation determines a direction of $\mathbf{n}$. If we were to replace v by $-v$, for example, we would replace $\mathbf{n}$ by $-\mathbf{n}$. We can think of this as changing the orientation on S.

We can generalise the above to the case where $\mathbf{r}$ is not assumed injective. But if there are two or more points mapped to p, then we need to specify which of the preimages is being used to define the tangent space.

We remark that, by Theorem 0.1, $\mathbf{r}$ will always be locally injective. That is, if $(u, v) \in U$, then there is always an open neighbourhood, V, of (u, v) in U such that $\mathbf{r}|V$ is injective. Thus, for the purpose of studying local properties, there is no loss in assuming injectivity. We say more about this in Section 5.

Examples :

(1) Plane. $U = \mathbf{R}^2$.

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

where $\mathbf{a}, \mathbf{b}, \mathbf{c}$ are fixed and $\mathbf{b} \wedge \mathbf{c} \neq 0$.

$$\mathbf{n} = (\mathbf{b} \wedge \mathbf{c})/|\mathbf{b} \wedge \mathbf{c}|.$$

(2) Cylinder. $U = \mathbf{R}^2$.

$$\mathbf{r}(u,v) = (a\cos v, a\sin v, u)$$
$$\mathbf{n} = (\cos v, \sin v, 0)$$

(3) More generally, we have a "product surface": Let $\gamma : I \longrightarrow \mathbf{R}^2$ be a curve. Write $\gamma(v) = (x(v), y(v))$. Set $U = I \times \mathbf{R}$ and set

$$\mathbf{r}(u,v) = (x(v), y(v), u)$$

.

(4) Cone. $U = (0,\infty) \times \mathbf{R}$.

$$\mathbf{r}(u,v) = (au\cos v, au\sin v, u)$$

(5) Sphere (minus poles) in spherical polar co-ordinates: $U = (-\pi/2, \pi/2) \times \mathbf{R}$.

$$\mathbf{r}(u,v) = (a\cos v\cos u, a\sin v\cos u, a\sin u)$$

Here, $\mathbf{n} = -\frac{1}{a}\mathbf{r}$.
(It is customary to use the notation $\theta = v$, $\phi = u$.)

(6) Helicoid. $U = \mathbf{R}^2$.

$$\mathbf{r}(u,v) = (au\cos v, au\sin v, v)$$

(7) Torus. $U = \mathbf{R}^2$, $a > b$ are constants.

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

(Note $\mathbf{r}$ is "doubly periodic": the plane wraps around the torus in both the u and v direcstions.)

(8) Surface of revolution. Suppose $I \subseteq \mathbf{R}$ is open and $U = I \times \mathbf{R}$, and that $\gamma : I \longrightarrow \mathbf{R}^2$ is a planar curve. Let $\gamma(u) = (\lambda(u), \mu(u))$ and suppose $\lambda(u) > 0$ for all u. Set

$$\mathbf{r}(u,v) = (\lambda(u)\cos v, \lambda(u)\sin v, \mu(u)).$$

Examples are the cylinder, cone, sphere and torus above.
Another, the "pseudosphere" obtained by spinning the tractrix will be discussed later in Section 8.
A third example, the "catenoid" is obtained from the catenery: $\lambda(u) = a\cosh u$, $\mu(u) = u$.

(9) Graph of a function. Suppose $f : U \longrightarrow \mathbf{R}$ is smooth. Let $\mathbf{r}(u,v) = (u, v, f(u,v))$.
e.g. paraboloid $f(u,v) = au^2 + bv^2$.

(10) Developable surface
Let γ be a unit speed curve in $\mathbf{R}^3$. Let

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

where $\mathbf{T}(u) = \gamma'(u)$. If $\kappa_\gamma(u) \neq 0$, write $\mathbf{N}(u)$, $\mathbf{B}(u)$ for the normal and binormal to γ respectively. Note that $\mathbf{r}_u = \mathbf{T} + v\mathbf{T}' = \mathbf{T} + v\kappa\mathbf{N}$ and $\mathbf{r}_v = \mathbf{T}$. If $\kappa \neq 0$, then $\mathbf{r}_u \wedge \mathbf{r}_v = v\kappa\mathbf{N} \wedge \mathbf{T} = -v\kappa\mathbf{B}$, so we get a parameterised surface for $v \neq 0$.

Reparameterisation.

Suppose $U, V \subseteq \mathbf{R}^2$ are open.

Definition : A *diffeomorphism* between V and U is a map $\sigma : V \longrightarrow U$ such that both σ and σ^{-1} are smooth.

Writing $\sigma(\eta, \theta) = (u(\eta, \theta), v(\eta, \theta))$ for $\eta, \theta \in V$, we have Jacobian

$$\begin{pmatrix} u_\eta & u_\theta \\ v_\eta & v_\theta \end{pmatrix}$$

invertible for all η, θ.

Suppose $\mathbf{r} : U \longrightarrow \mathbf{R}^3$ is a parameterised surface, and write $S = \mathbf{r}(U)$.

Definition : A *reparameterisation* of S is a composition $\mathbf{s} = \mathbf{r} \circ \sigma : V \longrightarrow \mathbf{R}^3$, where $\sigma : V \longrightarrow U$ is a diffeomorphism.

Note that $S = \mathbf{s}(V) = \mathbf{r}(U)$. We have

$$\begin{aligned} \mathbf{s}_\eta &= u_\eta \mathbf{r}_u + v_\eta \mathbf{r}_v \\ \mathbf{s}_\theta &= u_\theta \mathbf{r}_u + v_\theta \mathbf{r}_v \end{aligned}.$$

Thus $\mathbf{s}_\eta$ and $\mathbf{s}_\theta$ are linearly independent. (So $\mathbf{s} : V \longrightarrow \mathbf{R}^3$ is indeed a parametrised surface.) If $(u, v) = \sigma(\eta, \theta)$, we see that $\mathbf{s}_\eta(\eta, \theta)$ and $\mathbf{s}_\theta(\eta, \theta)$ span the same subspace of $\mathbf{R}^3$ as $\mathbf{r}_u(u, v)$ and $\mathbf{r}_v(u, v)$. Thus, the tangent space, $T_p(S)$ for $p = \mathbf{r}(u, v) = \mathbf{s}(\eta, \theta)$ is well defined, independently of such reparameterisation. (We say more about this in Section 5.)

We say that σ preserves or reverses orientation, depending on whether

$$\det \begin{pmatrix} u_\eta & u_\theta \\ v_\eta & v_\theta \end{pmatrix}$$

is positive or negative. The normal, $\mathbf{n}$, to S at p is then fixed or reversed accordingly.

Area.

Suppose that $D \subseteq S = \mathbf{r}(U)$ and let $\Delta = \mathbf{r}^{-1} D \subseteq U$. (We will assume that $\mathbf{r}$ is injective here.) We define the area of D by

$$\operatorname{area}(D) = \int_\Delta ||\mathbf{r}_u \wedge \mathbf{r}_v|| \, du dv.$$

For this, we need to assume that Δ is sufficiently "nice" for this integral to be defined. The technical term is "measurable" — see the measure theory course. However, it is sufficent that Δ should be open or closed, or equivalently that D be open or closed in S. This should be enough for our purposes.

Lemma 4.1 : *Area is invariant under reparameterisation.*

Proof : In other words, suppose that $\sigma : V \longrightarrow U$ is a diffeomorphism, and let $\mathbf{s} = \mathbf{r} \circ \sigma : V \longrightarrow \mathbf{R}^2$. Let $\Delta' = \sigma^{-1}\Delta = \mathbf{s}^{-1}D$. Then

$$\begin{aligned}\mathbf{s}_\eta \wedge \mathbf{s}_\theta &= (u_\eta \mathbf{r}_u + v_\eta \mathbf{r}_v) \wedge (u_\theta \mathbf{r}_u + v_\theta \mathbf{r}_v) \\ &= (u_\eta v_\theta - v_\eta u_\theta)\mathbf{r}_u \wedge \mathbf{r}_v,\end{aligned}$$

so

$$\begin{aligned}\int_{\Delta'} ||\mathbf{s}_\eta \wedge \mathbf{s}_\theta||\, d\eta d\theta &= \int_{\Delta'} ||\mathbf{r}_u \wedge \mathbf{r}_v||(u_\eta v_\theta - v_\eta u_\theta)\, d\eta d\theta \\ &= \int_{\Delta} ||\mathbf{r}_u \wedge \mathbf{r}_v||\, du dv\end{aligned},$$

since

$$u_\eta v_\theta - v_\eta u_\theta = \det \begin{pmatrix} u_\eta & u_\theta \\ v_\eta & v_\theta \end{pmatrix}.$$

$\diamondsuit$

The first fundamental form.

We introduce the "first fundamental form" of a parameterised surface. The idea behind it to describe the "intrinsic" geometry of the surface. First we recall some linear algebra.

Suppose that W is a vector space over $\mathbf{R}$. A *bilinear form* on W is a map of the form $\mathbf{x}, \mathbf{y} \mapsto \mathbf{x}^T P \mathbf{y}$, where $\mathbf{x}, \mathbf{y}$ are elements of W written as column vectors with respect to some basis of W, and P is a square matrix. (One should check that this notion is independent of the basis we have chosen for W.) The corresponding map $\mathbf{x} \mapsto \mathbf{x}^T P \mathbf{x}$ is a *quadratic form* on W. Such a form is *positive definite* if $\mathbf{x}^T P \mathbf{x} > 0$ for all $\mathbf{x} \neq 0$.

Example : If $W = \mathbf{R}^n$, then the inner product, $\mathbf{x}.\mathbf{y} = \mathbf{x}^T \mathbf{y} = \mathbf{x}^T I \mathbf{y}$ is a positive definite bilinear form. The square of the norm, $||\mathbf{x}||^2 = \mathbf{x}.\mathbf{x}$ is the corresponding quadratic from. If $X \subseteq W$ is any subspace, then this induces a form on X, simply by restriction.

Suppose that $U \longrightarrow \mathbf{R}^3$ is a parameterised surface, and $p = \mathbf{r}(u, v) \in \mathbf{r}(U) = S$.

Definition : The *first fundamental form* on the tangent space $T_p(S)$ is the quadratic form induced by the inner product on $\mathbf{R}^3$.

Clearly this is positive definite. We want to express this in terms of the intrinsic coordinates u, v of S.

Suppose $(\lambda_1, \mu_1), (\lambda_2, \mu_2) \in \mathbf{R}^2$. Then $\mathbf{x} = \lambda_1 \mathbf{r}_u + \mu_1 \mathbf{r}_v$ and $\mathbf{y} = \lambda_2 \mathbf{r}_u + \mu_2 \mathbf{r}_v$ lie in $T_p(S)$. We see that:

$$\mathbf{x}.\mathbf{y} = \begin{pmatrix} \lambda_1 & \mu_1 \end{pmatrix} \begin{pmatrix} E & F \\ F & G \end{pmatrix} \begin{pmatrix} \lambda_2 \\ \mu_2 \end{pmatrix}$$

where

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

Note that $\mathbf{x} = \mathbf{r}_*(\lambda_1, \mu_1)$ and $\mathbf{y} = \mathbf{r}_*(\lambda_2, \mu_2)$. Thus, under the identification of $\mathbf{R}^2$ with $T_p(S)$ via the derivative map, $\mathbf{r}_*$, the first fundamental form is the bilinear form with matrix:

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

Note that $E, G > 0$ and that

$$\det \begin{pmatrix} E & F \\ F & G \end{pmatrix} = EG - F^2 = ||\mathbf{r}_u||^2||\mathbf{r}_v||^2 - |\mathbf{r}_u.\mathbf{r}_v|^2 = ||\mathbf{r}_u \wedge \mathbf{r}_v||^2 > 0.$$

(From basic linear algebra, this gives another proof that the first fundamental form is positive definite.)

The first fundamental form is commonly written in the notation:

$$Edu^2 + 2Fdudv + Gdv^2.$$

One justification for this is as follows. Suppose $\sigma : V \longrightarrow U$ is a diffeomorphism, and $\mathbf{s} = \mathbf{r} \circ \sigma$. So $\mathbf{s}_\eta = u_\eta \mathbf{r}_u + v_\eta \mathbf{r}_v$ and $\mathbf{s}_\theta = u_\theta \mathbf{r}_u + v_\theta \mathbf{r}_v$. We check that in the new coordinates, the first fundamental form is given by the matrix

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

where:

$$\begin{aligned} \bar{E} &= \mathbf{s}_\eta.\mathbf{s}_\eta = Eu_\eta^2 + 2Fu_\eta v_\eta + Gv_\eta^2 \\ \bar{F} &= \mathbf{s}_\eta.\mathbf{s}_\theta = Eu_\eta u_\theta + F(u_\eta v_\theta + u_\theta v_\eta) + Gv_\eta v_\theta\ . \\ \bar{G} &= \mathbf{s}_\theta.\mathbf{s}_\theta = Eu_\theta^2 + 2Fu_\theta v_\theta + Gv_\theta^2 \end{aligned}$$

A more convenient way of carrying out the above calculation is to write "$du = u_\eta d\eta + u_\theta d\theta$" and "$dv = v_\eta d\eta + v_\theta d\theta$" and then expand:

$$Edu^2 + 2Fdudv + Gdv^2$$

$$= E(u_\eta d\eta + v_\theta d\theta)^2 + 2F(u_\eta d\eta + u_\theta d\theta)(v_\eta d\eta + v_\theta d\theta) + G(u_\eta d\eta + v_\theta d\theta)^2$$
$$= \bar{E}d\eta^2 + 2\bar{F}d\eta d\theta + \bar{G}d\theta^2.$$

The above should be viewed as a formal manipulation, with no particular intrinsic meaning attached to "du" or "dv" etc. Luckily for us, it gives the correct answer.

Definition : We say that a property of S is *intrinsic* if it can be expressed in terms of the first fundamental form.

(That is, the first fundamental form as a function on S — we are allowed to take derivatives of the form with respect to u and v etc.)

For example area. Suppose that $D \subseteq S$ is "nice". Then

$$\text{area}(D) = \int_D ||\mathbf{r}_u \wedge \mathbf{r}_v||\, dudv = \int_D \sqrt{EG - F^2}\, dudv$$

by the formula for $||\mathbf{r}_u \wedge \mathbf{r}_v||$ given earlier.

Another example is length. Suppose that $\xi : I \longrightarrow U$ is a smooth curve in $U \subseteq \mathbf{R}^2$. Then $\gamma = \mathbf{r} \circ \xi : I \longrightarrow S \subseteq \mathbf{R}^3$. Writing $\xi(t) = (u(t), v(t))$ its length is given by

$$\int_I ||\gamma'(t)||\, dt = \int_I ||\mathbf{r}_u u' + \mathbf{r}_v v'||\, dt = \int_I \sqrt{E(u')^2 + 2Fu'v' + G(v')^2}\, dt.$$

(This also gives another justification of the notation $Edu^2 + 2Fdudv + Gdv^2$.)

A third example are angles: Suppose that $\zeta, \xi : I \longrightarrow U \subseteq \mathbf{R}^2$ are regular curves, and $\zeta(t_0) = \xi(t_1) = q$ say. Let $\beta = \mathbf{r} \circ \zeta$ and $\gamma = \mathbf{r} \circ \xi$. The tangent vectors are $\beta'(t_0)$ and $\gamma'(t_1)$, so they meet at q at an angle

$$\cos^{-1}\left(\frac{\beta'(t_0).\gamma'(t_1)}{||\beta'(t_0)||\, ||\gamma'(t_1)||}\right)$$

which can again be expressed in terms of the first fundamental form of S at p.

Examples :

(1) Cartesian coordinates (x, y) in the plane: $\mathbf{r}(x, y) = (x, y, 0)$. Then $\mathbf{r}_x = \mathbf{i}$, $\mathbf{r}_y = \mathbf{j}$. So $E = \mathbf{i}.\mathbf{i} = 1$, $F = \mathbf{i}.\mathbf{j} = 0$, $G = \mathbf{j}.\mathbf{j} = 1$. We get the first fundamental form to be:

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

(2) Polar coordinates (r, θ) in the plane (minus the origin). Let $\mathbf{r}(r, \theta) = (r\cos\theta, r\sin\theta, 0)$. Then $\mathbf{r}_r = (\cos\theta, \sin\theta, 0)$, $\mathbf{r}_\theta = (-r\sin\theta, r\cos\theta, 0)$, so $E = 1$, $F = 0$, $G = r^2$. We get f.f.f. equal to

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

Note that we can view this as a reparameterisation of the plane minus the origin. The formal manipulation discussed above, with $x = r\cos\theta$ and $y = r\sin\theta$ gives: $dx^2 + dy^2 = (\cos\theta dr - r\sin\theta d\theta)^2 + (\sin\theta dr + r\cos\theta d\theta)^2 = dr^2 + r^2 d\theta^2$.

(3) The cylinder $\mathbf{r}(u, v) = (a\cos v, a\sin v, u)$.

So $\mathbf{r}_u = (0,0,1)$, $\mathbf{r}_v = (-a\sin v, a\cos v, 0)$, and $E = \mathbf{r}_u.\mathbf{r}_u = 1$, $F = \mathbf{r}_u.\mathbf{r}_v = 0$, $G = \mathbf{r}_v.\mathbf{r}_v = a^2$. Thus the f.f.f. is:

$$du^2 + a^2dv^2.$$

(4) The cone: The f.f.f. is:

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

.

(5) The sphere of radius a in spherical polars (u, v). We get

$$a^2du^2 + a^2\cos^2 u dv^2.$$

(6) Surface of revolution: $\mathbf{r}(u,v) = (\lambda(u)\cos v, \lambda(u)\sin v, \mu(u))$.
We get:

$$(\lambda'(u)^2 + \mu'(u)^2)du^2 + \lambda(u)^2dv^2.$$

(7) Helicoid. We get

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

(8) Graphs. We get $\mathbf{r}_u = (1, 0, f_u)$, $\mathbf{r}_v = (0, 1, f_v)$, so the f.f.f. is:

$$(1+f_u^2)du^2 + 2f_uf_vdudv + (1+f_v^2)du^2.$$

(8) Developable surfaces. $\mathbf{r}_u = \mathbf{T} + v\kappa\mathbf{N}$, $\mathbf{r}_v = \mathbf{T}$. The f.f.f. is:

$$(1+v^2\kappa^2)du^2 + 2dudv + dv^2.$$

Example : As an example, we can compute the area of a surface of revolution. Suppose that I and $J = (\theta_0, \theta_1) \subseteq \mathbf{R}$ are intervals. Let $U = I \times J$, and let $\mathbf{r} : U \longrightarrow \mathbf{R}^3$ be given by $\mathbf{r}(u,v) = (\lambda(u)\cos v, \lambda(u)\sin v, \mu(u))$. From the formula for the first fundamental form above, we calculate

$$EG - F^2 = \lambda(u)^2(\lambda'(u)^2 + \mu'(u)^2)$$

and so the total area is

$$(\theta_1 - \theta_0)\int_I \lambda(u)\sqrt{(\lambda'(u))^2 + \mu'(u)^2}du.$$

If $\theta_0 = 0$ and $\theta_1 = 2\pi$, we get

$$2\pi\int_I \lambda(u)\sqrt{(\lambda'(u))^2 + \mu'(u)^2}du.$$

Isometries.

Suppose that $\mathbf{r} : U \longrightarrow \mathbf{R}^3$ and $\mathbf{s} : U \longrightarrow \mathbf{R}^3$ are two parameterised surfaces with the same domain, U. Let $S = \mathbf{r}(U)$ and $\Sigma = \mathbf{s}(U)$. If $(u, v) \in U$, we will refer to $p \in \mathbf{r}(u, v) \in S$ and $q = \mathbf{s}(u, v) \in \Sigma$ as *corresponding points*.

We also have invertible linear maps $\mathbf{r}_* : \mathbf{R}^2 \longrightarrow T_p(S)$ and $\mathbf{s}_* : \mathbf{R}^2 \longrightarrow T_q(\Sigma)$. Thus $\mathbf{s}_* \circ \mathbf{r}_*^{-1} : T_p(S) \longrightarrow T_q(\Sigma)$ is linear and invertible. We shall refer to the image of a vector in $T_p(S)$ under this map as the *corresponding vector* in $T_p(\Sigma)$. If $\xi : I \longrightarrow U$ is a smooth path, then we get paths $\beta = \mathbf{r} \circ \xi : I \longrightarrow S$ and $\gamma = \mathbf{s} \circ \xi : I \longrightarrow \Sigma$. Again we refer to these as *corresponding paths*. For any t, $\beta(t)$ and $\gamma(t)$ are corresponding points, and $\beta'(t)$ and $\gamma'(t)$ are corresponding vectors.

Definition : We say $\mathbf{r} : U \longrightarrow \mathbf{R}^2$ and $\mathbf{s} : U \longrightarrow \mathbf{R}^2$ are isometric parameterised surfaces if the first fundamental forms on U are equal.

Note that, by the definition of first fundamental form, this is the same as saying that at any pair of corresponding points, $p \in S$ and $q \in \Sigma$, the natural correspondence between $T_p(S)$ and $T_q(\Sigma)$ respects the dot products induced from $\mathbf{R}^3$.

Lemma 4.2 : *The parametrised surfaces, $S = \mathbf{r}(U)$ and $\Sigma = \mathbf{s}(U)$ are isometric if and only if all pairs of corresponding curves have the same length.*

Proof : We write the first fundamental forms of S and Σ respectively as

$$Edu^2 + 2Fdudv + Gdu^2$$

$$\bar{E}du^2 + 2\bar{F}dudv + \bar{G}du^2.$$

If β and γ are corresponding curves, their lengths are respectively:

$$\int_I \sqrt{E(u')^2 + 2Fu'v' + G(v')^2}\, dt$$

$$\int_I \sqrt{\bar{E}(u')^2 + 2\bar{F}u'v' + \bar{G}(v')^2}\, dt$$

If the first fundamental forms are equal, these lengths are equal.

Conversely, suppose these are equal for all β and γ. In particular, given $(u, v) \in U$ and $(\lambda, \mu) \in \mathbf{R}^2$, can find a curve $\xi : [-a, a] \longrightarrow U$ with $\xi(0) = (u, v)$ and $\xi'(0) = (\lambda, \mu)$ (e.g. a short straight line segment.) Set $\beta = \mathbf{r} \circ \xi$ and $\gamma = \mathbf{s} \circ \xi$. Now length$(\beta|[-a, t]) =$ length$(\gamma|[-a, t])$, and so the integrals over $[-a, t]$ are equal. Differentiating at $t = 0$, we get:

$$\sqrt{E(u')^2 + 2Fu'v' + G(v')^2} = \sqrt{\bar{E}(u')^2 + 2\bar{F}u'v' + \bar{G}(v')^2}.$$

with $(u', v') = (\lambda, \mu)$. Thus

$$E\lambda^2 + 2F\lambda\mu + G\mu^2 = \bar{E}\lambda^2 + 2\bar{F}\lambda\mu + \bar{G}\mu^2.$$

Since this holds for all $(\lambda, \mu) \in \mathbf{R}^2$ we deduce that $E = \bar{E}$, $F = \bar{F}$ and $G = \bar{G}$. Moreover this holds for any pair of corresponding points. $\diamondsuit$

If $\mathbf{r}$ and $\mathbf{s}$ are injective and isometric, then the map $\mathbf{s} \circ \mathbf{r}^{-1} : S \longrightarrow \Sigma$ is referred to as an *isometry*.

Note that an intrinsic property can equivalently be viewed as one that is invariant under isometry.

Examples :

(1) Rigid motion:

If $\mathbf{r} : U \longrightarrow \mathbf{R}^3$ is a parameterised surface, and $\Phi : \mathbf{R}^3 \longrightarrow \mathbf{R}^3$ is a rigid motion, the $\mathbf{r}$ and $\Phi \circ \mathbf{r}$ are isometric.

(2) Cylinder and plane.

Consider:

$$(u, v) \mapsto (a \cos v, a \sin v, u)$$

$$(u, v) \mapsto (u, av, 0)$$

We get first fundamental forms $du^2 + a^2 dv^2$ in the case of the cylinder, and $du^2 + (adv)^2 = du^2 + a^2 dv^2$ in the case of the plane. These are equal, so the cylinder is isometric to the plane (locally).

This accords with the experience that one can bend a flat piece of paper into a cylinder without tearing it — their intrinsic geometries are the same.

(3) Cone and plane. Similar.

(4) Catenoid and helicoid. Consider the catenoid:

$$\mathbf{r}(u, v) = (\cosh u \cos v, \cosh u \sin v, u)$$

and helicoid:

$$(\bar{u}, v) \mapsto (\bar{u} \cos v, \bar{u} \sin v, v).$$

We begin by reparametrising the helicoid, by setting $\bar{u} = \sinh u$, and leaving v unchanged, so that it becomes:

$$\mathbf{s}(u, v) = (\sinh u \cos v, \sinh u \sin v, v).$$

We see that the f.f.f.s of both are given by:

$$\cosh^2 u (du^2 + dv^2)$$

so they are isometric.

Note we are thinking of each as a map from $\mathbf{R}^2$ into $\mathbf{R}^3$. The helicoid is injective, whereas the catenoid is periodic in the v direction, so the maps into $\mathbf{R}^3$ are quite different.

In fact, it's possible to find a smooth family of isometric surfaces which "morph" a catenoid to a helicoid. See example sheet. There are several animations of this on the web.

In dealing with isometries, it is often convenient to allow for reparameterisation at the same time. (In the above example, it would be more sensible to parameterise the plane with coordinates $x = u$ and $y = av$.)

Suppose that V, U are open, $\sigma : V \longrightarrow U$ is a diffeomorphism, and $\mathbf{r} : U \longrightarrow \mathbf{R}$, and $\mathbf{s} : V \longrightarrow \mathbf{R}$ are surfaces. Then $\mathbf{r} \circ \sigma$ and $\mathbf{s}$ are isometric if σ sends the first fundamental form of $\mathbf{s}$ on V to the first fundamental form of $\mathbf{r}$ on U.

Examples :

(1) In the above example (2), we could take

$$(u, v) \mapsto (a \cos v, a \sin v, u)$$

$$(x, y) \mapsto (x, y, 0)$$

and let $\sigma(u, v) = (u, av)$, i.e. $x = u$, $y = av$. This sends $dx^2 + dy^2$ to $du^2 + a^2 dv^2$, and so gives us an isometry.

(2) The cone and the plane.

These are respectively:

$$(u, v) \mapsto (au \cos v, au \sin v, u)$$

$$(r, \theta) \mapsto (r \cos \theta, r \sin \theta, 0)$$

(in polar coordinates). The f.f.f.s are:

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

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

Define $\sigma(u, v) = (r, \theta)$ where $r = u\sqrt{1 + a^2}$ and $\theta = av/\sqrt{1 + a^2}$. This gives an isometry of a subset of the cone to a subset of the plane.

(Exercise: describe suitable domains U and V so that σ is a diffeomorphism.)

(3) The catenoid and helicoid.

If we use the original parameterisation of the helicoid as $(u, v) \mapsto (u \cos v, u \sin v, v)$, then we can write $\sigma(u, v) = (\sinh u, v)$. This describes an isometry from the catenoid to the helicoid.

Conformal maps.

These are more general than isometries.

Definition : Suppose $\mathbf{r} : U \longrightarrow \mathbf{R}^3$ and $\mathbf{s} : U \longrightarrow \mathbf{R}^3$ are parameterised surfaces. Then $\mathbf{r}$ and $\mathbf{s}$ are *conformally equivalent* if there is a smooth function $\lambda : U \longrightarrow (0, \infty)$ such that for all $(u, v) \in U$, the first fundamental form of s at (u, v) is $\lambda(u, v)$ times that of $\mathbf{r}$ at (u, v).

In other words, it respect the first fundamental form pointwise up to scale. Note that angles depend on the first fundamental form up to scale. Thus, conformally equivalent surfaces respect angles between corresponding curves. (In fact, the converse is also true, so "conformal" is the same as "angle preserving".)

As with isometries, we can allow for reparametisation in dealing with conformal maps.

Examples :

(1) Any postcomposition with a dilation of $\mathbf{R}^3$ will be conformal.

(2) The plane and helicold or catenoid. Local scaling is $\cosh u$ (see above).

(3) Complex analytic maps are conformal. To intepret this in terms of surfaces in $\mathbf{R}^3$, we can identify the complex plane $\mathbf{C}$ with $\mathbf{R}^2$ and embed it in $\mathbf{R}^3$ by $(x, y) \mapsto (x, y, 0)$.

As an explicit example, set $V = \{(u, v) \mid u > 0\}$ and $U = \mathbf{R}^2 \setminus ((-\infty, 0] \times \{0\})$. Let $\sigma : V \longrightarrow U$ be the diffeomorphism that sends (u, v) to $(x, y) = (u^2 - v^2, 2uv)$. To compare first fundamental forms, we make the following formal manipulation:

$$dx = 2udu - 2vdv$$

$$dy = 2vdu + 2udv.$$

We get:

$$dx^2 + dy^2 = 4(u^2 + v^2)(du^2 + dv^2)$$

and so we set $\lambda(u, v) = 4(u^2 + v^2)$. We see that the map is conformal.

In terms of complex analysis, we have $x + iy = (u + iv)^2$, and so setting $w = u + iv$, the map σ is simply $[w \mapsto w^2]$.

(4) We will see other examples in Section 5 (stereographic projection, Mercator projection).

5. Embedded surfaces in $\mathbf{R}^3$.

In this section, we give a slightly different formulation of the notion of a surface, that does not require a particular parameterisation.

Recall that a "parameterised surface" was defined in terms of a map $\mathbf{r} : U \longrightarrow \mathbf{R}^3$.

This has the advantage that it allows the surface to cross itself, or to wrap around itself. (The latter is particularly useful in dealing with surfaces of revolution, for example.)

It has a number of disadvantages. It refers to a particular parametrisation. Also, many "surfaces" are not parametrised surfaces in this sense. With the sphere, for example, we saw that we had to remove the poles to make it conform naturally to this definition.

For this reason, we make the following definitions:

Definition : Suppose S is a subset of $\mathbf{R}^3$. By a *chart* for S, we mean a parameterised surface $\mathbf{r} : U \longrightarrow \mathbf{R}^3$ where $U \subseteq \mathbf{R}^2$ is open, such that there exists an open set $W \subseteq \mathbf{R}^3$ with $\mathbf{r}(U) = S \cap W$, and such that the inverse map $\mathbf{r}^{-1} : S \cap W \longrightarrow U$ is continuous.

We say that S is a *regular surface* if for each $p \in S$, there is some chart, $\mathbf{r} : U \longrightarrow S$ with $p \in \mathbf{r}(U)$.

We will refer to the image of a chart as a *coordinate patch*.

Thus, a regular surface is a subset of $\mathbf{R}^3$ that can be covered by such coordinate patches. These patches may overlap (and we will need to consider how the different coordinates behave on these overlaps). The collection of all coordinate charts is called an "atlas" for S.

The technical requirement of a chart that $\mathbf{r}^{-1}$ be continuous is required to rule out certain unwanted phenomena [picture]. One consequence of it will be that if $V \subseteq U$ is open, then $\mathbf{r}|V$ is also a chart (Lemma 5.1).

Examples :

(1) Let $\mathbf{r} : U \longrightarrow \mathbf{R}^3$ be a parameterised surface. Suppose $\mathbf{r}$ is injective and write $\Sigma = \mathbf{r}(U)$. Suppose also that the inverse map $\mathbf{r}^{-1} : \Sigma \longrightarrow U$ is continuous. Then Σ is a regular surface — we only need one chart, namely $\mathbf{r}$ itself.

Note: As a consequence, if $\mathbf{r} : U \longrightarrow S$ is a chart of a surface S, then then $\mathbf{r}(U)$ is itself a regular surface.

(2) The unit sphere (using steriographic projection).

Let $\mathbf{S}^2 = \{\mathbf{x} \mid ||\mathbf{x}|| = 1\} \subseteq \mathbf{R}^3$. (This is standard notation for the 2-sphere. It does not mean $\mathbf{S} \times \mathbf{S}$!) Let $n = (0, 0, 1) \in \mathbf{S}^2$ be the "north pole" and $s = (0, 0, -1) \in \mathbf{S}^2$ be the "south pole"

Define $\mathbf{r} : \mathbf{R}^2 \longrightarrow \mathbf{R}^3$ by the formula:

$$\mathbf{r}(x, y) = \left(\frac{2x}{x^2 + y^2 + 1}, \frac{2y}{x^2 + y^2 + 1}, \frac{x^2 + y^2 - 1}{x^2 + y^2 + 1} \right).$$

One can interpret this geometrically as follows. One can check that $P = \{(x, y, -1) \mid x, y \in \mathbf{R}\}$ is the tangent plane to the south pole (exercise). identify this with the plane $\mathbf{R}^2$ by $(x, y) \leftrightarrow (x, y, -1)$. The above map then sends a point $p \in P$ to the point $q \in \mathbf{S}^2$, where q is the intersection of $\mathbf{S}^2$ with the straight line in $\mathbf{R}^3$ connecting p to n. (Exercise). (The inverse of this map is called *steriographic projection*.)

One can now verify (exercise) that $\mathbf{r}$ us a chart of $\mathbf{S}^2$, with image $\mathbf{S}^2 \setminus \{n\}$.

Now swapping the roles of n and s, we get a chart with image $\mathbf{S}^2 \setminus \{s\}$. These cover $\mathbf{S}^2$, and so it is a regular surface. We have constructed an atlas with just two charts.

Exercise : Describe a simpler altas for the 2-sphere using 6 charts.

We begin with a couple of technical observations:

Lemma 5.1 : *If* $\mathbf{r} : U \longrightarrow S$ *is a chart, and* $V \subseteq U$ *is open, then* $\mathbf{r}|V : V \longrightarrow S$ *is a chart.*

Proof : We have $\mathbf{r}(U) = S \cap W$ for some open $W \subseteq \mathbf{R}^3$. The fact that $\mathbf{r}^{-1} : S \cap W \longrightarrow U$ is continuous implies that $\mathbf{r}(V)$ is open in $S \cap W$. This means that there is an open set $W' \subseteq W$ with $\mathbf{r}(V) = S \cap W'$. (These just use the definitions of subspace topology and continuity.) Now W' is also open in $\mathbf{R}^3$.

The properties of a chart follow easily. $\diamondsuit$

Lemma 5.2 : *Suppose* $S \subseteq \mathbf{R}^3$ *is a regular surface, and* $\mathbf{r} : U \longrightarrow S$ *is a chart (with* $\mathbf{r}(U) = W \cap S$*, with* $W \subseteq \mathbf{R}^3$ *open). Suppose that* $(u_0, v_0) \in U$ *and let* $p = \mathbf{r}(u_0, v_0) \in S$*. Then there are open sets* $U_0 \subseteq U$, $W_0 \subseteq W$*, with* $(u_0, v_0) \in U_0$, $p \in W_0$ *and a smooth map* $\mathbf{f} : W_0 \longrightarrow \mathbf{R}^2$ *with* $\mathbf{f}(W_0 \cap S) = U_0$*, and with* $\mathbf{f} \circ \mathbf{r} = 1_{U_0}$ *(the identity).*

In other words, we can find a smooth left inverse locally that is defined on a neighbourhood of S in $\mathbf{R}^3$.

Proof : Write $\mathbf{r}(u, v) = (x(u, v), y(u, v), z(u, v))$. The jacobian

$$\begin{pmatrix} x_u & x_v \\ y_u & y_v \\ z_u & z_v \end{pmatrix}$$

has rank 2 at (u_0, v_0). Thus (swapping the roles of y and z if necessary) we can assume that (x_u, x_v) and (y_u, y_v) are linearly independent, i.e.

$$\det \begin{pmatrix} x_u & x_v \\ y_u & y_v \end{pmatrix} \neq 0$$

Define a map $\mathbf{a} : U \times \mathbf{R} \longrightarrow \mathbf{R}^3$ by

$$\mathbf{a}(u, v, w) = \mathbf{r}(u, v) + (0, 0, w).$$

This has jacobian

$$\begin{pmatrix} x_u & x_v & 0 \\ y_u & y_v & 0 \\ z_u & z_v & 1 \end{pmatrix}$$

which is non-singular at $(u_0, v_0, 0)$. Thus, by the inverse function theorem (Theorem 0.2), there are neighbourhoods V_0 of $(u_0, v_0, 0)$ and W_0 of p in $\mathbf{R}^3$, with $\mathbf{a}(V_0) = W_0$, and with a smooth inverse map, $\mathbf{a}^{-1} : W_0 \longrightarrow V_0$. Let $\mathbf{f}$ be the composition of $\mathbf{a}^{-1}$ with the projection map $(u, v, w) \mapsto (u, v) : \mathbf{R}^3 \longrightarrow \mathbf{R}^2$. Since these are both smooth, $\mathbf{f}$ is smooth. Let $U_0 = \{(u, v) \in \mathbf{R}^2 \mid (u, v, 0) \in V_0\}$. This is open in $\mathbf{R}^2$ (exercise). If $(u, v) \in U_0$, then by construction, $\mathbf{a}^{-1} \circ \mathbf{r}(u, v) = \mathbf{a}^{-1} \circ \mathbf{a}(u, v, 0) = (u, v, 0)$, so $\mathbf{f} \circ \mathbf{r}(u, v) = (u, v)$, i.e. $\mathbf{f} \circ \mathbf{r}$ is the identity on U_0 as required. $\diamondsuit$

Transition maps.

Suppose that $\mathbf{r} : U \longrightarrow S$ and $\mathbf{s} : V \longrightarrow S$ are charts. Let $U_0 = \mathbf{r}^{-1} \circ \mathbf{s}(V)$ and $V_0 = \mathbf{s}^{-1} \circ \mathbf{r}(U)$. Thus, U_0 and V_0 are open subsets of U and V respectively. Moreover, by Lemma 5.1, $\mathbf{r}|U_0$ and $\mathbf{s}|V_0$ are charts. These subcharts have the same image in S (namely $\mathbf{r}(U) \cap \mathbf{s}(V)$, and so $\sigma = \mathbf{r}^{-1} \circ \mathbf{s} : V_0 \longrightarrow U_0$ is bijective. In fact it is a homeomorphism (i.e. it is continuous with continuous inverse).

Definition : The map $\sigma : V_0 \longrightarrow U_0$ is the *transition map* between these two charts.

(Note that we allow for the possibility that $U_0 = V_0 = \emptyset$, though there is not much to be said in that case.)

Lemma 5.3 : *A transition map is a diffeomorphism.*

Proof : In other words, we claim that $\sigma : V_0 \longrightarrow U_0$ and $\sigma^{-1} : U_0 \longrightarrow V_0$ are both smooth. Swapping the roles of $\mathbf{r}$ and $\mathbf{s}$, it's enough to prove the former statement.

Suppose that $(u_1, v_1) \in V_0$. Let $p = \mathbf{s}(u_1, v_1)$ and let $(u_0, v_0) = \mathbf{r}^{-1}(p) \in U_0$. Now Lemma 5.2 gives us a smooth inverse $\mathbf{f}$ of $\mathbf{r}$ defined on a neighbourhood of p in $\mathbf{R}^3$. Thus, for all (u, v) in a neighbourhood of (u_1, v_1), we have $\sigma(u, v) = \mathbf{r}^{-1} \circ \mathbf{s}(u, v) = \mathbf{f} \circ \mathbf{s}(u, v)$. In particular, $\sigma = \mathbf{f} \circ \mathbf{s}$ is smooth at $(u_1, v_1) \in V_0$. Since (u_1, v_1) was arbitrary, σ is smooth everywhere in V_0. $\diamondsuit$

Thus, we have $\mathbf{s} = \mathbf{r} \circ \sigma : V_0 \longrightarrow \mathbf{R}^3$ where σ is a diffeomorphism. In other words $\mathbf{s}|V_0$ is a reparameterisation of $\mathbf{r}|U_0$. We have shown that different charts differ by reparameterisation where they overlap. We can therefore bring the results of Section 4 into play.

Tangent spaces.

Suppose that $p \in S$. Let $\mathbf{r} : U \longrightarrow S$ be a chart with $p \in \mathbf{r}(U)$. This defines a tangent space $T_p(S)$. By Lemma 5.3, and the discussion in Section 4, we see that $T_p(S)$ is well defined, independently of the choice of chart.

Definition : $T_p(S)$ is the *tangent space* to S at p. We refer to $p + T_p(S)$ as the *tangent plane*.

Curves in S.

Definition : A *smooth curve* in S is a smooth curve, $\gamma : I \longrightarrow \mathbf{R}^3$ with $\gamma(I) \subseteq S$.

Lemma 5.4 : *Suppose that γ is a smooth curve in S, and $\mathbf{r} : U \longrightarrow S$ is a chart with $\gamma(I) \subseteq \mathbf{r}(U)$. Then there is a smooth curve, $\xi : I \longrightarrow U$ with $\gamma = \mathbf{r} \circ \xi$.*

Proof : Let $\xi = \mathbf{r}^{-1} \circ \gamma : I \longrightarrow U$. To see that ξ is smooth, note that it locally has the form $\xi = \mathbf{f} \circ \gamma$, where $\mathbf{f}$ is a local inverse to $\mathbf{r}$ from an open subset of $\mathbf{R}^3$ to $\mathbf{R}^2$ (cf. the proof of Lemma 5.3). $\diamondsuit$

Lemma 5.4 gives us the same notion of a path in a parameterised surface that we used in Section 4, so we can bring the results there into play.

If the image of γ does not lie in a single coordinate patch, we need to say a bit more. The idea to deal with this is to cut S can be cut into pieces, each of which lies in a coordinate patch, and then apply Lemma 5.4 to each piece. The argument uses the fact (Heine-Borel) that an interval of the form $[a, b]$ is compact. (Exercise for those who know the relevant topology.)

Lemma 5.5 : *If $\gamma : I \longrightarrow S$ is smooth, and $t \in I$, then $\gamma'(t) \in T_p(S)$, where $p = \gamma(t)$.*

Proof : Exercise: use a chart $\mathbf{r} : U \longrightarrow S$ with $p \in \mathbf{r}(U)$, and apply Lemma 5.4. $\diamondsuit$

In fact (exercise) each element of $T_p(S)$ is the derivative $\gamma'(t)$, for some curve γ in S with $\gamma(t) = p$. This allows us to make sense of "tangent space" without explicit reference to charts. Thus, $T_p(S)$ is the set of $\gamma'(0)$ such that γ is a smooth curve with $\gamma(0) = p$. This makes it clear that $T_p(S)$ is well defined as a subspace of $\mathbf{R}^3$. However, it is not imediate from this definition that $T_p(S)$ is a subspace of $\mathbf{R}^3$.

Smooth maps and diffeomorphisms.

Definition : A map $f : S \longrightarrow \mathbf{R}^n$ is *smooth* if, for every chart, $\mathbf{r} : U \longrightarrow S$, the composition $f \circ \mathbf{r} : U \longrightarrow \mathbf{R}^n$ is smooth.

Definition : We say that f is *smooth* at a point p if for any chart, $\mathbf{r} : U \longrightarrow S$ with $p \in \mathbf{r}(U)$, $f \circ \mathbf{r}$ is smooth at $\mathbf{r}^{-1}(p)$. (Note that by Lemma 5.3, it is enough to show that there is some such chart with $p \in \mathbf{r}(U)$ for which this is true.)

Thus f is smooth if and only if it is smooth at every point.

Example : If $\mathbf{r} : U \longrightarrow S$ is a chart, then $\mathbf{r}^{-1} : \mathbf{r}(U) \longrightarrow \mathbf{R}^2$ is smooth.

Definition : Suppose that S and Σ are surfaces. A *diffeomorphism* from S to Σ is a bijective map, $f : S \longrightarrow \Sigma$ such that f and f^{-1} are both smooth (as maps into $\mathbf{R}^3$).

Example : Let $S = \mathbf{S}^2$ be the unit sphere, and let $\Sigma = \{(x, y, z) \mid \frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1\}$ be an ellipsoid (where $a, b, c > 0$ are constant). Then $(x, y, z) \mapsto (ax, by, cz)$ is a diffeomorphism from S to Σ.

Lemma 5.6 : *Suppose that $f : S \longrightarrow \Sigma$ is a bijective map. Then f is a diffeomorphism if and only if, for all charts, $\mathbf{r} : U \longrightarrow S$ and $\mathbf{s} : V \longrightarrow \Sigma$, there is a diffeomorphism $\sigma : U_0 \longrightarrow V_0$ such that*

$$f \circ \mathbf{r}|U_0 = \mathbf{s} \circ \sigma$$

where $U_0 = U \cap (f \circ \mathbf{r})^{-1}\mathbf{s}(V)$ and $V_0 = V \cap \mathbf{s}^{-1}(f \circ \mathbf{r})(U)$.

Proof : Exercise, using Lemma 5.3. ♢

This allows us to deal with diffeomorphisms using charts. As an example, we consider tangent spaces.

Suppose $p \in S$ and let $q = f(p) \in \Sigma$. There is a bijective linear map from $T_p(S)$ to $T_q(\Sigma)$ defined as follows. Let $\mathbf{r} : U \longrightarrow S$ and $\mathbf{s} : V \longrightarrow \Sigma$ be charts with $p \in \mathbf{r}(U)$, $q \in \mathbf{s}(\Sigma)$. Let $\sigma : U_0 \longrightarrow V_0$ be defined by Lemma 5.6. Let $p = \mathbf{r}(u, v)$, so $q = f \circ \mathbf{r}(u, v) = \mathbf{s} \circ \sigma(u, v)$. Note that $(u, v) \in U_0$. Taking derivatives, we have bijective linear maps

$$\mathbf{r}_* : \mathbf{R}^2 \longrightarrow T_p(S)$$

$$\mathbf{s}_* : \mathbf{R}^2 \longrightarrow T_q(\Sigma)$$

$$\sigma_* : \mathbf{R}^2 \longrightarrow \mathbf{R}^2.$$

Taking a composition $\mathbf{s}_* \circ \sigma_* \circ \mathbf{r}_*^{-1}$, we get a bijective linear map

$$T_p(S) \longrightarrow T_q(\Sigma)$$

which we denote by f_*. This is the *derivative* of f at p.

Exercise : Let $f : I \longrightarrow \Sigma$ be a diffeomorphism. If $\gamma : I \longrightarrow S$ is a smooth curve with $\gamma(0) = p$, then $f \circ \gamma : I \longrightarrow \Sigma$ is a smooth curve, and $f_*(\gamma'(0)) = (f \circ \gamma)'(0)$.
Explain how this gives another equivalent definition of the derivative f_*. What advantages and disadvantages does it have over that given above?

Isometries.

Definition : An *isometry* from S to Σ is a diffeomorphism $f : S \longrightarrow \Sigma$ such that for all $p \in S$, the map $f_* : T_p(S) \longrightarrow T_{f(p)}(\Sigma)$ sends the first fundamental form on $T_p(S)$ to the first fundamental form on $T_{f(p)}(\Sigma)$.

Lemma 5.7 : *A diffeomorphism $f : S \longrightarrow \Sigma$ is an isometry if and only if for any curve $\gamma : I \longrightarrow S$ we have* $\text{length}(f \circ \gamma) = \text{length}(\gamma)$.

Proof : We saw in Section 4 that the first fundamental form of a surface determines and is determined by the lengths of curves there.

To be more precise, that result applied to parameterised surfaces, so we need the additional information that a curve in S can be split into subpaths, such that each path lies in a coordinate patch in S and its image under f also lies in a coordinate patch in Σ. This is an exercise in compactness arguments, for those who have the relevant background in topology. $\diamondsuit$

Note that an isometry preserves all the intrinsic properties of a surface, such as area and angles.

We also note the following more general definition.

Definition : A diffeomorphism $f : S \longrightarrow \Sigma$ is *conformal* if there is a smooth function, $\lambda : S \longrightarrow (0, \infty)$ such that for all $p \in S$, the map $f_* : T_p(S) \longrightarrow T_{f(p)}(\Sigma)$ multiplies the first fundamental form by a factor of $\lambda(p)$.

Thus a diffeomorphism is conformal if and only if it preserves angles.

Exercise : A diffeomorphism that is conformal and preserves area is an isometry.

Level sets.

Suppose that $f : \mathbf{R}^n \longrightarrow \mathbf{R}$ is smooth. For each $\mathbf{x} \in \mathbf{R}^n$, we have a derivative map with jacobian $(\, f_{x_1} \quad f_{x_2} \quad \cdots \quad f_{x_3}\,)$, where $f_{x_i} = \partial f / \partial x_i$.

Definition : We say that f is *critical* at $\mathbf{x}$ if $f_{x_i} = 0$ for all i.
Otherwise it is *regular*

Definition : A *level set* of f is a set of the form $f^{-1}(t)$ for some $t \in \mathbf{R}$.

Definition : A level set, $f^{-1}(t)$ is *regular* if f is regular at each point of $f^{-1}(t)$.
In this case, we refer to t as a *regular value*.

As an example think of a smooth map $f : \mathbf{R}^2 \longrightarrow \mathbf{R}$. We can imagine $f(\mathbf{x})$ as measuring the "height" at the location $\mathbf{x}$. The level sets are then the contours of this function. Note that typically on a contour map "most" contours are either circles or lines. One dimension up, one might expect "most" contours to be 2-dimensional surfaces. Here is a precise formulation:

Lemma 5.8 : *Suppose that $f : \mathbf{R}^3 \longrightarrow \mathbf{R}$ is smooth and that $S = f^{-1}(t)$ is regular. Then S is an regular surface in $\mathbf{R}^3$.*

Proof : Let $p \in S$ (so $f(p) = t$). We can suppose (after permuting coordinates) that $f_z \neq 0$ at p. Define $F : \mathbf{R}^3 \longrightarrow \mathbf{R}^3$ by

$$F(x, y, z) = (x, y, f(x, y, z)).$$

Thus F is smooth, and its jacobian is:

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ f_x & f_y & f_z \end{pmatrix}$$

which is invertable at p.

Thus, by the inverse function theorem, there is an inverse $F^{-1} : V \longrightarrow \mathbf{R}^3$ to F defined in an open neighbourhood, V, of $F(p)$ in $\mathbf{R}^3$. Let $U = \{(x, y) \in \mathbf{R}^2 \mid (x, y, t) \in V\}$. We see that U is open in $\mathbf{R}^2$. Define $\mathbf{r} : U \longrightarrow \mathbf{R}^3$ by $\mathbf{r}(x, y) = F^{-1}(x, y, t)$. Then $\mathbf{r}$ is smooth. Moreover, $F(\mathbf{r}(x, y)) = F \circ F^{-1}(x, y, t) = (x, y, t)$, so by the definition of F, $f(\mathbf{r}(x, y)) = t$, and so $\mathbf{r}(x, y) \in S$. In other words, $\mathbf{r}(U) \subseteq S$. In fact, $\mathbf{r}(U) = S \cap W$, where $W = F^{-1}(V)$ is an open subset of $\mathbf{R}^3$. Moreover, $\mathbf{r}^{-1} : \mathbf{r}(U) \longrightarrow \mathbf{R}^2$ is continuous (since F is). It follows that $\mathbf{r} : U \longrightarrow S$ is a chart with $p \in \mathbf{r}(U)$. Since $p \in S$ was arbitrary, it follows that S is a surface. $\diamondsuit$

Examples :

(1) $f(x, y, z) = x^2 + y^2 + z^2$.

$f_x = 2x,\ f_y = 2y,\ f_z = 2z$.

Thus, $f^{-1}(t)$ is an embedded surface for all $t > 0$. (A sphere of radius $\sqrt{t}$.)

More generally, $f(x, y, z) = \frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2}$.

The level sets for $t > 0$ are ellipsoids.

(2) $f(x, y, z) = x^2 + y^2 - z^2$.

$f^{-1}(0)$ is a cone which is singular at the origin.

$f^{-1}(t)$ is non-singular for all $t \neq 0$. It is a hyperboloid — 1-sheeted for $t > 0$ and 2-sheeted for $t < 0$.

Cartography.

We can apply some of these ideas to describing "maps" of the earth's surface. Let the earth be:

$$S = \mathbf{S}^2 = \{\mathbf{x} \in \mathbf{R}^3 \mid ||\mathbf{x}|| = 1\}.$$

Definition : A *projection* of S is the inverse of a chart, i.e.: a map $\mathbf{w} : V \longrightarrow \mathbf{R}^2$ where $V \subseteq S$ and $\mathbf{r} = \mathbf{w}^{-1} : U \longrightarrow S$ is a chart with $\mathbf{r}(U) = V$.

(This is geographers' terminology: it's not completely standard mathematically.)

We will see (in Section 8) that such a map cannot be an isometry to $\mathbf{R}^2$ — so it cannot preserve distances (lengths). We need to do the best we can. We can try to preserve angles or area, etc. (Note that by the above fact, and earlier exercise we cannot simultanously preserve angles and area.)

Recall spherical polar coordinates:

$$(\theta, \phi) \mapsto (\cos\theta\cos\phi, \sin\theta\cos\phi, \sin\phi)$$

from $\mathbf{R} \times (-\pi/2, \pi/2) \longrightarrow S$. If we restrict to $U = (-\pi, \pi) \times (-\pi/2, \pi/2)$, we get a chart with image $S \setminus L$ where L is the idealised date line, i.e. the 180 degree longitude. The first fundamental form is:

$$\cos^2\phi\, d\theta^2 + d\phi^2.$$

Its inverse is a projection (see the map of "The Living Earth" outside room B1.28). Here are some other examples:

(1) Stereographic projection.
Suppose that (θ, ϕ) is projected to the point (x, y) in the plane. An argument using similar triangles (see lectures) shows that the distance of (x, y) from the origin is $\frac{2\cos\phi}{1-\sin\phi}$, and so

$$(x, y) = \left(\frac{2\cos\theta\cos\phi}{1-\sin\phi}, \frac{2\sin\theta\cos\phi}{1-\sin\phi}\right).$$

from which one calculates

$$dx^2 + dy^2 = \frac{4}{(1-\sin\phi)^2}(\cos^2\phi\, d\theta^2 + d\phi^2).$$

It follows that stereographic projection is conformal.

(2) [See M, p.126] We consider projections that send latitudes and longitudes to straight lines. To this end, define a projection by: $(x, y) \equiv (\theta, v(\phi))$. Thus, $x_\theta = 1$, $x_\phi = 0$, $y_\theta = 0$, $y_\phi = v'$. We get the euclidean first fundamental form:

$$dx^2 + dy^2 \equiv d\theta^2 + (v')^2 d\phi^2$$

which we want to compare with the spherical first fundamental form:

$$\cos^2\phi\, d\theta^2 + d\phi^2.$$

(2a) Lambert's cylindrical projection.

If we want our projection to be area preserving, we need $(v')^2 = \cos^2\phi$ (so that $dudv \equiv d\theta d\phi$). So $v' = \cos\phi$ and we integrate to give $v = \sin\theta$. That is,

$$(x, y) \equiv (\theta, \sin\phi).$$

Geometrically, this is the same as projecting the sphere to a cylinder radially outward from the north-south axis, and then unwrapping the cylinder into the plane.

This is sometimes seen as a "politically correct" projection, since countries far from the equator don't appear bigger than they deserve to be. But they end up getting distorted quite dramatically.

It was described in detail by Lambert in 1772, though was probably known in some form to Archimedes. A variation where the vertical direction is stretched linearly so as to distribute distortion more evenly was advocated by Gall in the mid nineteenth century, and more recently publicised by some bloke called Peters.

(2b) Mercator's projection.

If we want our projection to be conformal we need $(v')^2 = 1/\cos^2\phi$. So $v' = \sec\phi$ and we integrate to give $v = \log\tan\left(\frac{\pi}{4} + \frac{\phi}{2}\right)$. that is:

$$(x, y) \equiv \left(\theta, \log\tan\left(\frac{\pi}{4} + \frac{\phi}{2}\right)\right).$$

Exercise: This sends loxodromes (paths of constant compass bearing) to straight lines.

The Mercator projection gives an accurate picture locally, but it grossly misrepresents area on a large scale. It makes Greenland look about the same size as Africa.

This projection was introduced by Gheert Cremer (a.k.a. Gerardus Mercator) in 1569. It's still used today.

(3) Other projections.

There are loads of other projections in use. For example, the "sinusoidal projection" is another example of an area-preserving map.

Exercise : Check these out on the web, and verify they have the properties claimed of them.

6. The second fundamental form.

Unlike the first fundamental form, the second fundamental form is an extrinsic property of the surface. It describes how the surface is bent into 3-space. It can be viewed in a number of different ways. For example, it describes how the first fundamental form varies if we move the surface in the normal direction. We will introduce it from this viewpoint.

Let $\mathbf{r} : U \longrightarrow \mathbf{R}^3$ be a parameterised surface. We write $S = \mathbf{r}(U)$. Given $t \in \mathbf{R}$, let

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

(Recall that $\mathbf{n}(u,v)$ is the unit normal.) We assume that $\mathbf{r}(t) : U \longrightarrow \mathbf{R}^3$ is itself a parameterised surface for small t. (There is no essential loss in making this assumption. Exercise: given any $(u,v) \in U$ there is a neighbourhood U_0 of (u,v) in U and some $t_0 > 0$ such that $\mathbf{r}(t)|U_0$ is a parameterised surface for all $t \in [-t_0, t_0]$. Since the second fundamental form is a local property, so this is good enough.) We will refer to $\mathbf{r}(t)$ as the "normal displacement" of $\mathbf{r}$. We get

$$\begin{aligned} \mathbf{f}_u &= \mathbf{r}_u - t\mathbf{n}_u \\ \mathbf{f}_v &= \mathbf{r}_v - t\mathbf{n}_v. \end{aligned}$$

The first fundamental form of $\mathbf{r}(t)$ is

$$E(t)du^2 + 2F(t)dudv + G(t)dv^2,$$

where

$$E(t) = \mathbf{f}_u.\mathbf{f}_u = \mathbf{r}_u.\mathbf{r}_u - 2t\mathbf{r}_u.\mathbf{n}_u + t^2\mathbf{n}_u.\mathbf{n}_u,$$

and so

$$\frac{1}{2}\frac{d}{dt}E(t)\bigg|_{t=0} = -\mathbf{r}_u.\mathbf{n}_u.$$

We similarly get:

$$\begin{aligned} \frac{1}{2}\frac{d}{dt}F(t)\bigg|_{t=0} &= -\frac{1}{2}(\mathbf{r}_u.\mathbf{n}_v + \mathbf{r}_v.\mathbf{n}_u) \\ \frac{1}{2}\frac{d}{dt}G(t)\bigg|_{t=0} &= -\mathbf{r}_v.\mathbf{n}_v. \end{aligned}$$

Thus, half the first variation of the first fundamental form under normal displacement is given by:

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

where

$$\left.\begin{aligned} L &= -\mathbf{r}_u.\mathbf{n}_u \\ M &= -\frac{1}{2}(\mathbf{r}_u.\mathbf{n}_v + \mathbf{r}_v.\mathbf{n}_u) \\ N &= -\mathbf{r}_v.\mathbf{n}_v \end{aligned}\right..$$

Definition : The expression

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

is the *second fundamental form* of the parameterised surface at (u,v).

More formally, it should be viewed as a quadratic form with matrix:

$$\begin{pmatrix} L & M \\ M & N \end{pmatrix}$$

with respect to the standard basis of $\mathbf{R}^2$.

Remark : : Some books use the letters e, f, g in place of L, M, N.

Unlike the first fundamental form, it need not be positive definite.
Here is another expression. Note that $\mathbf{r}_u.\mathbf{n} = 0$, so

$$\frac{\partial}{\partial u}(\mathbf{r}_u.\mathbf{n}) = \mathbf{r}_{uu}.\mathbf{n} + \mathbf{r}_u.\mathbf{n}_u = 0.$$

Similarly,

$$\mathbf{r}_{vu}.\mathbf{n} + \mathbf{r}_v.\mathbf{n}_u = 0$$

$$\mathbf{r}_{uv}.\mathbf{n} + \mathbf{r}_u.\mathbf{n}_v = 0$$

$$\mathbf{r}_{vv}.\mathbf{n} + \mathbf{r}_v.\mathbf{n}_v = 0.$$

Note that $\mathbf{r}_{uv} = \mathbf{r}_{vu}$ and so $\mathbf{r}_v.\mathbf{n}_u = \mathbf{r}_u.\mathbf{n}_v$. Thus:

$$\begin{aligned} L &= -\mathbf{r}_u.\mathbf{n}_u = \mathbf{r}_{uu}.\mathbf{n} \\ M &= -\mathbf{r}_u.\mathbf{n}_v = -\mathbf{r}_v.\mathbf{n}_u = \mathbf{r}_{uv}.\mathbf{n}\,. \\ N &= -\mathbf{r}_v.\mathbf{n}_v = \mathbf{r}_{vv}.\mathbf{n} \end{aligned}$$

Examples :
(1) The plane. $\mathbf{r}(u, v) = \mathbf{a} + u\mathbf{b} + v\mathbf{c}$. Here $\mathbf{r}_{uu} = \mathbf{r}_{uv} = \mathbf{r}_{vv} = 0$, so the second fundamental form is identically 0.
(2) The sphere of radius a:
Here $\mathbf{r} = -a\mathbf{n}$, so $\mathbf{r}_u = -a\mathbf{n}_u$, $\mathbf{r}_v = -a\mathbf{n}_v$, and it follows that $L = -\mathbf{r}_u.\mathbf{n}_u = a^{-1}E$. Similarly $M = a^{-1}F$ and $N = a^{-1}G$. and so

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

So, for example, in spherical polars $(u, v) = (\theta, \phi)$, we get the s.f.f. as

$$a^{-1}(a^2\cos^2\phi d\theta^2 + a^2 d\phi^2) = a\cos^2\phi d\theta^2 + ad\phi^2.$$

(3) Graphs:
Let $\mathbf{r}(u, v) = (u, v, f(u, v))$. Then $\mathbf{r}_{uu} = (0, 0, f_{uu}), \mathbf{r}_{uv} = (0, 0, f_{uv}), \mathbf{r}_{vv} = (0, 0, f_{vv})$.
Suppose $f_u = f_v = 0$ at (u, v). Then $\mathbf{n} = (0, 0, 1)$ and so:

$$\begin{pmatrix} L & M \\ M & N \end{pmatrix} = \begin{pmatrix} f_{uu} & f_{uv} \\ f_{uv} & f_{vv} \end{pmatrix}$$

In other words, at a critical point, the s.f.f. is just the hessian of the function f.
For example $f(u, v) = au^2 + bv^2$ gives $2adu^2 + 2bdv^2$ at the origin.

Exercise : If the second fundamental form is identically zero, then the surface is planar.

The shape operator.

Another interpretation allows us to view the second fundamental form in terms of variation of normals.

Note that $\mathbf{n}_u$ and $\mathbf{n}_v$ are both orthogonal to $\mathbf{n}$, and so lie in the tangent space, that is the subspace, $T_p(S)$, of $\mathbf{R}^3$ spanned by $\mathbf{r}_u$ and $\mathbf{r}_v$. There is therefore a linear map, $\Theta : T_p(S) \longrightarrow T_p(S)$, which sends $\mathbf{r}_u$ to $-\mathbf{n}_u$ and $\mathbf{r}_v$ to $-\mathbf{n}_v$. With respect to the basis $\mathbf{r}_u, \mathbf{r}_v$ for $T_p(S)$ we write its matrix as

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}.$$

In other words $\Theta(\mathbf{r}_u) = a\mathbf{r}_u + c\mathbf{r}_v$ and $\Theta(\mathbf{r}_v) = b\mathbf{r}_u + d\mathbf{r}_v$. This is the same as the matrix for Θ with respect the standard basis of $\mathbf{R}^2$ under the identification $\mathbf{r}^* : \mathbf{R}^2 \longrightarrow T_p(S)$. (Recall that $\mathbf{r}_*$ sends the standard basis, $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$, $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$ of $\mathbf{R}^2$ to the basis $\mathbf{r}_u$, $\mathbf{r}_v$ of $T_p(S)$.)

The map Θ is sometimes called the *shape operator*.

We claim that

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} E & F \\ F & G \end{pmatrix}^{-1} \begin{pmatrix} L & M \\ M & N \end{pmatrix}.$$

(This is a form of the "Weingarten equations".)
By definition we have:

$$-\mathbf{n}_u = a\mathbf{r}_u + c\mathbf{r}_v$$

$$-\mathbf{n}_v = b\mathbf{r}_u + d\mathbf{r}_v.$$

Taking dot products with $\mathbf{r}_u$ and $\mathbf{r}_v$, we get the four equalities:

$$\begin{aligned} L &= aE + cF \\ M &= bE + dF \\ M &= aF + cG \\ N &= bF + dG \end{aligned}$$

which can be written as:

$$\begin{pmatrix} E & F \\ F & G \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} L & M \\ M & N \end{pmatrix},$$

so the above statement follows.

Interpretation in terms of linear algebra.

First we recall some abstract linear algebra. The following makes sense in $\mathbf{R}^n$, though we restrict here to $\mathbf{R}^2$.

Suppose that $\langle .,. \rangle$ is a symmetric bilinear form on $\mathbf{R}^2$ given by the symmetric 2×2 matrix A. That is:

$$\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}^T A \mathbf{y} = \mathbf{y}^T A^T \mathbf{x} = \mathbf{y}^T A \mathbf{x} = \langle \mathbf{y}, \mathbf{x} \rangle$$

for all $\mathbf{x}, \mathbf{y} \in \mathbf{R}^2$.

Suppose that $\langle .,. \rangle'$ is another symmetric bilinear form given by a matrix B. That is

$$\langle \mathbf{x}, \mathbf{y} \rangle' = \mathbf{x}^T B \mathbf{y}.$$

Suppose that A is positive definite. Then there is a (unique) linear map $\Phi : \mathbf{R}^2 \longrightarrow \mathbf{R}^2$ such that for all $\mathbf{x}, \mathbf{y}$

$$\langle \mathbf{x}, \mathbf{y} \rangle' = \langle \mathbf{y}, \mathbf{x} \rangle' = \langle \Phi(\mathbf{x}), \mathbf{y} \rangle = \langle \mathbf{x}, \Phi(\mathbf{y}) \rangle.$$

In fact, we set Φ to be the transformation given by the matrix $P = A^{-1}B$. Thus

$$\langle \mathbf{x}, \Phi(\mathbf{y}) \rangle = \mathbf{x}^T A P \mathbf{y} = \mathbf{x}^T A A^{-1} B \mathbf{y} = \mathbf{x}^T B \mathbf{y} = \langle \mathbf{x}, \mathbf{y} \rangle'$$

$$\langle \Phi(\mathbf{x}), \mathbf{y} \rangle = (P\mathbf{x})^T A \mathbf{y} = \mathbf{x}^T B A^{-1} A \mathbf{y} = \mathbf{x}^T B \mathbf{y} = \langle \mathbf{x}, \mathbf{y} \rangle'.$$

The map Φ is diagonalisable. (This is because it is self-adjoint with respect to $\langle .,. \rangle$. If we take a basis for $\mathbf{R}^2$ that is orthornormal with respect to $\langle .,. \rangle$ then Φ will be represented by a symmetric matrix.) Let $\mathbf{w}_1, \mathbf{w}_2$ be eigenvectors, and λ_1, λ_2 the corresponding eigenvalues. That is, $\Phi(\mathbf{w}_1) = \lambda_1 \mathbf{w}_1$ and $\Phi(\mathbf{w}_2) = \lambda_2 \mathbf{w}_2$. We can normalise so that $\langle \mathbf{w}_1, \mathbf{w}_1 \rangle = \langle \mathbf{w}_2, \mathbf{w}_2 \rangle = 1$. If $\lambda_1 \neq \lambda_2$ then necessarily $\langle \mathbf{w}_1, \mathbf{w}_2 \rangle = 0$. If $\lambda_1 = \lambda_2$, we can always choose $\mathbf{w}_1, \mathbf{w}_2$ so that $\langle \mathbf{w}_1, \mathbf{w}_2 \rangle = 0$. In other words, we can assume that $\mathbf{w}_1$, $\mathbf{w}_2$ are orthonormal with respect to $\langle .,. \rangle$.

In the case where $\langle .,. \rangle$ and $\langle .,. \rangle'$ are respectively the first and second fundamental forms, then we get:

$$A = \begin{pmatrix} E & F \\ F & G \end{pmatrix} \qquad B = \begin{pmatrix} L & M \\ M & N \end{pmatrix}$$

and so

$$P = A^{-1}B = \begin{pmatrix} a & b \\ c & d \end{pmatrix}.$$

In other words, Φ corresponds under the derivative $\mathbf{r}_* : \mathbf{R}^2 \longrightarrow T_p(S)$ to the map Θ described above.

In summary, we have a map

$$\Theta : T_p(S) \longrightarrow T_p(S).$$

We have shown that, with respect to the basis $\mathbf{r}_u, \mathbf{r}_v$ of $T_p(S)$ it has matrix:

$$\begin{pmatrix} E & F \\ F & G \end{pmatrix}^{-1} \begin{pmatrix} L & M \\ M & N \end{pmatrix}.$$

Equivalently, the map induced from $\mathbf{R}^2$ to $\mathbf{R}^2$ under the derivative identification $\mathbf{r}_* : \mathbf{R}^2 \longrightarrow T_p(S)$ is represented by the above matrix with respect to the standard basis for $\mathbf{R}^2$.

As far as we know, for the moment, the map Θ might depend on the parameterisation, $\mathbf{r}$, of S, since it was defined in these terms. In Section 7, we will see that it is in fact independent of parameterisation (see Corollary 7.2).

Now Θ is diagonalisable. As discussed above, we can find $\mathbf{w}_1, \mathbf{w}_2 \in T_p(S)$ which are othonormal with respect to the first fundamental form (that is $||\mathbf{w}_1|| = ||\mathbf{w}_2|| = 1$ and $\mathbf{w}_1.\mathbf{w}_2 = 0$ with respect to the standard dot product on $\mathbf{R}^3$) and $\kappa_1, \kappa_2 \in \mathbf{R}$ such that $\Theta(\mathbf{w}_1) = \kappa_1 \mathbf{w}_1$ and $\Theta(\mathbf{w}_2) = \kappa_2 \mathbf{w}_2$.

Definition : κ_1 and κ_2 are the *principal curvatures* of $S = \mathbf{r}(U)$ at the point $p = \mathbf{r}(u, v)$.

Definition : The product, $\kappa = \kappa_1 \kappa_2$, is the *Gauss curvature* of $S = \mathbf{r}(U)$ at the point $p = \mathbf{r}(u, v)$.

(Again, we will see in Section 7 that these notions are invariant under reparameterisation of S.)

Note that

$$\begin{aligned} \kappa &= \kappa_1 \kappa_2 \\ &= \det \begin{pmatrix} \kappa_1 & 0 \\ 0 & \kappa_2 \end{pmatrix} \\ &= \det \begin{pmatrix} a & b \\ c & d \end{pmatrix} \\ &= \det \begin{pmatrix} E & F \\ F & G \end{pmatrix}^{-1} \begin{pmatrix} L & M \\ M & N \end{pmatrix} \\ &= \frac{LN - M^2}{EG - F^2}. \end{aligned}$$

(To get between lines 2 and 3 above, we have used the fact that these two matrices represent the same linear map, Θ, with respect to different bases.)

We have shown:

Theorem 6.1 : *The Gauss curvature is given by:*

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

$\diamondsuit$

We will return to Gauss curvature in Section 8.

Examples :

(1) Plane.

Clearly $\kappa_1 = \kappa_2 = \kappa = 0$.

(2) Sphere of radius a.

We saw that the s.f.f. is a^{-1} times the f.f.f. Thus

$$\begin{pmatrix} E & F \\ F & G \end{pmatrix}^{-1} \begin{pmatrix} L & M \\ M & N \end{pmatrix} = a^{-1} I.$$

and so

$$\kappa_1 = \kappa_2 = a^{-1} \qquad \kappa = a^{-2}.$$

(3) Graphs.

$(u, v) \mapsto (u, v, f(u, v))$, with critical point at the origin. The f.f.f. is the identity and the s.f.f. is the hessian. Thus, κ_1, κ_2 are the eigenvalues of the hessian matrix, and

$$\kappa = \kappa_1 \kappa_2 = \det \begin{pmatrix} f_{uu} & f_{uv} \\ f_{uv} & f_{vv} \end{pmatrix} = f_{uv}^2 - f_{uu} f_{vv}.$$

Note that if f has a maximum or minimum at the critical point, then $\kappa \geq 0$. If it has a saddle, then $\kappa \leq 0$.

(4) Developable surface.

$$\begin{aligned} \mathbf{r} &= \gamma(u) + v\mathbf{T}(u) \\ \mathbf{r}_u &= \mathbf{T} + v\kappa\mathbf{N} \\ \mathbf{r}_v &= \mathbf{T}. \end{aligned}$$

Here $\mathbf{N}$ is the (principal) normal to γ. (Recall that $\mathbf{T}' = \kappa\mathbf{N}$.) The normal to the surface is $\mathbf{B} = \mathbf{T} \wedge \mathbf{N}$, i.e. the binormal to γ.

$$\mathbf{r}_{uu} = \kappa\mathbf{N} + v\kappa'\mathbf{N} + v\kappa(-\kappa\mathbf{T} + \tau\mathbf{B})$$

since $\mathbf{N}' = -\kappa\mathbf{T} + \tau\mathbf{B}$, and

$$\mathbf{r}_{uv} = \mathbf{T}_u = \kappa\mathbf{N} \qquad \mathbf{r}_{vv} = 0.$$

Thus

$$L = \mathbf{r}_{uu}.\mathbf{B} = v\kappa\tau \qquad M = \mathbf{r}_{uv}.\mathbf{B} = 0 \qquad N = 0.$$

And so the second fundamental form is

$$v\kappa\tau\, du^2.$$

Also

$$\begin{pmatrix} E & F \\ F & G \end{pmatrix}^{-1} \begin{pmatrix} L & M \\ M & N \end{pmatrix} = \begin{pmatrix} 1 + v^2\kappa^2 & 1 \\ 1 & 1 \end{pmatrix}^{-1} \begin{pmatrix} v\kappa\tau & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} \tau/v\kappa & 0 \\ -\tau/v\kappa & 0 \end{pmatrix}$$

so the principal curvatures are $0, \tau/v\kappa$, with the 0 eigenvalue corresponding to the r_v direction. In particular, we see that $\kappa = 0$.

Exercise : The principal (hence Gauss) curvatures are invariant under postcomposition with a rigid motion of $\mathbf{R}^3$.

Mean curvature.

Let $H = \frac{1}{2}(\kappa_1 + \kappa_2)$, where κ_1, κ_2 are the principal curvatures.

Definition : H is the *mean curvature* of S at p.

Like Gauss curvature, this is a smooth function on the surface.

Note that this is the same as half the trace of the shape operator.

Exercise :

$$H = \frac{GL + EN - 2FM}{2(EG - F^2)}.$$

(Recall that the trace of a matrix is invariant under conjugation.)

The mean curvature measures the first variation in area under normal displacement of S (see exercise sheet).

Definition : A *minimal surface* is one for which the mean curvature vanishes.

In other words, $GL + EN = 2FM$ everywhere.

For example a surface that locally minimises area is a minimal surface. Hence soap films are minimal surfaces (provided the air pressures on both sides are equal).

Examples : The plane, helicoid and catenoid are all minimal surfaces. (Exercise — by tedious calculation, or see exercise sheet for a simpler way.)

7. Covariant derivatives and parallel transport.

We shall formulate this notion in terms of embedded surfaces in $\mathbf{R}^3$.

Let S be a smooth surface. Recall that a map $f : S \longrightarrow \mathbf{R}^n$ is smooth if and only if $f \circ \mathbf{r} : U \longrightarrow \mathbf{R}^n$ is smooth for all charts $\mathbf{r} : U \longrightarrow S$.

Definition : An *ambient vector field* on S is a smooth map $\mathbf{a} : S \longrightarrow \mathbf{R}^3$.

Here we imagine $\mathbf{a}(p)$ as a vector based at the point p. We can also speak of ambient tangent fields restricted to subsets of S.

Example : The normal vector field $p \mapsto \mathbf{n}(p)$.

Definition : A *tangential vector field* is an ambient vector field, $\mathbf{a}$, such that $\mathbf{a}(p) \in T_p(S)$ for all $p \in S$.

Here we can imagine $\mathbf{a}(p)$ as lying "on" the surface S.

Example : If $\mathbf{r} : U \longrightarrow S$ is a chart, then $\mathbf{r}_u$ and $\mathbf{r}_v$ are tangential vector fields on $\mathbf{r}(U)$.

Suppose that $\mathbf{a}$ is an ambient vector field on S. If $\mathbf{r} : U \longrightarrow S$ is a chart, then the derivatives $\mathbf{a}_u, \mathbf{a}_v$ are also vector fields (on $\mathbf{r}(U)$). (More precisely, we define $\mathbf{a}_u = \frac{\partial}{\partial u}\mathbf{a} \circ \mathbf{r}(u, v)$ and $\mathbf{a}_v$ similarly.) Suppose that $\mathbf{e} \in T_p(S)$ is a vector (which we imagine to be based at p). We write $\mathbf{e} = \lambda \mathbf{r}_u + \mu \mathbf{r}_v$, and set

$$D_{\mathbf{e}}(\mathbf{a}) = \lambda \mathbf{a}_u + \mu \mathbf{a}_v.$$

[This is not standard notation, though the notation $\nabla_{\mathbf{e}}$ derived from it below, is fairly standard.]

Lemma 7.1 : *$D_{\mathbf{e}}\mathbf{a}$ is independent of the chart $\mathbf{r}$.*

Proof : Suppose we have another chart, say $\mathbf{s}$, with coordinates η, θ. By Lemma 5.3, transition maps are smooth, so we can write:

$$\mathbf{s}_\eta = u_\eta \mathbf{r}_u + v_\eta \mathbf{r}_v$$

$$\mathbf{s}_\theta = u_\theta \mathbf{r}_u + v_\theta \mathbf{r}_v.$$

In terms of these new coordinates, write

$$\mathbf{e} = \bar{\lambda} \mathbf{s}_\eta + \bar{\mu} \mathbf{s}_\theta.$$

We verify that $\lambda = \bar{\lambda} u_\eta + \bar{\mu} u_\theta$ and $\mu = \bar{\lambda} v_\eta + \bar{\mu} v_\theta$.

Now, in terms of the new coordinates we have:

$$\begin{aligned} D_{\mathbf{e}}\mathbf{a} &= \bar{\lambda} \mathbf{a}_\eta + \bar{\mu} \mathbf{a}_\theta \\ &= \bar{\lambda}(u_\eta \mathbf{a}_u + v_\eta \mathbf{a}_v) + \bar{\mu}(u_\theta \mathbf{a}_u + v_\theta \mathbf{a}_v) \\ &= (\bar{\lambda} u_\eta + \bar{\mu} u_\theta)\mathbf{a}_u + (\bar{\lambda} v_\eta + \bar{\mu} v_\theta)\mathbf{a}_v \\ &= \lambda \mathbf{a}_u + \mu \mathbf{a}_v \end{aligned}$$

which agrees with the original formula. $\diamondsuit$

Note also that the map $\mathbf{e} \mapsto D_{\mathbf{e}}\mathbf{a}$ is linear in $\mathbf{e}$.

Application to the second fundamental form.

Before continuing, we relate this to the shape operator, as described in Section 6. Recall that $\Theta : T_p(S) \longrightarrow T_p(S)$ was the linear map defined by $\Theta(\mathbf{r}_u) = -\mathbf{n}_u$ and $\Theta(\mathbf{r}_v) = -\mathbf{n}_v$. In the above notation, this becomes $\Theta(\mathbf{r}_u) = -D_{\mathbf{r}_u}\mathbf{n}$ and $\Theta(\mathbf{r}_v) = -D_{\mathbf{r}_v}\mathbf{n}$. By linearity of both sides we see that

$$\Theta(\mathbf{e}) = -D_{\mathbf{e}}\mathbf{n}$$

for all tangent vectors in $\mathbf{e} \in T_p(S)$. In particular, using Lemma 7.1, we deduce the result promised in Section 6, namely:

Corollary 7.2 : *The map $\Theta : T_p(S) \longrightarrow T_p(S)$ is invariant under reparametisation.* $\Diamond$

Note that in these terms, we can write the second fundamental form as:

$$\Pi(\mathbf{e}, \mathbf{f}) = \mathbf{e}.\Theta(\mathbf{f}) = -\mathbf{e}.D_{\mathbf{f}}\mathbf{n},$$

where $\mathbf{e}, \mathbf{f} \in T_p(S)$. It follows that the second fundamental form on $T_p(S)$ is also well defined.

Exercise : Verify directly that $\Pi(\mathbf{e}, \mathbf{f}) = \Pi(\mathbf{f}, \mathbf{e})$.

Remark : We can also view the shape operator as (minus) the derivative of the Gauss map. This is just a shift of perspective. Instead of imagining $\mathbf{n}$ as an ambient vector field on S, we can think of it as a map $\mathbf{n} : S \longrightarrow \mathbf{S}^2$ to the unit 2-sphere, $\mathbf{S}^2$. This is called the *Gauss map*. We check that the tangent spaces $T_p(S)$ and $T_{\mathbf{n}(p)}(\mathbf{S}^2)$ are equal, and that $-\Theta$ is just the derivative, $\mathbf{n}_*$, of this map.

Continuing where we left off :

Suppose that $\mathbf{e}$ is now a tangent vector field on S. We define the ambient vector field $D_{\mathbf{e}}\mathbf{a}$ by setting $D_{\mathbf{e}}\mathbf{a}(p)$ to be $D_{\mathbf{e}(p)}\mathbf{a}$ evaluated at p. Note that, by definition, $D_{\mathbf{r}_u}\mathbf{a} = \mathbf{a}_u$.

(Strictly speaking, $\mathbf{r}_u$ is not a vector field on $\mathbf{r}(U)$, since its domain is U, not $\mathbf{r}(U)$. Formally we should write it as $\mathbf{r}_u \circ \mathbf{r}^{-1} : \mathbf{r}(U) \longrightarrow \mathbf{R}^3$. When precomposed with the chart $\mathbf{r}$, we recover $(\mathbf{r}_u \circ \mathbf{r}^{-1}) \circ \mathbf{r} = \mathbf{r}_u : U \longrightarrow \mathbf{R}^3$. Informally when the chart is clear, it is often omitted from the notation. We informally think of u and v as co-ordinates locally in the surface. Thus we can think of (u, v) as determining the point $p = \mathbf{r}(u, v)$ without explicitly mentioning $\mathbf{r}$.)

If $f : S \longrightarrow \mathbf{R}$ is a smooth function, then we can similarly define $D_{\mathbf{r}_u} f = f_u = \partial f / \partial u$ and $D_{\mathbf{r}_v} f = f_v = \partial f / \partial v$. As with vector fields, we can extend linearly to define $D_{\mathbf{e}} f$, which is another smooth function on S. As before, we can verify that this is independent of parametisation. Moreover, a simple calculation gives the product formula:

$$D_{\mathbf{e}}(f\mathbf{a}) = (D_{\mathbf{e}} f)\mathbf{a} + f D_{\mathbf{e}}\mathbf{a}.$$

(There are also obvious formulae for differentiating products of functions, and dot products of vector fields.)

Note that $D_{\mathbf{e}}\mathbf{a}$ need not be a tangential field, even if $\mathbf{a}$ is. To rectify this, we define:

$$\nabla_{\mathbf{e}}\mathbf{a} = D_{\mathbf{e}}\mathbf{a} - ((D_{\mathbf{e}}\mathbf{a}).\mathbf{n})\mathbf{n}.$$

In other words, $\nabla_{\mathbf{e}}\mathbf{a}$ is the orthogonal projection of $D_{\mathbf{e}}\mathbf{a}$ to $T_p(S)$. By definition, this is a tangential field. (Of course, one needs to verify that all these constructions are smooth.) We don't need $\mathbf{a}$ itself to be tangential in the above.

If $\mathbf{a}$ is tangential we refer to $\nabla_{\mathbf{e}}\mathbf{a}$ as the *covariant derivative* of $\mathbf{a}$ in the direction of $\mathbf{e}$. Viewed as a function with two arguments, $\mathbf{e}$ and $\mathbf{a}$, ∇ is sometimes referred to as a *connection*. (More precisely, it is the "Levi-Cevita connection" in this case.)

If f is a smooth function, we will simply write: $\nabla_{\mathbf{e}}f = D_{\mathbf{e}}f$.

Lemma 7.3 : *Suppose that λ is a smooth function on S, and that $\mathbf{a}$ is a tangential vector field. Then*

$$\nabla_{\mathbf{e}}(\lambda\mathbf{a}) = (\nabla_{\mathbf{e}}\lambda)\mathbf{a} + \lambda\nabla_{\mathbf{e}}\mathbf{a}.$$

Proof : By linearity in $\mathbf{e}$, it's enough to show that:

$$\nabla_u(\lambda\mathbf{a}) = \lambda_u\mathbf{a} + \lambda\nabla_u\mathbf{a}.$$

Using $\mathbf{a}.\mathbf{n} = 0$, we have:

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

as required. $\diamondsuit$

By a similar argument, if n is the unit normal to S, then

$$\nabla_{\mathbf{e}}(\lambda\mathbf{n}) = \lambda\nabla_{\mathbf{e}}\mathbf{n}.$$

Intuitively, the contribution to the derivative due to varying λ is in the normal direction, and so this term vanishes on projecting to the tangent space.

We next aim to prove:

Theorem 7.4 : *Suppose that* $\mathbf{a}, \mathbf{e}$ *are tangential vector fields. Then* $\nabla_{\mathbf{e}}\mathbf{a}$ *depends only on the first fundamental form of* S.

In other words, it is an intrinsic construction. In fact, if $\mathbf{r} : U \longrightarrow S$ is a chart, then we can express $\nabla_{\mathbf{e}}\mathbf{a}$ in terms of the coordinates of $\mathbf{e}$ and $\mathbf{a}$ with respect to $\mathbf{r}_u$ and $\mathbf{r}_v$, the derivatives of these coordinates with respect to u and v, the functions E, F, G, featuring in the first fundamental form, and their derivatives, $E_u, F_u, G_u, E_v, F_v, G_v$. We shall not derive an explicit expression here, though this is in principle possible by working carefully through the argument given below. (See [???] for a formula.)

For the proof we make the following observation:

Lemma 7.5 : *Suppose that* $\mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathbf{R}^3$ *are vectors and that* $\mathbf{x}$ *and* $\mathbf{y}$ *are linearly independent. Then the orthogonal projection of* $\mathbf{z}$ *to the plane spanned by* $\mathbf{x}$ *and* $\mathbf{y}$ *is given by*

$$\frac{((\mathbf{y}.\mathbf{y})(\mathbf{x}.\mathbf{z}) - (\mathbf{x}.\mathbf{y})(\mathbf{y}.\mathbf{z}))\mathbf{x} + ((\mathbf{x}.\mathbf{x})(\mathbf{y}.\mathbf{z}) - (\mathbf{x}.\mathbf{y})(\mathbf{x}.\mathbf{z}))\mathbf{y}}{(\mathbf{x}.\mathbf{x})(\mathbf{y}.\mathbf{y}) - (\mathbf{x}.\mathbf{y})^2}$$

Proof : Exercise. ♢

All we really need to note from this is that the coordinates are expressible in terms of $\mathbf{x}.\mathbf{x}$, $\mathbf{y}.\mathbf{y}$, $\mathbf{x}.\mathbf{y}$, $\mathbf{x}.\mathbf{z}$ and $\mathbf{y}.\mathbf{z}$.

Proof of Theorem 7.4 : By linearity in $\mathbf{e}$, it is enough to express $\nabla_u\mathbf{a}$ and $\nabla_v\mathbf{a}$ in terms of the first fundamental form. We will just deal with $\nabla_u\mathbf{a}$. Now $\mathbf{a} = \lambda\mathbf{r}_u + \mu\mathbf{r}_v$ where λ, μ are smooth functions on S. Also, by Lemma 7.3,

$$\nabla_u(\lambda\mathbf{r}_u) = \lambda_u\mathbf{r}_u + \lambda\nabla_u\mathbf{r}_u$$

$$\nabla_u(\mu\mathbf{r}_v) = \mu_u\mathbf{r}_v + \mu\nabla_u\mathbf{r}_v.$$

Thus, it's enough to express $\nabla_u\mathbf{r}_u$ and $\nabla_u\mathbf{r}_v$ in terms of the first fundamental form.

Consider $\nabla_u\mathbf{r}_u$. This is the projection of $\mathbf{r}_{uu}$ to the plane spanned by $\mathbf{r}_u$ and $\mathbf{r}_v$. Thus, by Lemma 7.5, its coefficients are expressible in terms of $\mathbf{r}_u.\mathbf{r}_u = E$, $\mathbf{r}_u.\mathbf{r}_v = F$, $\mathbf{r}_v.\mathbf{r}_v = G$ together with $\mathbf{r}_{uu}.\mathbf{r}_u$ and $\mathbf{r}_{uu}.\mathbf{r}_v$. But $\mathbf{r}_{uu}.\mathbf{r}_u = \frac{1}{2}(\mathbf{r}_u.\mathbf{r}_u)_u = \frac{1}{2}E_u$, and one checks that $\mathbf{r}_{uu}.\mathbf{r}_v = F_u - \frac{1}{2}E_v$.

Similarly for $\nabla_u\mathbf{r}_v$, it's enough to note that $\mathbf{r}_{uv}.\mathbf{r}_u = \mathbf{r}_{vu}.\mathbf{r}_u = \frac{1}{2}E_v$ etc. ♢

Lemma 7.6 : *Suppose that* $\mathbf{e}, \mathbf{a}, \mathbf{b}$ *are tangent vector fields on* S. *Then*

$$\nabla_{\mathbf{e}}(\mathbf{a}.\mathbf{b}) = \mathbf{a}.\nabla_{\mathbf{e}}\mathbf{b} + \mathbf{b}.\nabla_{\mathbf{e}}\mathbf{a}.$$

Proof : Exercise. $\diamondsuit$

Suppose that $f : S \longrightarrow \Sigma$ is a diffeomorphism. For each $p \in S$, we have an induced map $f_* : T_p(S) \longrightarrow T_{f(p)}(\Sigma)$ (as described in Section 5). A tangential vector field, $\mathbf{a}$, on S thus gives rise to a tangential vector field on Σ, which we denote by $f_*\mathbf{a}$. If f is an isometry, then it respects the first fundamental forms of S and Σ. If $\mathbf{e}$ and $\mathbf{a}$ are tangential vector fields, then applying Lemma 7.5, it follows that $\nabla_{f_*\mathbf{e}} f_*\mathbf{a} = f_*(\nabla_{\mathbf{e}}\mathbf{a})$. This can be expressed informally by saying that "∇ commutes with f_*".

Let γ be a smooth curve in S. If $\mathbf{a}$ is a tangent vector field on S we define:

$$\frac{D}{dt}\mathbf{a}(\gamma(t)) = \nabla_{\gamma'(t)}\mathbf{a}.$$

We will frequently abbreviate this to $\frac{D}{dt}\mathbf{a}$ when the curve γ is implicitly assumed.

Remark : We only really need $\mathbf{a}$ to be defined on the image of γ for this to make sense. More precisely, if $\mathbf{a}, \mathbf{b}$ are tangential vector fields, and $\mathbf{a}(\gamma(t)) = \mathbf{b}(\gamma(t))$ for all t, then $\frac{D}{dt}\mathbf{a} = \frac{D}{dt}\mathbf{b}$ (exercise).

In this notation, Lemma 7.6 implies that, for any vector fields $\mathbf{a}$, $\mathbf{b}$ on S, we have

$$\frac{d}{dt}(\mathbf{a}.\mathbf{b}) = \mathbf{a}.\frac{D}{dt}\mathbf{b} + \mathbf{b}.\frac{D}{dt}\mathbf{a}$$

(where $\frac{d}{dt}(\mathbf{a}.\mathbf{b})$ is an abbreviation for $\frac{d}{dt}(\mathbf{a}(\gamma(t).\mathbf{b}(\gamma(t)))$.

From the definition, it follows easily that $D_{\gamma'(t)}\mathbf{a} = \frac{d}{dt}(\mathbf{a} \circ \gamma(t))$ and so $\frac{D}{dt}\mathbf{a}$ is the othogonal projection of $\frac{d}{dt}(\mathbf{a} \circ \gamma(t))$ to S, or more precisely, to the tangent space of S at $\gamma(t)$.

Suppose now that s is an arc-length parameterisation of γ. That is $\gamma'(s).\gamma'(s) = 1$, and so, by Lemma 7.6, $\gamma'(s).\frac{D}{ds}\gamma'(s) = 0$. Therefore we can write

$$\frac{D}{ds}\gamma'(s) = \kappa_S \mathbf{N}_S$$

where $\mathbf{N}_S$ is a unit tangent vector in S normal to $\gamma'(s)$.

Definition : κ_S is the *geodesic curvature* of γ in S.

(In fact, it can be given a sign, depending on where $\mathbf{N}_S$ is to the left or right of γ in the orientation of S.)

Here is another description of geodesic curvature. We have observed that $\frac{D}{ds}\gamma'(s)$ is the projection of $\frac{d}{ds}\gamma'(s) = \gamma''(s)$ to $T_{\gamma(s)}S$. But $\gamma''(s) = \kappa_\gamma \mathbf{N}_\gamma$, where $\mathbf{N}_\gamma$ is the (principal) normal to γ and κ_γ is the curvature of γ (as defined in Section 2). Thus, $\kappa_S \mathbf{N}_S$ is the othogonal projection of $\kappa_\gamma \mathbf{N}_\gamma$ to $T_{\gamma(p)}S$.

Examples :

(1) If γ is a curve in the plane, $S = \mathbf{R}^2 \hookrightarrow \mathbf{R}^3$, then $\gamma''(s)$ lies in $\mathbf{R}^2$ so $\kappa_S = \kappa_\gamma$ and $\mathbf{N}_S = \mathbf{N}_\gamma$.

(2) The unit 2-sphere.

The geodesic curvature of the latitude of angle ϕ is $\tan\phi$.

Remark: one can view this in terms of the cone that touches the latitude tangentially.

Parallel transport.

Suppose that $\mathbf{a}$ is a tangential vector field on S, and that γ is a curve in S.

Definition : We say that $\mathbf{a}$ is *parallelly transported* along γ if $\frac{D}{dt}\mathbf{a} = 0$, for all t.

Remark : We only need $\mathbf{a}$ to be defined on the image of γ for this to make sense.

Note that by lemma 7.6, $\frac{D}{dt}(\mathbf{a}.\mathbf{a}) = 2\mathbf{a}.\frac{D}{dt}\mathbf{a} = 0$, and so $||\mathbf{a}(\gamma(t))||$ is constant along γ.

The following give the existence and uniqueness of parallel transport of a vector along a given curve:

Theorem 7.7 : *Suppose that $\gamma : [t_0, t_1] \longrightarrow S$ is a smooth path and $\mathbf{a}_0 \in T_{\gamma(t_0)}(S)$. Then there is a unique smooth vector field, $\mathbf{a}$, along γ, tangential to S, with $\mathbf{a}(t_0) = \mathbf{a}_0$ and $\frac{D}{dt}\mathbf{a}(\gamma(t)) = 0$ for all t.*

Proof : (Sketch) Let $\mathbf{r} : U \longrightarrow S$ be a chart. Assume first that $\gamma([t_0, t_1]) \subseteq \mathbf{r}(U)$. Write $\gamma'(t) = \lambda(t)\mathbf{r}_u + \mu(t)\mathbf{r}_v$. It follows from Lemma 5.4 that λ and μ are smooth. Let $\mathbf{a}(\gamma(t)) = a_1(t)\mathbf{r}_u(\gamma(t)) + a_2(t)\mathbf{r}_v(\gamma(t))$. The functions a_1, a_2 are smooth if and only if $\mathbf{a}$ is smooth (directly from the definition of vector fields). We saw (as in the proof of Theorem 7.4) that $\frac{D}{dt}\mathbf{a}(t) = \nabla_{\gamma(t)}\mathbf{a}(\gamma(t))$ can be expressed as a function of $\lambda, \mu, E, F, G, a_1, a_2$, and their derivatives with respect to u and v. (Following that argument as given there, we need to assume that $\mathbf{a}$ is actually defined on a neighbourhood of the image of γ, but this is not essential to the proof.) Looking closely at the argument we see that the derivatives of a_i with respect to u, v only appear as part of expessions of the form $(a_i)_u\mathbf{r}_u + (a_i)_v\mathbf{r}_v$ and so these can be replaced by $\frac{d}{dt}(a_i(t)) = a_i'(t)$. Moreover, the resulting expression is linear in the terms $a_1(t), a_2(t), a_1'(t), a_2'(t)$. In summary, we get a system of first order linear differential equations for $a_1(t)$ and $a_2(t)$. The coefficients are expressible in terms of $\lambda(t), \mu(t), E(\gamma(t)), E_u(\gamma(t))$ etc. and are hence smooth functions of t. Thus, by Theorem 3.4, they have a unique solution with the given initial condition.

To deal with the general case, we cut γ into a finite number of subpaths, such that the image of each subpath lies in one coordinate patch. (This uses compactness of the closed interval.) We then apply the above inductively to each subpath in turn. $\diamondsuit$

Suppose that γ is a curve parameterised by arc length, s, and that $\mathbf{a}$ is a parallelly transported vector field along γ. Then

$$\frac{D}{ds}(\gamma'(s) - \mathbf{a}(\gamma(s))) = \frac{D}{ds}\gamma'(s) = \kappa_S \mathbf{N}_S.$$

Thus we can think of the geodesic curvature, κ_S as measuring the rate at which the curve turns in relation to parallel transport.

Total turning.

Suppose that γ is a closed curve in a surface S. Then (applying the same definition as in Section 2), we can define the *total turning* of γ as $\int_\gamma \kappa_S(s)ds$.

Examples :
(1) For a curve in the plane $S = \mathbf{R}^2 \hookrightarrow \mathbf{R}^3$, this agrees with the definition given in Section 2.
As an example, we can compute the total turning of a round circle, γ, of radius r as $\int_\gamma \frac{1}{r} ds = (2\pi r)\frac{1}{r} = 2\pi$.
Indeed, we remarked there that the total turning of any Jordan curve was $\pm 2\pi$, depending on orientation.

(2) The ϕ latitude on the unit 2-sphere has length $2\pi \cos\phi$, and hence total turning

$$2\pi \cos\phi \tan\phi = 2\pi \sin\phi.$$

Foucault's pendulum

The direction of oscillation of a pendulum carried along the earth's surface will move by parallel transport. If the suspension of the pendulum is stationary relative to the earth, then in absolute terms it is being carried once around a latitude every day (or every 23 hours and 56 minutes, if one takes into account the earth's motion relative to the sun in that time). The observed angle it turns through is therefore equal to the total turning of the latitude. That is $2\pi \sin\phi$ at latitude ϕ. As expected this is 0 at the equator and $\pm 2\pi$ at the poles.

Foucault hung his pendulum from the dome of the Panthéon in Paris in 1851. It was the first dynamical demonstration of the Earth's rotation. The original pendulum was reinstalled there in 1995. Foucault also suggested the use of a vertical gyroscope to similar effect, though this was harder to implement in practice.

Geodesics.

Definition : A path γ in S is *geodesic* if $\frac{D}{dt}\gamma'(t) = 0$.

This can equivalenty be expressed by saying any one of the following:
(a) The tangent vector is parallelly transported.

(b) The geodesic curvature is identically 0, or

(c) $\gamma''(t)$ is always orthogonal to $T_{\gamma(t)}(S)$.

Note that $||\gamma'(t)||$ is constant, that is, γ has constant speed.

Examples :

It is generally complicated to write out explicit formulae for geodesics. However, certain examples can be seen by symmetry: where the surface admits an (orientation reversing) isometry fixing a given curve, thereby showing that its geodesic curvature must be 0. For example:

Straight lines in the plane.

Great circles of the sphere.

Generators of surfaces of revolution.

Remarks : (We won't prove these here.)

(1) Constant speed length minimising paths are geodesics. In fact any geodesic is locally length minimising.

(2) Given any $p \in S$ and any $\mathbf{x} \in T_p(S)$ there is a geodesic path $\gamma : [-a, a] \longrightarrow \mathbf{R}$ with $\gamma(0) = 0$ and $\gamma'(0) = \mathbf{x}$, where $a > 0$. Any two such paths agree on the intersection of their domains.

(3) We say that S is *complete* if any geodesics can be extended indefinitely (that is, with domain $\mathbf{R}$). If S is complete, then any two points can be connected by a geodesic (the "Hopf-Rinow Theorem"). One can show that if $S \subseteq \mathbf{R}^3$ is closed (as a subset of $\mathbf{R}^3$) then S is complete.

8. Gauss curvature.

Let us recall some definitions.

Let S be a regular surface and $p \in S$. We have defined the linear map:

$$\Theta : T_p(S) \longrightarrow T_p(S)$$

by

$$\mathbf{e} \mapsto -\nabla_{\mathbf{e}}\mathbf{n}$$

where $\mathbf{n}$ the unit normal ambient vector field. (Note that $\nabla_{\mathbf{e}}\mathbf{n} = D_{\mathbf{e}}\mathbf{n}$ since $\mathbf{n}.D_{\mathbf{e}}\mathbf{n} = 0$.) This has two real eigenvalues, κ_1 and κ_2. The Gauss curvature is defined as $\kappa = \kappa_1\kappa_2$. (Our sign convention regarding the shape operator is not going to matter here, since $\kappa = \kappa_1\kappa_2 = (-\kappa_1)(-\kappa_2)$.) We have seen that it can be expressed in terms of the first and second fundamental forms as:

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

In fact, it was shown by Gauss to be intrinsic to S, that is, it depends only on the first fundamenal form (and its derivatives) and is therefore preserved by isometry. He called this a "Whopping Great Theorem" or "Theorema Egregium" in Latin:

Theorem 8.1 : (Theorema Egregium) *The Gauss curvature, κ, depends only on the first fundamental form.*

In fact, if $\mathbf{r} : U \longrightarrow S$ is a coordinate chart, then κ can be expressed in terms of the functions E, F, G and their first and second derivatives with respect to u and v. Recall that $\mathbf{n} = (\mathbf{r}_u \wedge \mathbf{r}_v)/||\mathbf{r}_u \wedge \mathbf{r}_v||$.

The proof we present here is taken from Hitchin's notes [H] (with some elaboration).

Proof : We will consider how the "operator"

$$\nabla_v \nabla_u - \nabla_u \nabla_v$$

behaves when applied to smooth tangential vector fields on S.

Let $\mathbf{a}$ be a tangential vector field. Recall that

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

So

$$\begin{aligned} \nabla_v \nabla_u \mathbf{a} &= \nabla_v (\mathbf{a}_u + (\mathbf{a}.\mathbf{n}_u)\mathbf{n}) \\ &= \nabla_v \mathbf{a}_u + \nabla_v ((\mathbf{a}.\mathbf{n}_u)\mathbf{n}) \\ &= \mathbf{a}_{uv} - (\mathbf{a}_{uv}.\mathbf{n})\mathbf{n} + \nabla_v ((\mathbf{a}.\mathbf{n}_u)\mathbf{n}). \end{aligned}$$

As observed after Lemma 7.3, we have $\nabla_v((\mathbf{a}.\mathbf{n}_u)\mathbf{n}) = (\mathbf{a}.\mathbf{n}_u)\nabla_v \mathbf{n} = (\mathbf{a}.\mathbf{n}_u)\mathbf{n}_v$. (Since $\nabla_v \mathbf{n} = \mathbf{n}_v - (\mathbf{n}_v.\mathbf{n})\mathbf{n} = \mathbf{n}_v$. Thus,

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

Swapping u and v, and using $\mathbf{a}_{uv} = \mathbf{a}_{vu}$, we get:

$$\begin{aligned} \nabla_v \nabla_u \mathbf{a} - \nabla_u \nabla_v \mathbf{a} &= (\mathbf{a}.\mathbf{n}_u)\mathbf{n}_v - (\mathbf{a}.\mathbf{n}_v)\mathbf{n}_u \\ &= (\mathbf{n}_u \wedge \mathbf{n}_v) \wedge \mathbf{a}. \end{aligned}$$

Now $\mathbf{n}_u, \mathbf{n}_v \in T_p(S)$ and so $\mathbf{n}_u \wedge \mathbf{n}_v$ is normal to S. Thus, we can write

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

, where λ is a smooth real valued function on S. We have:

$$(\nabla_v \nabla_u - \nabla_u \nabla_v)\mathbf{a} = \lambda \mathbf{n} \wedge \mathbf{a}.$$

(In other words, $\nabla_u \nabla_v - \nabla_v \nabla_u$ rotates $\mathbf{a}$ though $\pi/2$ in each tangent space, and then scales it by a variable factor of λ.) By Theorem 7.4, the action of $\nabla_v \nabla_u - \nabla_u \nabla_v$ is intrinsic to S, and so therefore is the function λ.

Now

$$\begin{aligned}
\lambda \mathbf{n}.\mathbf{r}_u \wedge \mathbf{r}_v &= (\mathbf{n}_u \wedge \mathbf{n}_v).(\mathbf{r}_u \wedge \mathbf{r}_v) \\
&= (\mathbf{n}_u.\mathbf{r}_u)(\mathbf{n}_v.\mathbf{r}_v) - (\mathbf{n}_u.\mathbf{r}_v)(\mathbf{n}_v.\mathbf{r}_u) \\
&= LN - M^2
\end{aligned}$$

and

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

Thus,

$$\lambda = \frac{LN - M^2}{\sqrt{EG - F^2}}.$$

It follows that

$$LN - M^2 = \lambda \sqrt{EG - F^2}$$

is intrinsic, and so

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

is intrinsic also. $\diamondsuit$

Going through the above argument carefully one can derive an explicit formula for κ, though we won't attempt that here. A formula can be found, for example in [M p.213]. Apparently this is known as "Brioschi's formula".

Note that it follows that if $f : S \longrightarrow \Sigma$ is an isometry, then the Gauss curvature of S at p is equal to the Gauss curvature of Σ at $f(p)$.

For example, we have seen that the Gauss curvature of the sphere is non-zero. Thus, there can be no isometry from the sphere to the plane, even locally. This is why any planar map of the earth's surface must inevitably distort distances.

Surfaces of revolution.

Let $\gamma(t) = (\lambda(t), \mu(t))$ be a curve in the plane with $\lambda(t) > 0$ for all t. Recall that the associated surface of revolution can be defined by:

$$\mathbf{r}(u, v) = (\lambda(u) \cos v, \lambda(u) \sin v, \mu(u)).$$

From the formula $\kappa = (LN - M^2)/(EF - G^2)$ one can derive:

$$\kappa = \frac{\lambda' \mu' \mu'' - \lambda''(\mu')^2}{\lambda((\lambda')^2 + (\mu')^2)^2}.$$

Exercise: or see or [M p.141].

If t is arc-length, then $(\lambda')^2 + (\mu')^2 = 1$ and so $\mu'\mu'' = -\lambda'\lambda''$. In this case, the above formula reduces to:

$$\kappa = -\frac{\lambda''}{\lambda}.$$

As an example, the unit sphere (minus the poles) can be viewed as the surface of revolution of the semicircle $\gamma(t) = (\cos t, \sin t)$, for $t \in (\pi/2, \pi/2)$. In this case, we get $\lambda(t) = \cos t = -\lambda''(t)$, $\kappa = 1$ as before.

The pseudosphere.

The pseudosphere is the surface of revolution of the tractrix, discussed in Section 2. Recall that a unit speed parameterisation of the tractix has the form

$$\gamma(s) = (e^{-s}, y(s))$$

and so

$$\lambda(s) = e^{-s}$$

giving

$$\kappa = -1.$$

In other words, the pseudosphere has constant Gauss curvature -1.

One can also get the same result by a more laborious calculation using the parameterisation

$$\gamma(t) = (\sin t, -\log\tan(t/2) - \cos t)$$

(Exercise, or see [M p142].)

Remark : Note that the pseudosphere is not complete. In fact, it cannot be extended beyond the boundary circle (obtained by spinning $\gamma(0)$) in such a way as to have curvature -1.

In fact, it is a theorem of Hilbert that there is no complete embedded surface in $\mathbf{R}^3$ of constant curvature -1. (See [M p.203].)

This is one reason to develop a theory of "intrinsic geometry" of surfaces without reference to embeddings in 3-dimensional space. This is the subject of riemannian geometry. It is discussed in Section 11.

9. The local Gauss-Bonnet Theorem.

Here we give a local form of the Gauss-Bonnet theorem. This relates the Gauss curvature of a surface to the geodesic curvature of a curve. In the next section we will give a global form for closed surfaces.

Let S be a regular surface, and $\Delta \subseteq S$ be a smooth disc in S. Its boundary, γ, is a smooth closed curve, which we parameterise by arc-length, s. We assume that γ is positively oriented, that is Δ is to the "left" of γ when proceeding in a positive direction.

Recall the definitions of Gauss curvature, κ, of S, and geodesic curvature, κ_S, of γ, from earlier sections.

Theorem 9.1 : (Gauss-Bonnet)

$$\int_\gamma \kappa_S \, ds = 2\pi - \int_\Delta \kappa \, dA.$$

Here dA denotes the area element (as discussed in Section 4). Note that the LHS is the total turning of γ in S as defined in Section 7.

By a *smooth disc* in S, we mean a closed subset, $\Delta \subseteq S$ that is the image of a Jordan domain Ω in the plane under an injective smooth regular map. In fact, it will be convenient to insist that the map is defined on a neighbourhood of the Jordan domain. More precisely, there is a chart, $\mathbf{r} : U \longrightarrow S$ such that $\Delta \subseteq \mathbf{r}(U)$ and $\Omega = \mathbf{r}^{-1}\Delta$ is a Jordan domain. The boundary curve, β, of Ω gets sent to the boundary curve, γ, of Δ under $\mathbf{r}$. Note that, by the smooth Schoenflies Theorem (2.2), there is no loss in assuming that Ω is, in fact, the unit disc.

The argument is again based on Hitchin's notes [H].

Let $\mathbf{r} : U \longrightarrow S$, with $\Delta \subseteq \mathbf{r}(U)$. Let $\Omega = \mathbf{r}^{-1}\Delta$ be a Jordan domain as above. We write β for its boundary, viewed as a smooth curve.

We recall Green's theorem from Section 2. Namely that if P, Q are smooth functions defined on Ω, then

$$\int_\beta (Pu' + Qv') \, dt = \int_\Omega (Q_u - P_v) \, dudv \qquad (*)$$

In what follows, to simplify notation, we will generally suppress mention of the map, $\mathbf{r}$, and just work with (u, v) coordinates on $\mathbf{r}(U) \subseteq S$. We can then apply Green's theorem $(*)$ directly to $\mathbf{r}(U)$.

Proof of Theorem 9.1 : Set $\mathbf{e} = \mathbf{r}_u/\sqrt{E}$. Thus $\mathbf{e}$ is a tangential vector field on $\mathbf{r}(U)$ with $||\mathbf{e}|| = 1$. Let $\mathbf{f} = \mathbf{n} \wedge \mathbf{e}$. Thus, $\mathbf{f}$ is another unit tangential field on $\mathbf{r}(U)$, orthogonal to $\mathbf{e}$. Note that $\mathbf{e}, \mathbf{f}, \mathbf{n}$ is an orthonormal frame in $\mathbf{R}^3$. Now since $\mathbf{e}.\mathbf{e} = 1$, we have $\mathbf{e}.\nabla_u \mathbf{e} = \mathbf{e}.\nabla_v \mathbf{e} = 0$ so we can write

$$\nabla_u \mathbf{e} = P\mathbf{f} \qquad \nabla_v \mathbf{e} = Q\mathbf{f},$$

where P and Q are smooth functions, namely $P = (\nabla_u \mathbf{e}).\mathbf{f}$ and $Q = (\nabla_v \mathbf{e}).\mathbf{f}$.

We parameterise γ by arc length, s, and write $\mathbf{t} = \gamma'(s)$ for the unit tangent. Given any vector field, $\mathbf{a}$, along Γ, we write We write $\mathbf{a}' = \frac{D}{ds}\mathbf{a} = \nabla_{\mathbf{t}}\mathbf{a}$.

Now the left hand side of $(*)$ is:

$$\begin{aligned} \int_\gamma (Pu' + Qv') \, ds &= \int_\gamma (u'\nabla_u \mathbf{e} + v'\nabla_v \mathbf{e}).\mathbf{f} \, ds \\ &= \int_\gamma \mathbf{e}'.\mathbf{f} \, ds \end{aligned} .$$

Since $||\mathbf{t}|| = 1$, we can write $\mathbf{t} = \cos\theta\,\mathbf{e} + \sin\theta\,\mathbf{f}$, where θ is a smooth function on the image of γ. In fact, θ itself is only defined up to addition by an integer multiple of 2π, though the derivative θ' with respect to t, is well defined.

We claim that

$$\mathbf{e}'.\mathbf{f} = \kappa_S - \theta'.$$

To see this, note that

$$\begin{aligned}\mathbf{n}\wedge\mathbf{t} &= \cos\theta\,\mathbf{n}\wedge\mathbf{e} + \sin\theta\,\mathbf{n}\wedge\mathbf{f}\\ &= \cos\theta\,\mathbf{f} - \sin\theta\,\mathbf{e}\end{aligned}$$

is the unit vector in $T_p(S)$ obtained by rotating t through $\pi/2$. Now, differentiating $\mathbf{t}$ using Lemma 7.3, we get

$$\mathbf{t}' = \cos\theta\,\mathbf{e}' - \theta'\sin\theta\,\mathbf{e} + \sin\theta\,\mathbf{f}' + \theta'\cos\theta\,\mathbf{f}.$$

From the definition of geodesic curvature, κ_S, and using $\mathbf{e}.\mathbf{f} = \mathbf{f}'.\mathbf{f} = \mathbf{e}'.\mathbf{f} + \mathbf{e}.\mathbf{f}' = 0$, we get

$$\begin{aligned}\kappa_S &= \mathbf{t}'.(\mathbf{n}\wedge\mathbf{t})\\ &= (\cos^2\theta\,\mathbf{e}'.\mathbf{f} - \sin^2\theta\,\mathbf{e}.\mathbf{f}') + (cos^2\theta + \sin^2\theta)\theta'\\ &= \mathbf{e}'.\mathbf{f} + \theta'.\end{aligned}$$

verifying the claim.

Thus, the left hand side of $(*)$ is:

$$\int_\gamma \mathbf{e}'.\mathbf{f}\,ds = \int_\gamma (\kappa_S - \theta')\,ds.$$

Now $\int_\gamma \theta' ds$ measures the turning of the curve γ relative to the field $\mathbf{e}$, which must be equal to is equal to 2π. (A formal statement of this is given as Lemma 9.2 below.) Thus, the left hand side of $(*)$ becomes:

$$\int_\gamma \kappa_S\,ds - 2\pi.$$

For the right hand side, first note that $\mathbf{f}.\nabla_u\mathbf{f} = 0$, and so $\nabla_u\mathbf{f} = H\mathbf{e}$ for some real valued function, H. But $(\mathbf{e}.\mathbf{f})_u = \mathbf{f}.\nabla_u\mathbf{e} + \mathbf{e}.\nabla_u\mathbf{f} = 0$, and so $P + H = 0$, that is $H = -P$ and we get:

$$\nabla_u\mathbf{f} = -P\mathbf{e}.$$

Similarly

$$\nabla_v\mathbf{f} = -Q\mathbf{e}.$$

Thus:

$$\begin{aligned}\nabla_v\nabla_u\mathbf{e} &= \nabla_v(P\mathbf{f})\\ &= P_v\mathbf{f} + P\nabla_v\mathbf{f}\\ &= P_v\mathbf{f} - PQ\mathbf{e}.\end{aligned}$$

Similarly

$$\nabla_u \nabla_v \mathbf{e} = Q_u \mathbf{f} - PQ\mathbf{e}$$

and so

$$(\nabla_v \nabla_u - \nabla_u \nabla_v)\mathbf{e} = (P_v - Q_u)\mathbf{f}.$$

But as in the proof of Theorem 8.1 we have:

$$(\nabla_v \nabla_u - \nabla_u \nabla_v)\mathbf{e} = \lambda \mathbf{f}$$

where

$$\lambda = \kappa\sqrt{EG - F^2}.$$

In other words,

$$P_v - Q_u = \kappa\sqrt{EG - F^2}$$

and so the right hand side of $(*)$ becomes:

$$-\int_\Delta \kappa\sqrt{EG - F^2}\, dudv.$$

But $\sqrt{EG - F^2}\, dudv$ is precisely the area element in S, and so

$$\int_\Delta \kappa\sqrt{EG - F^2}\, dudv = \int_\Delta \kappa\, dA.$$

Equating the two sides, we have shown:

$$\begin{aligned}\int_\Delta \kappa\, dA &= -\left(\int_\gamma \kappa_S\, ds - 2\pi\right)\\ &= 2\pi - \int_\gamma \kappa_S\, ds.\end{aligned}$$

$\diamondsuit$

Here is a more formal statement regarding the total turning of a curve relative to a vector field, needed in the proof.

Let $\Omega \subseteq S$ is a smooth disc with boundary curve $\gamma : \mathbf{S}^1 \longrightarrow S$ where $\mathbf{S}^1$ is the unit circle. Suppose that $\mathbf{e}$ is a unit tangent vector field on Ω. Given $\phi \in \mathbf{S}^1$, let $\theta(\phi)$ be the angle $\gamma'(\phi)$ makes with $\mathbf{e}(\gamma(\phi))$. More formally, we can write:

$$\gamma'(\phi)/||\gamma'(\phi)|| = \cos\theta\, \mathbf{e} + \sin\theta\, (\mathbf{n} \wedge \mathbf{e}).$$

Note that this determines $\cos\theta$ and $\sin\theta$ completely, and so $\theta(\phi)$ is well defined up to adding some element of $2\pi\mathbf{Z}$. It follows that $\theta'(\phi)$ is completely well defined. (We can assume, for any particular ϕ, that θ is chosen to be continuous in a small neighbourhood of ϕ.) Let $T(\gamma) = \int_\gamma \theta'(\phi)\, d\phi$. (One can check that this is invariant under reparameterisation of γ.) We can think of it as the turning of γ relative to the vector field $\mathbf{e}$.

Lemma 9.2 : $T(\gamma) = 2\pi$.

Proof : (Sketch) [There must be a better way of doing this...]

We begin by observing that $T(\gamma) \in 2\pi\mathbf{Z}$. To see this, we parameterise $\mathbf{S}^1$ by radial angle ϕ, varying between 0 and 2π. We fix a particular value of $\theta(0)$. We define $\theta(\phi) = \theta(0) + \int_0^{2\pi} \theta'(\phi)\, d\phi$. This gives a particular value of $\theta(\phi)$ for all $\phi \in [0, 2\pi]$. By definition of the closed curve integral, we have $T(\gamma) = \int_0^{2\pi} \theta'(\phi)\, d\phi = \theta(2\pi) - \theta(0)$. Now, from the definition of θ we see that $\gamma'(\phi)/||\gamma'(\phi)|| = \cos\theta(\phi)\, \mathbf{e} + \sin\theta(\phi)\, (\mathbf{n} \wedge \mathbf{e})$ holds for all $\phi \in [0, 2\pi]$. In particular, $\cos\theta(0) = \cos\theta(2\pi)$ and $\sin\theta(0) = \sin\theta(2\pi)$, and so $\theta(2\pi) - \theta(0) \in 2\pi\mathbf{Z}$ as claimed.

Let $D = \{\mathbf{x} \in \mathbf{R} \mid ||\mathbf{x}|| \leq 1\}$. Let ρ, ϕ be polar co-ordinates. Then there is a smooth injective map $\mathbf{h} : D \longrightarrow S$ which is a diffeomorphism to $\mathbf{h}(D)$. Given $\rho \in (0, 1]$, define $\gamma_\rho : \mathbf{S}^1 \longrightarrow S$ by $\gamma_\rho(\theta) = \mathbf{h}(\rho, \phi)$. This is regular curve. Let $\theta_\rho(\phi)$ be the angle between $\gamma'_\rho(\theta)$. and $\mathbf{e}(\gamma_\rho(\theta))$. Let $T(\gamma) = \int_{\gamma_\rho} \theta'_\rho(\phi)\, d\phi$. As for γ, we see that $T(\gamma_\rho) \in 2\pi\mathbf{Z}$.

Note that $[\rho \longrightarrow T(\gamma_\rho)]$ is continuous, and always lies in $2\pi\mathbf{Z}$. Thus, $T(\gamma_\rho) = T(\gamma)$ for all $\rho \in (0, 1]$.

Now let $c = \mathbf{h}(0) \in S$, and let $\mathbf{h}_* : \mathbf{R}^2 \longrightarrow T_c(S)$ be the derivative at this point. As $\rho \to 0$, $\theta_\rho(\phi)$ converges uniformly to $\theta_0(\phi)$, where $\theta_0(\phi)$ is the angle $h_*(t(\phi))$ makes with the fixed tangent vector, $\mathbf{e}(c)$, where $t(\phi) = (-\sin\phi, \cos\phi)$. (Note that $t(\phi)$ is the unit tangent to $\mathbf{S}^1$ at ϕ. For given ϕ, the unit tangent vector $\gamma'(\phi)/||\gamma'(\phi)||$ tends to $h_*(t(\phi))$.) As ϕ goes once around S^1, so does $\theta'_0(\phi)$. (This can be written out explicity, given that $\mathbf{h}_*$ is just a linear map. It is the statement that the total turning of an ellipse in the plane is equal to 2π.) So $T(\gamma_\rho) \to \int_{\mathbf{S}^1} \theta'_0(\phi)\, d\phi = 2\pi$, and so $T(\gamma_\rho) = 2\pi$ for all ρ. ♢

Examples : (of the local Gauss-Bonnet theorem).

(1) A Jordan curve in the plane.

In this case $\kappa = 0$ and we get $\int_\gamma \kappa_S ds = 2\pi$ which is just the formula for total turning.

(2) Latitude of the unit sphere.

Let γ be the ϕ-latitude. This bounds the northerly spherical cap, Δ. We have already seen that the total turning of γ is:

$$\int_\gamma \kappa_S\, ds = 2\pi \sin\phi.$$

We also have

$$\text{area}(\Delta) = 2\pi(1 - \sin\phi).$$

(This can be conveniently seen using radial projection to a cylinder, namely Lambert's projection discussed in Section 5. We saw this was area preserving, and so the area of the cap is the area of the portion of the cylinder lying above the horizontal plane of through the latitude, i.e. of height $\sin\phi$.)

Now $\kappa = 1$ and so:

$$\begin{aligned}\int_\Delta \kappa\, dA &= \text{area}(\Delta) \\ &= 2\pi - 2\pi\sin\theta \\ &= 2\pi - \int_\gamma \kappa_S\, ds\end{aligned}$$

which agrees with the Gauss-Bonnet formula.

Piecewise smooth curves.

We finish this section by remarking that the main result can be generalised to a piecewise smooth Jordan curve. If γ is such a curve, we can define its exterior angles $\phi_1, \phi_2, \ldots, \phi_n \in [-\pi, \pi]$, at the vertices, cf. the discussion of plane curves in Section 2. (The angles are defined with respect to the first fundamental form on S.) We can define the total turning, $\Theta(\gamma)$, of γ as $\int_\gamma \kappa_S\, ds + \sum_{i=1}^n \phi_i$. Suppose that γ bounds a disc Δ. The Gauss-Bonnet theorem holds in the form $\Theta(\gamma) = 2\pi - \int_\Delta \kappa\, dA$. It can be proven, for example, by approximating γ by smooth curves, though the details are a bit technical. (Here the term "disc" should be interpreted to be the image of a disc under an injective continuous map. Such a map can be taken to be smooth away from the vertices. We will not worry too much about these technicalities here.)

10. Some global properties of surfaces.

Much of the discussion here will be less than rigorous. The necessary formal background could be found in courses on measure theory and differential geometry.

The following bit about orientations could have come earlier, but we've not needed it until now.

Orientations.

Let $U, V \subseteq \mathbf{R}^2$ be open. Suppose $\sigma : U \longrightarrow V$ is a diffeomorphism. Given $x \in U$, write $D(x) = \det \sigma_*(x)$, i.e. the determinant of the jacobian. Note that $D(x) \neq 0$ for all x.

Definition : We say that σ is *orientation preserving* (resp. *orientation reversing*) if $D(x) > 0$ for all $x \in U$ (resp. $D(x) < 0$ for all $x \in U$).

We note that U is connected if and only if any two points of U can be connected by a smooth path. (This is an exercise if you know about the general notion of connectedness, otherwise it can serve as a definition here.)

Exercise : If U is connected, then V is connected, and any such diffeomorphism σ is either orientation preserving or reversing.

Definition : Suppose $\mathbf{r} : V \longrightarrow \mathbf{R}^3$ is a parameterised surface, and $\mathbf{s} = \mathbf{r} \circ \sigma$ is a reparameterisation. We say they are *consistently oriented* if σ is orientation preserving.

An atlas is *oriented* if all pairs charts are consistently oriented on their overlap.

S is *orientable* if it admits an oriented atlas.

Lemma 10.1 : *S is orientable if and only if there is a smooth ambient normal field,* $\mathbf{n}$*. (i.e. $\mathbf{n}(p)$ is perpendicular to $T_p(S)$ for all $p \in S$).*

Note that we can always assume $\mathbf{n}$ to be a unit normal field.

We have noted that any surface has a locally defined normal field, i.e. defined in the neighbourhood of any point. The point here is that it can be defined globally, i.e. on all of S.

Proof : Suppose S is orientable, and choose an oriented atlas. For each chart $\mathbf{r}$ set $\mathbf{n}(\mathbf{r}((u,v))) = (\mathbf{r}_u \wedge \mathbf{r}_v)/||\mathbf{r}_u \wedge \mathbf{r}_v||$. This is well defined on the overlaps.
We leave the converse as an exercise. $\diamondsuit$

By an *orientation* on S, we mean a unit normal field.

Remark : Orientation can be expressed purely in terms of the atlas of S.

Definition : We say that S is *smoothly path connected* if and only if every pair of points are connected by a smooth path.

Exercise : (for topologists) S is smoothly path connected if any only if it is path connected. (If you don't know what this means, don't worry.)

Exercise : If S is smoothly path connected and orientable then it has exactly two orientations.

The following terminology is confusing, but standard.

Definition : A *closed surface* in $\mathbf{R}^3$ is surface $S \subseteq \mathbf{R}^3$ that is closed (in the usual topological sense), bounded, and smoothly path connected.

Examples : Sphere, torus etc.

Remark : (for topologists) A subset of $\mathbf{R}^3$ is (topologically) closed and bounded if and only if it is compact. In this case, S is smoothly path connected if and only if it is path connected and if and only if it is connected.

Fact : A any closed surface in $\mathbf{R}^3$ is orientable.

If we don't assume closed, this is false: Möbius band.

In fact, if S is a closed surface in $\mathbf{R}^3$, then $\mathbf{R}^3 \setminus S$ has exactly two components – one bounded "inside" and another unbounded "outside". This requires some topology to prove.

Euler characteristic

This can be defined for a topological space as an alternating sum of Betti numbers — the dimensions of the homology groups with real coefficients. For a closed surface S, it can be described as follows:

A *graph* embedded in S consists of a finite set, $V \subseteq S$, of "vertices", and a finite set, E, of "edges", such that each edge is a smooth regular injective path connecting two distinct points of V and meeting V only at its endpoints, and such that and any two distinct edges meet at most at their endpoints. We write Γ for the union of the edges. A *face* is a component of $S \setminus \Gamma$. We write F for the set of faces. Let $v = |V|$, $e = |E|$ and $f = |F|$. We say that Γ is a *cellulation* if the closure of each face is homeomorphic to a disc. A *triangulation* is a cellulation in which each face has exactly three edges in its boundary.

Fact : A closed surface admits a triangulation. (In fact, we can assume that each triangle lies in a co-ordinate patch.)

Fact : The number

$$\chi = v - e + f$$

is same for any cellulation.

If you believe these facts, then this gives us a well defined number, $\chi = \chi(S)$ called the *Euler characteristic*

Examples : The Euler characteristic of the sphere is 2 ("Euler's formula") and that of the torus is 0.

In fact we get an infinite sequence of surfaces with Euler characteristic 2, 0, -2, -4....

Theorem 10.2 : (Classification of closed (orientable) surfaces) *The Euler characteristic of any closed surface is one of the above numbers (i.e.* $2 - 2g$ *for some* $g \in \mathbf{N}$*. (i.e.* $2 - 2g$ *for some* $g \in \mathbf{N}$*. Any two surfaces with the same Euler characteristic are diffeomorphic.*

In other words any closed surface is diffeomorphic to exacty one of the above examples: i.e. a surface with g "holes", where g is called the "genus" and $\chi = 2 - 2g$.

Proof : For the general idea, see a book on topology (though these don't always discuss diffeomorphisms). $\diamondsuit$

Remarks :

(1) We say two surfaces, S and Σ are "homeomorphic" if there is a continuous bijective map $f : S \longrightarrow \Sigma$ such that f^{-1} is also continuous. It follows that they have the same euler characterictic and are hence diffeomorphic. (i.e. we can replace f by another map h such that h and h^{-1} are both smooth.

(2) If S is diffeomorphic to a 2-sphere then there is a diffeomorphism of $\mathbf{R}^3$ to itself sending S to the round sphere. (This is the 3-dimensional "Schoenflies theorem".) The analogous statement is also true an $(n-1)$-sphere in n space for $n \geq 5$ but unresolved for $n = 4$. This fails for higher genus sufaces: we can knot tori etc.

Note that if we have a triangulation, then $2e = 3f$ and so

$$\chi = v - \frac{3}{2}f + f = v - \frac{f}{2}.$$

Integration.

Suppose $f : S \longrightarrow \mathbf{R}$ is a smooth function, and $Q \subseteq S$ is a "nice subset" with $Q \subseteq \mathbf{r}(U)$ for some chart $\mathbf{r} : U \longrightarrow S$. We define:

$$\int_Q f\, dA = \int_{\mathbf{r}^{-1}Q} f \circ \mathbf{r}(u, v)\, ||\mathbf{r}_u \wedge \mathbf{r}_v||\, dudv.$$

Here dA is the "area element". (Recall that $||\mathbf{r}_u - \mathbf{r}_v|| = \sqrt{EG - F^2}$, where E, F, G are the coefficients of the first fundamental form.) In the above, "nice" means "measurable" in the sense of measure theory. It is sufficient that Q be open or closed, for example.

Exercise : This is independent of the chart: If $Q \subseteq \mathbf{s}(V)$, where $\mathbf{s} : V \longrightarrow S$ is another chart, then

$$\int_{\mathbf{r}^{-1}Q} f \circ \mathbf{r}(u, v)\, ||\mathbf{r}_u \wedge \mathbf{r}_v||\, dudv = \int_{\mathbf{s}^{-1}Q} f \circ \mathbf{s}(\eta, \theta)\, ||\mathbf{s}_\eta \wedge \mathbf{s}_\theta||\, d\eta d\theta.$$

Now given a triangulation of S, we can define:

$$\int_S f\, dA = \sum_\Delta \int_\Delta f\, dA$$

as Δ ranges over all faces of the triangulation.

Fact : This is independent of the choice of triangulation.

One would expect this at least to be invariant under subdivision of the triangulation. If it were true that any two triangulations had a common subdivision, the fact would then follow. Unfortunately, life is more complicated. To do this properly it would be better to define integration using "partitions of unity" instead of triangulations — see the differential geometry course.

Theorem 10.3 : (Gauss-Bonnet) Let S be a closed surface. Let $\kappa : S \longrightarrow \mathbf{R}$ be the Gauss curvature. Then

$$\int_S \kappa \, dA = 2\pi\chi(S).$$

Proof : Let V, E, F be the sets of vertices, edges and faces of some triangulation of S. Suppose that $\Delta \in F$. Let $\theta_1, \theta_2, \theta_3$ be the interior angles. Then the total turning of $\partial\Delta$ is

$$3\pi - \theta_1 - \theta_2 - \theta_3 + \sum_\epsilon \int_\epsilon \kappa_S \, ds,$$

where ϵ ranges over the three edges of Δ, positively oriented, κ_S is the geodesic curvature, and s is arc length. Thus, the local version of the Gauss-Bonnet theorem (as discussed at the end of Section 9) gives:

$$\int_\Delta \kappa \, dA = \theta_1 + \theta_2 + \theta_3 - \pi - \sum_\epsilon \int_\epsilon \kappa_S \, ds.$$

We now sum over all $\Delta \in F$. The LHS becomes $\int_S \kappa \, dA$. For the RHS, summing the angles gives a contribution of 2π for each vertex of V, that is, giving $2\pi v$ in total. Summing the π's gives $-\pi f$. The path integrals cancel out: for each edge of E, we get two contributions, one for each of the two incident faces. But these are oriented in opposite directions, and so the signs of the geodesic curvatures are opposite. Equating, we get:

$$\int_S \kappa \, dA = 2\pi v - \pi f = 2\pi(v - f/2) = 2\pi\chi$$

as required. ◊

Examples :

(1) For example for the round 2-sphere, $\kappa \equiv 1$, and so we recover the formula:

$$\text{area}(\mathbf{S}^2) = (2\pi)2 = 4\pi.$$

(2) Suppose that $\gamma(u) = (\lambda(u), \mu(u))$ is a closed embedded curve in the plane, with $\lambda(u) > 0$ for all u. The surface of revolution, S, is diffeomorphic to a torus, so $\chi(S) = 0$. If u is arc length, recall that $\kappa(u, v) = -\lambda(u)''/\lambda(u)$ and the f.f.f. is

$$du^2 + \lambda(u)^2 dv^2$$

so the area element is $dA = \lambda\, du\, dv$. Thus
$$\int_S \kappa\, dA = -\int_S \lambda''(u)\, du\, dv = -2\pi \int_\gamma \gamma''(u) du = 0$$
since we are intergrating around a closed curve.

11. 2-manifolds.

This section may or may not be examinable. We'll decide depending on how much time we have at the end of the course.

Here's how to define a surface (or "2-manifold") without reference to embeddings in $\mathbf{R}^3$.

Let Σ be a metric space. (We're only really interested in it as a topological space.)

Definition : A *(topological) chart* is a bijective map $r : U \longrightarrow S$, where $U \subseteq \mathbf{R}^2$ and $r(U)$ are open, and r and r^{-1} are continuous. (i.e r is a homeomorphism from U to $r(U)$.) The image of a chart is a *patch*.

An *(topological) atlas* is a collection of charts such that each point lies in a patch.

Σ is a *topological surface* if it admits an atlas.

If $r : U \longrightarrow \Sigma$, and $s : V \longrightarrow \Sigma$ are two charts, we refer to the map
$$s \circ r^{-1} : V \cap s^{-1}(r(U)) \longrightarrow U \cap r^{-1}(s(V))$$
as a *transition map*.

(Note that a transition map is a homeomorphism.)

Definition : An atlas is *smooth* if every transition function is smooth.

A *smooth surface* is a surface together with a smooth atlas.

Fact : Every topological surface admits a smooth atlas.

Examples :

(1) The plane, $\mathbf{R}^2$, is a smooth surface. We only need one chart: the identity map. Any open subset is also a surface.

(2) Any regular surface is a smooth surface. We can take the atlas to consist of the set of all smooth charts (as defined earlier). We use Lemma 5.3.

(3) Define the torus as the quotient of $\mathbf{R}^2$ under the action of $\mathbf{Z}^2$ (i.e. $(x, y) \sim (z, w)$ if $x - z, y - w \in \mathbf{Z}$). Take an atlas to consist of all inclusions of open discs of radius less than $1/4$ composed with the quotient map. All transition functions are translations of subsets of the plane, and are in particular smooth. This is a smooth atlas. (We refer to this construction as the "square torus".)

Definition : A map $f : S \longrightarrow \Sigma$ between smooth surfaces is smooth if $s^{-1} \circ f \circ r$ is smooth for all charts r and s of S and Σ respectively, wherever and whenever this is defined.

We say f is a *diffeomorphism* if it is bijective and both f and f^{-1} are smooth. We say that S and Σ are *diffeomorphic* if there is a diffeomorphism between them.

Fact : If two surfaces are homeomorphic, then they are diffeomorphic. (The homeomorphism we first thought of need not be smooth of course, so this is highly non-trivial.)

Let $\beta : I \longrightarrow \Sigma$ be a continuous path.

Definition : β is *smooth* if $r^{-1} \circ \beta$ is smooth for all charts r, whenever and wherever it's defined.

Note, for regular surfaces, this agrees with the definition already given, using Lemma 5.4.

Tangent spaces.

Suppose we have two smooth paths, $\beta : I \longrightarrow \mathbf{R}^2$ and $\gamma : J \longrightarrow \mathbf{R}^2$, with $0 \in \operatorname{int} I \cap \operatorname{int} J$.

Fix some $a \in \mathbf{R}^2$, and let $C(a)$ be the set of curves $\beta : I \longrightarrow \mathbf{R}^2$ with $0 \in \operatorname{int} I$ and with $\beta(0) = \mathbf{R}^2$. Given $\beta, \gamma \in C(a)$, write $\beta \sim \gamma$ if $\beta'(0) = \gamma'(0)$.

It's easy to see that this an equivalence relation with $C(a)/\sim$ identified with $\mathbf{R}^2$ via $\beta \leftrightarrow \beta(0)$.

Suppose $a \in U \subseteq \mathbf{R}^2$ is open and $\sigma : U \longrightarrow \mathbf{R}^2$ and $\sigma : U \longrightarrow V$ is a diffeomorphism. Let $b = \sigma(a) \in V$. We identify the tangent spaces at a and b with $\mathbf{R}^2$. Let $\sigma_* : \mathbf{R}^2 \longrightarrow \mathbf{R}^2$ be the derivative at a.

Lemma 11.1 : *If $\beta \sim \gamma \in C(a)$ then $\sigma \circ \beta \sim \sigma \circ \gamma \in C(b)$. Moreover, under the idenfications $C(a)/\sim \equiv C(a)/\sim \equiv \mathbf{R}^2$, σ_* sends the equivalence class $[\beta]$ to $[\sigma \circ \beta]$.*

Proof : Exercise. $\diamondsuit$

Now suppose that Σ is a smooth surface and $p \in \Sigma$. Let $C(p)$ be the set of smooth curves $\beta : I \longrightarrow S$ with $\gamma(0) = p$. Define a relation $\sim$ on $C(p)$ as follows. Let $r : U \longrightarrow \Sigma$ be a chart with $p \in r(U)$. Let $a = r^{-1}p \in U$. If $\beta, \gamma \in C(p)$, we write $\beta \sim \gamma$ if $r^{-1} \circ \beta \sim r^{-1} \circ \gamma$ where these are defined.

Exercise : This is well defined, i.e. independent of the choice of chart.

We write $T_p(\Sigma) = C(p)/\sim$.

Note that we can identify $T_p(\Sigma)$ with $C(a)/\sim \equiv \mathbf{R}^2$. If s is another chart, with $s^{-1}p = b$, we can also identify $T_p(\Sigma)$ with $C(b)/\sim \equiv \mathbf{R}^2$. The compositions of these idenfications, as a map $\mathbf{R}^2 \longrightarrow \mathbf{R}^2$, is precisely σ_*, where σ is the transition map. Since σ_* is linear, we see that the vector spaces stuctures agree, and so $T_p(\Sigma)$ has a natural structure as a vector space.

Definition : $T_p(\Sigma)$ is the *tangent space* of Σ at p.

Note that if β is a smooth path in Σ, then we have a well defined tangent vector $\beta'(t) \in T_{\beta(t)}\Sigma$.

Exercise : If Σ is a regular surface in $\mathbf{R}^3$, we can naturally identify $T_p(\Sigma)$ with the tangent space already defined.

Exercise : If $f : \Sigma \longrightarrow S$ is smooth, then there is natural linear *derivative* map $f_* : T_p(\Sigma) \longrightarrow T_{f(p)}(S)$.

We can now go on to redefine many of notions used for regular sufaces in $\mathbf{R}^2$ by interpreting them in terms of charts. For example:

Definition : A *riemannian metric* on Σ associates to each tangent space a positive definite symmetric bilinear form such that it varies smoothly over Σ.
A *riemannian surface* is a smooth surface with a riemannian metric. [Note, this is not the same thing as a "Riemann surface"!]

(To define smoothness, here, we refer back to coordinate charts and coordinates, and observe that it is independent of choices of charts.)

Examples :
(1) If Σ is a regular surface, then the first fundamental form is a riemannian metric.
(2) Suppose that $U \subseteq \mathbf{R}^2$ is open, and $\lambda : U \longrightarrow (0, \infty)$ is smooth, then we can define a riemannian metric as $\lambda(u,v)^2(du^2 + dv^2)$. (i.e., the form is given by the matix λI.) In other words, we rescale the euclidan metric locally by λ.
(3) The square torus described above has a metric defined by $du^2 + dv^2$, where u, v are the euclidean coordinates (Check this is well-defined.) This is locally isometric to (or "modelled on") the euclidian plane — every point as a neigbourhood isometric to a subset of $\mathbf{R}^2$.

A riemannian metric on Σ allows us to define lengths of smooth curves, angles, area, etc, using the same formulae as before, and reinterpreting the f.f.f. as the riemannian metric. One needs to check that they are well defined, independently of the charts, but that follows by essentially the same argument.

Example : In example (2) above, the length of a curve, $\beta : I \longrightarrow U$ is given by

$$\int_I \lambda(\beta(t)) ||\beta'(t)|| \, dt$$

Definition : An *isometry* between two riemannian surfaces Σ and S is a smooth map $f : \Sigma \longrightarrow S$ such that for all $p \in \Sigma$, f_* sends the riemannian form on $T_p(\Sigma)$ to that on $T_{f(p)}(S)$.

An isometry respects lengths, angles and area. (In fact, it's equivalent to length-preserving.) $\blacksquare$

We can now go on to define parallel transport and Gauss curvature. One (very inefficient) way to do this would be to go carefully through the proofs that these were intrinsic properties of regular surfaces, derive explicit formulae for these in terms of the f.f.f. use the same formulae as definitions in the riemannian case. One needs to check that they are independent of charts. In principle, this will work, but there are better ways. This is the subject of a course in riemannian geometry.

One has (the same) classification of compact orientable surfaces. The Gauss-Bonnet theorem remains true. (For example, for the euclidean square torus above, we get 0 on both sides.) It also works in the non-orientable case.

Remark : For almost everything we said we can replace 2 by n, and talk about smooth and riemannian n-manifolds. These are not classified, and curvature is more subtle. There is a generalisation of the "Gauss-Bonnet" theorem when n is even.

Example : (Hyperbolic plane). Let $D = \{\mathbf{x} \in \mathbf{R}^2 \mid ||\mathbf{x}|| < 1\}$. Rescale the metric by a factor of

$$\lambda(\mathbf{x}) = \frac{2}{1 - ||\mathbf{x}||^2}$$

as above. This is the Poincaré model of the hyperbolic plane. We denote it $\mathbf{H}^2$.

Exercise : All rotations of the disc are ismetries of $\mathbf{H}^2$.

Facts :
(1) $\mathbf{H}^2$ is homogeneous, that is, given any $x, y \in \mathbf{H}^2$, there is an isometry taking x to y.
(2) $\mathbf{H}^2$ has constant Gauss curvature -1.
(3) $\mathbf{H}^2$ cannot be realised as a regular surface in $\mathbf{R}^3$. (Though bits of it can — the pseudosphere is locally isometric to $\mathbf{H}^2$.)

Waffle : Any surface of genus at least 2 admits a metric of constant curvature -1, locally modelled on $\mathbf{H}^2$. (cf. flat torus described above).

These surfaces cannot be realised as regular surfaces in $\mathbf{R}^3$.

THE END