Math 162B - Notes

Neil Donaldson

March 10, 2009

1 Exterior calculus

1.1 Wedge products and *n***-forms**

Recall the notion of a 1-form α on \mathbb{R}^n : if x_1, \ldots, x_n are co-ordinates on \mathbb{R}^n , then $\alpha = \sum_{i=1}^n a_i \, dx_i$ is a 1-form, where $a_1, \ldots, a_n : \mathbb{R}^n \to \mathbb{R}$ are smooth functions.

We introduce an operation \land on 1-forms which satisfies the properties

 $\mathrm{d} x_i \wedge \mathrm{d} x_j = - \mathrm{d} x_i \wedge \mathrm{d} x_i, \qquad \mathrm{d} x_i \wedge \mathrm{d} x_i = 0,$

and study the algebra generated by 1-forms under \wedge .

Definition 1.1. A *multi-index* of length *k* is a list of numbers $I = (i_1, ..., i_k)$, $1 \le i_1, ..., i_k \le n$ (when in \mathbb{R}^n). We write

$$\mathrm{d} x_I = \mathrm{d} x_{i_1} \wedge \cdots \wedge \mathrm{d} x_{i_k}.$$

A multi-index is *increasing* if $p < q \Rightarrow i_p < i_q$. A *k*-form α on \mathbb{R}^n is an object of the form

$$\alpha = \sum_{\substack{\text{All multi-indices } I \\ \text{of length } k}} \alpha_I \, \mathrm{d} x_I,$$

where each $\alpha_I : \mathbb{R}^n \to \mathbb{R}$ is a smooth function. We also call α a *form of degree k*.

At the moment the symbol \land is just a symbol for denoting *k*-forms. It will shortly become an operation. It is clear that at the expense of a minus sign, every multi-index can be put in increasing order. This is the standard way of writing *k*-forms. The definition also includes 0-forms: with no multi-indices, a 0-form is just a function on \mathbb{R}^n .

Example.

 \mathbb{R}^2 \mathbb{R}^3 \mathbb{R}^4 k function *f* f f 0 1 $\alpha_1 dx_1 + \alpha_2 dx_2 \qquad \alpha_1 dx_1 + \alpha_2 dx_2 + \alpha_3 dx_3$ $\alpha_1 dx_1 + \cdots + \alpha_4 dx_4$ $\beta_1 dx_1 \wedge dx_2 + \cdots + \beta_6 dx_3 \wedge dx_4$ $\beta_1 dx_1 \wedge dx_2 + \cdots + \beta_3 dx_2 \wedge dx_3$ 2 $\beta dx_1 \wedge dx_2$ 3 None $\gamma \, \mathrm{d} x_1 \wedge \mathrm{d} x_2 \wedge \mathrm{d} x_3$ $\gamma_1 dx_1 \wedge dx_2 \wedge dx_3 + \cdots + \gamma_4 dx_2 \wedge dx_3 \wedge dx_4$ 4 None None $\delta dx_1 \wedge dx_2 \wedge dx_3 \wedge dx_4$

Generally, there are $\binom{n}{k}$ components for a *k*-form on \mathbb{R}^n . Note that there are no *k*-forms on \mathbb{R}^n for k > n. The set of *k*-forms at $p \in U \subset \mathbb{R}^n$ form a vector space: this should be clear if you view *k*-forms as alternating *k*-linear maps. If x_1, \ldots, x_n are co-ordinates on *U*, then the set of *k*-forms $\{ dx_{i_1} \land \cdots \land dx_{i_k} : i_1 < \cdots < i_k \}$ forms a basis of this vector space which thus has dimension $\binom{n}{k}$. More concretely, if $\{\alpha_1, \ldots, \alpha_n\}$ is a basis of 1-forms at *p*, then the set $\{\alpha_i \land \alpha_j : i < j\}$ is a basis of the set of 2-forms at *p*, and $\{\alpha_{i_1} \land \cdots \land \alpha_{i_k} : i_1 \leq \cdots \leq i_k \}$ a basis of the set of *k*-forms at *p*.

Definition 1.2. Suppose that $k + l \le n$. The *wedge product* of a *k*-form α and an *l*-form β on \mathbb{R}^n is the (k + l)-form $\alpha \land \beta$ and is formed in the obvious way.

The 'obvious' way here can get complicated for large degree forms as the following formula shows: if $\alpha = \sum \alpha_I dx_I$ and $\beta = \sum \beta_I dx_I$ are *k*- and *l*-forms respectively, then

$$\alpha \wedge \beta = \sum_{\substack{\text{All multi-indices } I, J \\ \text{of lengths } k, l \\ \text{respectively}}} \alpha_I \beta_J \, dx_I \wedge dx_J,$$

where $dx_I \wedge dx_J = dx_{i_1} \wedge \cdots \wedge dx_{i_k} \wedge dx_{j_1} \wedge \cdots \wedge dx_{j_l}$. Thankfully in this course, we will rarely consider higher than 3-forms.

Examples. 1. $\alpha = 2 dx_1 - 3x_1 dx_2$ and $\beta = (1 - x_2^2) dx_1 + x_2 dx_2$ are 1-forms on \mathbb{R}^2 . Then

$$\begin{split} \alpha \wedge \beta &= (2 \,\mathrm{d} x_1 - 3 x_1 \,\mathrm{d} x_2) \wedge ((1 - x_2^2) \,\mathrm{d} x_1 + x_2 \,\mathrm{d} x_2) \\ &= 2(1 - x_2^2) \,\mathrm{d} x_1 \wedge \mathrm{d} x_1 + 2 x_2 \,\mathrm{d} x_1 \wedge \mathrm{d} x_2 - 3 x_1(1 - x_2^2) \,\mathrm{d} x_2 \wedge \mathrm{d} x_1 - 3 x_2 \,\mathrm{d} x_2 \wedge \mathrm{d} x_2 \\ &= (2 x_2 - 3 x_1 + 3 x_1 x_2^2) \,\mathrm{d} x_1 \wedge \mathrm{d} x_2. \end{split}$$

2. $\alpha = dx_1 + 2 dx_2 + x_1 dx_3$ and $\beta = 3x_3 dx_1 \wedge dx_2 - dx_2 \wedge dx_3$ are 1- and 2-forms on \mathbb{R}^3 respectively. Here we have

$$\alpha \wedge \beta = (3x_1x_3 - 1) dx_1 \wedge dx_2 \wedge dx_3.$$

Proposition 1.3. *If* α , β *are forms, then*

$$\alpha \wedge \beta = (-1)^{\deg \alpha \deg \beta} \beta \wedge \alpha.$$

Note that a function *f* is a 0-form and that $f \wedge \alpha = f\alpha = \alpha \wedge f$, so that the proposition still holds.

1.2 The exterior derivative

We are used to thinking of d of a function: of a co-ordinate function as dx_1 say, or of a more general function on an open set U; $df = \sum \frac{\partial f}{\partial x_i} dx_i$. d is a more general operator on forms of any degree.

Definition 1.4. Given a *k*-form $\alpha = \sum_{I} \alpha_{I} dx_{I}$ on \mathbb{R}^{n} , the exterior derivative of α is the (k + 1)-form

$$\mathrm{d}\alpha=\sum_{I}\mathrm{d}\alpha_{I}\wedge\mathrm{d}x_{I},$$

where $d\alpha_I$ is ' d' of a function.

Example. In \mathbb{R}^3 , let $\alpha = x_1 x_2^2 dx_1 - x_2 dx_3$. Then

$$d\alpha = d(x_1x_2^2) \wedge dx_1 - d(x_2) \wedge dx_3 = (x_2^2 dx_1 + 2x_1x_2 dx_2) \wedge dx_1 - dx_2 \wedge dx_3$$

= $-2x_1x_2 dx_1 \wedge dx_2 - dx_2 \wedge dx_3$.

Proposition 1.5. *For all forms* α *,* β *we have*

- 1. $d(\alpha + \beta) = d\alpha + d\beta$.
- 2. $d(\alpha \wedge \beta) = d\alpha \wedge \beta + (-1)^{\deg \alpha} \alpha \wedge d\beta$.
- 3. $d(d\alpha) = 0$.

The final result is often written as $d^2 = 0$.

Example. Let $f(x_1, x_2) = x_1^2 x_2$ on \mathbb{R}^2 . Then $df = 2x_1 x_2 dx_1 + x_1^2 dx_2$, and so,

$$d(df) = d(2x_1x_2 dx_1 + x_1^2 dx_2)$$

= 2 d(x_1x_2) \lapha dx_1 + d(x_1^2) \lapha dx_2
= 2x_2 dx_1 \lapha dx_1 + 2x_1 dx_2 \lapha dx_1 + 2x_1 dx_1 \lapha dx_2 = 0.

Co-ordinate invariance in \mathbb{R}^2

One of the main advantages that comes with thinking about forms is that they have an inbuilt coordinate invariance; otherwise said, when you change co-ordinates, forms automatically change in the correct way. Here is an example in $\mathbb{R}^2 \setminus \{0\}$.



Figure 1: Polar co-ordinates

With respect to the standard co-ordinates x, y, any 1-form may be written $\alpha = a \, dx + b \, dy$. The same 1-form may also be written $\alpha = A \, dr + B \, d\theta$ in polar co-ordinates. Similarly, any 2-form is a multiple of $dx \wedge dy$ and simultaneously a multiple of $dr \wedge d\theta$. These must correspond somehow. Indeed

$$\begin{array}{l} x = r\cos\theta \\ y = r\sin\theta \end{array} \Rightarrow \begin{array}{l} dx = \cos\theta \, dr - r\sin\theta \, d\theta \\ dy = \sin\theta \, dr + r\cos\theta \, d\theta. \\ \therefore a \, dx + b \, dy = r(a\cos\theta \, dr + b\sin\theta \, d\theta). \end{array}$$

Similarly,

$$dr = \frac{x}{\sqrt{x^2 + y^2}} \, dx + \frac{y}{\sqrt{x^2 + y^2}} \, dy, \qquad d\theta = \frac{-y}{\sqrt{x^2 + y^2}} \, dx + \frac{x}{x^2 + y^2} \, dy.$$

Hence

$$\mathrm{d}x \wedge \mathrm{d}y = \cos\theta \cdot r \cos\theta \,\mathrm{d}r \wedge \mathrm{d}\theta - r \sin\theta \cdot \sin\theta \,\mathrm{d}\theta \wedge \mathrm{d}r = r \,\mathrm{d}r \wedge \mathrm{d}\theta.$$

Recall this from change of variables in integration: if $f(x, y) = g(r, \theta)$, then

$$\int f(x,y) \, \mathrm{d}x \, \mathrm{d}y = \int g(r,\theta) r \, \mathrm{d}r \, \mathrm{d}\theta.$$

The change of variables from integration is thus already built into forms. We will come back to forms and integration earlier, though we've already seen its first steps: 1-forms are integrated along curves, 2-forms will be what we integrate over surfaces, 3-forms over 3-dimensional regions, etc..

1.3 Forms as multi-linear maps

A *k*-form α at a point $p \in \mathbb{R}^n$ is a multi-linear, alternating map

$$\alpha|_p: \underbrace{T_p \mathbb{R}^n \times \cdots \times T_p \mathbb{R}^n}_{k \text{ times}} \to \mathbb{R}$$

from *k* copies of the tangent space at *p* to \mathbb{R} . The following formula can be taken either as a definition of a *k*-form, or as the definition of the determinant in terms of forms, depending on your preference. If $\alpha = \alpha_1 \wedge \cdots \wedge \alpha_k$ is a *k*-form where each α_i is a 1-form, and $v_1, \ldots, v_k \in T_p \mathbb{R}^n$ are vectors, then

$$\alpha(v_1,\ldots,v_k) = \begin{vmatrix} \alpha_1(v_1) & \cdots & \alpha_1(v_k) \\ \alpha_2(v_1) & \cdots & \alpha_2(v_k) \\ \vdots & & \vdots \\ \alpha_k(v_1) & \cdots & \alpha_k(v_k) \end{vmatrix}.$$

Alternating means that if you swap the positions of any two of the v_i , the result changes sign. This is equivalent to swapping two rows in the determinant.

Example. Let $\beta = dx_1 \wedge dx_2 + x_3 dx_2 \wedge dx_3$ be a 2-form on \mathbb{R}^3 , and let $u = \frac{\partial}{\partial x_1} - \frac{\partial}{\partial x_3}$, $v = x_2 \frac{\partial}{\partial x_2} + \frac{\partial}{\partial x_3}$. Now,

$$(dx_1 \wedge dx_2)(u,v) = \begin{vmatrix} dx_1(u) & dx_1(v) \\ dx_2(u) & dx_2(v) \end{vmatrix} = x_2, \qquad (dx_2 \wedge dx_3)(u,v) = x_2$$

$$\therefore \beta(u,v) = x_2(1+x_3).$$

Definition 1.6. In \mathbb{E}^2 , the 2-form $dx \wedge dy$ is the *standard area form*. Indeed if **u**, **v** are two vectors, then

$$\mathrm{d}x \wedge \mathrm{d}y(\mathbf{u}, \mathbf{v}) = \begin{vmatrix} u_1 & v_1 \\ u_2 & v_2 \end{vmatrix} = u_1 v_2 - v_1 u_2$$

is the signed area of the parallelogram spanned by **u**, **v**.

Similarly $dx \wedge dy \wedge dz$ is the standard volume form on \mathbb{E}^3 . These terms will become clearer when we study integration.

1.4 An aside on vector calculus

The standard vector calculus operations of div, grad and curl in \mathbb{E}^3 are closely related to 'd'. For example, the curl of a vector field $\mathbf{v} = \alpha_1 \mathbf{i} + \alpha_2 \mathbf{j} + \alpha_3 \mathbf{k}$ is

$$\nabla \times \mathbf{v} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ \alpha_1 & \alpha_2 & \alpha_3 \end{vmatrix} = \left(\frac{\partial \alpha_3}{\partial y} - \frac{\partial \alpha_2}{\partial z} \right) \mathbf{i} + \left(\frac{\partial \alpha_1}{\partial z} - \frac{\partial \alpha_3}{\partial x} \right) \mathbf{j} + \left(\frac{\partial \alpha_2}{\partial x} - \frac{\partial \alpha_1}{\partial y} \right) \mathbf{k},$$

while the exterior derivative of the 1-form $\alpha = \alpha_1 dx + \alpha_2 dy + \alpha_3 dz$ is

$$d\alpha = \left(\frac{\partial \alpha_2}{\partial x} - \frac{\partial \alpha_1}{\partial y}\right) dx \wedge dy - \left(\frac{\partial \alpha_1}{\partial z} - \frac{\partial \alpha_3}{\partial x}\right) dx \wedge dz + \left(\frac{\partial \alpha_3}{\partial y} - \frac{\partial \alpha_2}{\partial z}\right) dy \wedge dz.$$

Comparing coefficients gives part of the following table: start on any line and compare what d does to the form with what the corresponding vector calculus operation does to the object on the right hand side.

Forms

Traditional vector fields

$$\begin{array}{cccc} \text{function} f & \longleftrightarrow & f \\ & & & & & \\ & & & & \\ & & & & \\ & & & \\ 1 \text{-form } \alpha_1 \, dx + \alpha_2 \, dy + \alpha_3 \, dz & \longleftrightarrow & & \\ & & & & \\ & & & & \\ & &$$

The single 'd' operator is grad, div and curl all in one!

The differential form notation has two distinct advantages over traditional vector calculus: it works in all co-ordinate systems and all dimensions.

The forms result $d^2 = 0$ translates to the 2 theorems,

$$abla imes (
abla f) = 0, \qquad
abla \cdot (
abla imes \mathbf{v}) = 0.$$

With a little calculation it can be seen that the wedge product of 1-forms translates to taking the cross product of vectors, while the wedge product of a 1-form and a 2-form corresponds to taking the dot product. Complicated formulae from vector calculus can be easily proved this way; e.g. let f be a function and α a 1-form. Then,

$$d(f \wedge \alpha) = df \wedge \alpha + d\alpha \wedge f = df \wedge \alpha + f d\alpha$$

$$\uparrow$$

$$\nabla \times f\mathbf{v} = \nabla f \times \mathbf{v} + f \nabla \times \mathbf{v}.$$

Similarly, if α , β are 1-forms, we have

$$d(\alpha \wedge \beta) = d\alpha \wedge \beta - \alpha \wedge d\beta$$

$$\uparrow$$

$$\nabla \cdot (\mathbf{u} \times \mathbf{v}) = (\nabla \times \mathbf{u}) \cdot \mathbf{v} - \mathbf{u} \cdot (\nabla \times \mathbf{v}).$$

2 Moving frames and the structure equations

2.1 Maps $\mathbb{R}^m \to \mathbb{E}^3$

Definition 2.1. A *moving frame* for a smooth map $\mathbf{x} : U \subset \mathbb{R}^m \to \mathbb{E}^3$, written

$$\mathbf{x}(u_1,\ldots,u_m) = \begin{pmatrix} x_1(u_1,\ldots,u_m) \\ x_2(u_1,\ldots,u_m) \\ x_3(u_1,\ldots,u_m) \end{pmatrix},$$

is a triple of maps $\mathbf{e}_i : U \to \mathbb{E}^3$, i = 1, 2, 3, such that $(\mathbf{e}_1(p), \mathbf{e}_2(p), \mathbf{e}_3(p))$ is an oriented orthonormal basis of \mathbb{E}^3 for each $p = (u_1, \ldots, u_m) \in U$.



Figure 2: A moving frame

A moving frame is usually chosen in a way that is suited to the map **x**.

Examples. 1. m = 1. $\mathbf{x} : U \to \mathbb{E}^3$ is a smooth curve. If \mathbf{x} is biregular, we can choose $(\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3) = (\mathbf{T}, \mathbf{N}, \mathbf{B})$ to be the Frenet frame.

2. m = 3. If **x** is the cylindrical polar co-ordinate map

$$\mathbf{x}:(r,\phi,z)\mapsto \begin{pmatrix} r\cos\phi\\ r\sin\phi\\ z \end{pmatrix}$$

then it would be sensible to choose the moving frame

$$\mathbf{e}_r := \frac{\partial \mathbf{x}}{\partial r} = \begin{pmatrix} \cos \phi \\ \sin \phi \\ 0 \end{pmatrix}, \qquad \mathbf{e}_\phi := r^{-1} \frac{\partial \mathbf{x}}{\partial \phi} = \begin{pmatrix} -\sin \phi \\ \cos \phi \\ 0 \end{pmatrix}, \qquad \mathbf{e}_z = \frac{\partial \mathbf{x}}{\partial z} = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix},$$

so that the moving frame's axes point in the directions of change with respect to the three variables.¹

3. m = 2. **x** is a parameterized surface. One direction at least is specified: fix an orientation on the surface and choose a moving frame with $\mathbf{e}_3 = \mathbf{U}$, the unit normal.

Definition 2.2. If $\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$: $U \to \mathbb{E}^3$ is a smooth function, the we write dx for its exterior deriva-

tive:

$$\mathbf{dx} = \begin{pmatrix} \mathbf{dx}_1 \\ \mathbf{dx}_2 \\ \mathbf{dx}_3 \end{pmatrix}$$

dx is a *matrix of 1-forms*, or an \mathbb{E}^3 -valued 1-form, since it maps tangent vectors in U to vectors in \mathbb{E}^3 .

Example. Recalling the cylindrical co-ordinate map, we have

$$d\mathbf{x} = \begin{pmatrix} \cos\phi \, dr - r \sin\phi \, d\phi \\ \sin\phi \, dr + r \cos\phi \, d\phi \\ dz \end{pmatrix} = \mathbf{x}_r \, dr + \mathbf{x}_\phi \, d\phi + \mathbf{x}_z \, dz = \mathbf{e}_r \, dr + r \mathbf{e}_\phi \, d\phi + \mathbf{e}_z \, dz.$$

2.2 Connection forms and the structure equations

In this section we consider a map $\mathbf{x} : U \subset \mathbb{R}^m \to \mathbb{E}^3$, and a moving frame $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$. Instead of simply thinking about how the map \mathbf{x} changes (i.e. about $d\mathbf{x}$), we split the problem into two: describe how \mathbf{x} moves with respect to the moving frame, and describe how the frame itself moves. If the frame is chosen sensibly with respect to the map, then the answer to the first problem should be simple, and we transfer the difficulty to thinking about how the frame moves. While this may seem to increase the complexity, it in fact improves matters, even allowing the application of group theory to the problem.²

First we define 1-forms on U which encode how x changes with respect to the moving frame:

$$\theta_k := \mathbf{d} \mathbf{x} \cdot \mathbf{e}_k, \ k = 1, 2, 3.$$

Proposition 2.3. $d\mathbf{x} = \sum_{k=1}^{3} \theta_k \mathbf{e}_k$.

Proof. Since, at each point $p \in U$, \mathbf{e}_1 , \mathbf{e}_2 , \mathbf{e}_3 form a basis of \mathbb{E}^3 , it is clear that $d\mathbf{x} = \sum_{k=1}^3 (d\mathbf{x} \cdot \mathbf{e}_k) \mathbf{e}_k = \sum_{k=1}^3 \theta_k \mathbf{e}_k$.

Examples. 1. If $\mathbf{x} : U \to \mathbb{E}^3$ is a parameterized curve $\mathbf{x}(t)$, then $d\mathbf{x} = \mathbf{x}' dt$. If we take $(\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3) = (\mathbf{T}, \mathbf{N}, \mathbf{B})$ to be the Frenet frame, then $\theta_1 = |\mathbf{x}'| dt$, while $\theta_2 = \theta_3 = 0$.

2. In the cylindrical co-ordinate example

$$\theta_1 = \mathbf{d} \mathbf{x} \cdot \mathbf{e}_1 = \mathbf{d} r, \quad \theta_2 = r \, \mathbf{d} \phi, \quad \theta_3 = \mathbf{d} z.$$

¹The fact that these directions are orthogonal is a special property of cylindrical co-ordinates which is not true for general co-ordinate systems.

²Indeed this process and its generalizations to higher dimensions and more specialized situations is one of the main ways that group theory finds applications in Physics.

If we now assemble the moving frame into a row matrix $\mathbf{E} = (\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3)$, and introduce the vector of 1-forms $\theta = \begin{pmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \end{pmatrix}$, then Proposition 2.3 can be written,

$$d\mathbf{x} = (\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3) \begin{pmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \end{pmatrix} = \mathbf{E}\theta.$$
(2.1)

Now we consider how the frame moves with respect to itself.

Definition 2.4. Define the *connection forms* ω_{ik} of the moving frame **E** by,

$$\omega_{ik} = \mathbf{e}_i \cdot \mathbf{d} \mathbf{e}_k,$$

where \cdot is the standard dot product on \mathbb{E}^3 .

Lemma 2.5. $\omega_{jk} = -\omega_{kj}$ (hence $\omega_{jj} = 0$) and moreover,

$$\mathbf{d}\mathbf{e}_k = \sum_{j=1}^3 \mathbf{e}_j \omega_{jk}.$$

Proof. Just apply 'd' to the dot product $\mathbf{e}_{j} \cdot \mathbf{e}_{k} = \delta_{jk}$.

The proof of the following theorem is then a simple calculation from Lemma 2.5.

Theorem 2.6. Let ω be the 3 \times 3 skew-symmetric matrix of 1-forms,

$$\omega = \begin{pmatrix} 0 & \omega_{12} & \omega_{13} \\ -\omega_{12} & 0 & \omega_{23} \\ -\omega_{13} & -\omega_{23} & 0 \end{pmatrix}.$$

Then the term-by-term exterior derivative of the moving frame is given by,

$$\mathbf{dE} = (\mathbf{de}_1, \mathbf{de}_2, \mathbf{de}_3) = \mathbf{E}\omega. \tag{2.2}$$

Compare (2.1) and (2.2): the first writes the differential of the map x in terms of the frame E, while the second writes the differential of the frame with respect to itself.

Examples. 1. Returning once again to our cylindrical co-ordinate example, we have,

$$d\mathbf{e}_{3} = 0 \Rightarrow \omega_{13} = \mathbf{e}_{1} \cdot d\mathbf{e}_{3} = 0 = \omega_{23} = \mathbf{e}_{2} \cdot d\mathbf{e}_{3},$$
$$d\mathbf{e}_{2} = \begin{pmatrix} -\cos\phi \, d\phi \\ -\sin\phi \, d\phi \\ 0 \end{pmatrix} \Rightarrow \omega_{12} = -d\phi.$$

2. Let $\mathbf{x}(t)$ be a biregular curve with speed v, curvature κ , torsion τ , and choose the moving frame $(\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3) = (\mathbf{T}, \mathbf{N}, \mathbf{B})$ to be the Frenet frame of \mathbf{x} . In the language of this section,

$$d\mathbf{x} = \frac{d\mathbf{x}}{dt} dt = \mathbf{x}' dt = v\mathbf{T} dt = v\mathbf{e}_1 dt \Rightarrow \theta_1 = v dt, \quad \theta_2 = \theta_3 = 0.$$

Moreover, the Frenet–Serret equations imply that,

 $\omega_{12} = -v\kappa \,\mathrm{d}t, \quad \omega_{13} = 0, \quad \omega_{23} = -v\tau \,\mathrm{d}t.$

In this context, Theorem 2.6 is exactly the Frenet–Serret equations multiplied through by the 1-form dt.

Inspired by the link to the Frenet equations for a curve, we now turn to the analogues for the map **x**.

Theorem 2.7. θ and ω satisfy the first structure equations $d\theta + \omega \wedge \theta = 0$. This is equivalent to the three equations,

$$d\theta_1 + \omega_{12} \wedge \theta_2 + \omega_{13} \wedge \theta_3 = 0,$$

$$d\theta_2 - \omega_{12} \wedge \theta_1 + \omega_{23} \wedge \theta_3 = 0,$$

$$d\theta_3 - \omega_{13} \wedge \theta_1 - \omega_{23} \wedge \theta_2 = 0.$$

Proof. Since $d^2 = 0$, we have,

$$0 = d^{2}\mathbf{x} = d(d\mathbf{x}) = d(\mathbf{E}\theta) = d\mathbf{E} \wedge \theta + \mathbf{E} d\theta = \mathbf{E}(\omega \wedge \theta + d\theta).$$

Writing $\underline{\alpha} = d\theta + \omega \wedge \theta$ and multiplying out, we see that $\sum_{i=1}^{3} \mathbf{e}_{i} \underline{\alpha}_{ij} = 0$ for each j = 1, 2, 3. The linear independence of the \mathbf{e}_{i} forces all 9 coefficients $\underline{\alpha}_{ij}$ to be zero, hence $d\theta + \omega \wedge \theta = 0$.

Theorem 2.8. The connection form ω satisfies the second structure equations $d\omega + \omega \wedge \omega = 0$. This is equivalent to the three equations,

$$\begin{aligned} &\mathrm{d}\omega_{12}-\omega_{13}\wedge\omega_{23}=0,\\ &\mathrm{d}\omega_{13}+\omega_{12}\wedge\omega_{23}=0,\\ &\mathrm{d}\omega_{23}-\omega_{12}\wedge\omega_{13}=0. \end{aligned}$$

Proof. $0 = d(dE) = d(E\omega) = dE \wedge \omega + E d\omega = E\omega \wedge \omega + E d\omega$.

The first and second structure equations are easier to remember if we use the fact that $\omega_{ij} = -\omega_{ji}$, for then they read,

$$d\theta_i + \sum_{j \neq i} \omega_{ij} \wedge \theta_j = 0, \tag{2.3}$$

$$d\omega_{ij} + \omega_{ik} \wedge \omega_{kj} = 0, \quad i, j, k \text{ distinct.}$$
(2.4)

It is straightforward to see the forms for our cylindrical co-ordinate example satisfy both structure equations. It is less easy to see that the ω_{ij} are in fact determined by the structure equations and the θ_i . In this example we have,

$$\theta_1 = dr, \quad \theta_2 = r d\phi, \quad \theta_3 = dz, \quad d\theta_1 = d\theta_3 = 0, \quad d\theta_2 = dr \wedge d\phi.$$

The first structure equations then give us,

$$\left. \begin{array}{l} \mathrm{d}\theta_1 = 0 = -\omega_{12} \wedge r \, \mathrm{d}\phi - \omega_{13} \wedge \mathrm{d}z \\ \mathrm{d}\theta_2 = \mathrm{d}r \wedge \mathrm{d}\phi = \omega_{12} \wedge \mathrm{d}r - \omega_{23} \wedge \mathrm{d}z \\ \mathrm{d}\theta_3 = 0 = \omega_{13} \wedge \mathrm{d}r + \omega_{23} \wedge r \, \mathrm{d}\phi \end{array} \right\}.$$

A little thought leads us through the following:

- ω_{13} is a combination of $d\phi$ and dz by the first equation, and of dr and $d\phi$ by the third. Since dr, $d\phi$, dz are linearly independent at each point we must have ω_{13} a multiple of $d\phi$ only. Write $\omega_{13} = ra d\phi$ for some function *a*.
- The first and third equations now say that $\omega_{12} = b d\phi + a dz$, $\omega_{23} = a dr + c d\phi$, for some functions *b*, *c*.
- Plugging all this into the second equation yields

$$dr \wedge d\phi = b \, d\phi \wedge dr + a \, dz \wedge dr - a \, dr \wedge dz - c \, d\phi \wedge dz,$$

$$\therefore (1+b) \, dr \wedge d\phi + 2a \, dr \wedge dz + c \, d\phi \wedge dz = 0.$$

• The linear independence of the above three 2-forms at each point guarantees that all coefficients are zero so that a = c = 0 and b = -1, thus recovering ω_{ij} .

The point of the above exercise is to see that the connection 1-forms of a moving frame are generally determined directly by the forms θ_i . In practical examples (surfaces later for instance), the θ_i 's are synonymous with the induced metric of the surface. This method of imposing a metric (choice of first fundamental form), and calculating the connection 1-forms from it is critical in physical applications — we shall do a little of this later. It will be seen that the Gauss curvature of a surface can be calculated directly from the connection 1-forms. The generalization of this to higher dimensions is the method by which the *Riemann curvature tensor* of the (Levi–Civita) *connection* of a metric is calculated. Indeed the relationship between a surface/manifold, metric, connection and Gauss/Riemann curvature is precisely what Physicists are talking about when they say that 'Spacetime is curved'.

The structure equations are important in the same way that the Frenet–Serret equations are important for curves: they tell you everything there is to know about a moving frame. As such, it is a standard method in differential geometry to use the method of the moving frame, reducing geometric problems to differential equations. The trick, of course, is to choose your frame so that the equations are not too difficult. The following Theorem may be regarded as the analogue of the Fundamental Theorem of biregular spacecurves.

Theorem 2.9. Let the domain U be simply connected. Given forms ω_{jk} satisfying the second structure equations, and given a frame $\mathbf{e}_1(p)$, $\mathbf{e}_2(p)$, $\mathbf{e}_3(p)$ at a point $p \in U$, there exists a unique moving frame \mathbf{e}_1 , \mathbf{e}_2 , \mathbf{e}_3 on U which agrees with the given frame at p and has the ω_{jk} as connection forms. Furthermore, if we are also given forms θ_k satisfying the first structure equations and one specifies $\mathbf{x}(p)$, then there is a unique map $\mathbf{x}: U \to \mathbb{E}^3$ with $\mathbf{x}(p)$ as specified, and for which $\theta_k = \mathbf{d}\mathbf{x} \cdot \mathbf{e}_k$.

The structure equations in other dimensions

The first and second structure equations are equations relating 2-forms. Since these vanish when m = 1, the structure equations tell us nothing about curves.

In the case of maps $\mathbf{x} : U \subset \mathbb{R}^m \to \mathbb{E}^2$, any moving frame has only two directions $\mathbf{e}_1, \mathbf{e}_2$, hence there is only one connection 1-form ω_{12} . The first structure equations then reduce to,

$$\mathrm{d} heta_1 + \omega_{12} \wedge heta_2 = 0 = \mathrm{d} heta_2 - \omega_{12} \wedge heta_1.$$

When m = 2, we have θ_1 , θ_2 forming a basis of 1-forms at each point, hence $\omega_{12} = a\theta_1 + b\theta_2$ for some functions *a*, *b*. Plugging this into the first structure equations yields

$$\mathrm{d}\theta_1 + a\theta_1 \wedge \theta_2 = 0 = \mathrm{d}\theta_2 + b\theta_1 \wedge \theta_2.$$

Since $d\theta_1$, $d\theta_2$ are also multiples of $\theta_1 \wedge \theta_2$, it is clear that *a*, *b* are uniquely determined by the first structure equations, and thus so is ω_{12} . The second structure equation(s) simply read $d\omega_{12} = 0$.

In higher dimensions the first structure equation stays exactly as in (2.3), while the second becomes only slightly more complicated, with (2.4) being replaced with

$$\mathrm{d}\omega_{ij} + \sum_{k
eq i,j} \omega_{ik} \wedge \omega_{kj} = 0.$$

Understanding this one equation in a given geometric context is the key to understanding that geometry.

Group theory and differential geometry One of the main places group theory appears in geometry is in the study of moving frames. We have seen that knowing a moving frame together with the 1-forms θ is equivalent to knowing the map **x**. Thus it is often desirable to study the frame itself as a single object: but where does it live? In our examples, we are think about the frame $\mathbf{E} = (\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3)$ as taking values in the set of 1×3 matrices whose three entries form an orthonormal basis of \mathbb{E}^3 . If instead we think of the \mathbf{e}_i as column vectors with respect to some (indeed any!) fixed basis of \mathbb{E}^3 , then **E** may be viewed as taking values in the Special Orthogonal group SO(3). As such, the study of maps into \mathbb{E}^3 is often reduced to the study of maps into SO(3). This idea can be generalized and many different groups can be considered as the universe of choice for a problem.

2.3 Surfaces and moving frames

Definition 2.10. Let $\mathbf{x} : U \to \mathbb{E}^3$ be an oriented local surface. An *adaptive frame* to the surface is any moving frame s.t. $\mathbf{e}_3 = \mathbf{U}$.

Note that there are an enormous number of adaptive frames since we are still free to rotate the other two basis vectors about **U**.

In an adaptive frame, $\theta_3 = 0$ and thus $d\mathbf{x} = \mathbf{e}_1\theta_1 + \mathbf{e}_2\theta_2$. The first and second structure equations now become the *structure equations for a surface*,

$egin{array}{l} \mathrm{d} heta_1+\omega_{12}\wedge heta_2=0\ \mathrm{d} heta_2-\omega_{12}\wedge heta_1=0 \end{array}$	First structure equations
$\omega_{13} \wedge \theta_1 + \omega_{23} \wedge \theta_2 = 0$	The Symmetry equation
$\mathrm{d}\omega_{12} - \omega_{13} \wedge \omega_{23} = 0$	The Gauss equation
$\mathrm{d}\omega_{13} + \omega_{12} \wedge \omega_{23} = 0$	The Codezzi equations
$\mathrm{d}\omega_{23} - \omega_{12} \wedge \omega_{13} = 0$	The Couazzi equations.

Furthermore, we have:



Figure 3: An adaptive frame on the sphere

Proposition 2.11. In an adaptive frame,

$$\mathbf{I} = \theta_1^2 + \theta_2^2, \quad \mathbf{I} = -(\theta_1 \omega_{13} + \theta_2 \omega_{23}).$$

Proof.

$$\mathbf{I} = \mathbf{d}\mathbf{x} \cdot \mathbf{d}\mathbf{x} = |\theta_1 \mathbf{e}_1 + \theta_2 \mathbf{e}_2|^2 = \theta_1^2 + \theta_2^2.$$
$$\mathbf{I} = -\mathbf{d}\mathbf{x} \cdot \mathbf{d}U = -(\theta_1 \mathbf{e}_1 + \theta_2 \mathbf{e}_2) \cdot \mathbf{d}\mathbf{e}_3 = -(\theta_1 \mathbf{e}_1 + \theta_2 \mathbf{e}_2) \cdot (\mathbf{e}_1 \omega_{13} + \mathbf{e}_2 \omega_{23}) = -(\theta_1 \omega_{13} + \theta_2 \omega_{23}).$$

Example. Consider the sphere $x^2 + y^2 + z^2 = 1$ parameterized by

$$\mathbf{x} = \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} \sin\theta\cos\phi \\ \sin\theta\sin\phi \\ \cos\theta \end{pmatrix}, \ \theta \in (0,\pi), \ \phi \in [0,2\pi).$$

together with the moving frame

$$\mathbf{e}_1 = \begin{pmatrix} \cos\theta\cos\phi\\ \cos\theta\sin\phi\\ -\sin\theta \end{pmatrix}, \quad \mathbf{e}_2 = \begin{pmatrix} -\sin\phi\\ \cos\phi\\ 0 \end{pmatrix}, \quad \mathbf{e}_3 = \mathbf{U} = \begin{pmatrix} \sin\theta\cos\phi\\ \sin\theta\sin\phi\\ \cos\theta \end{pmatrix}.$$

Here $d\mathbf{x} = \mathbf{e}_1 d\theta + \sin \theta \mathbf{e}_2 d\phi$ so that $\theta_1 = d\theta$, $\theta_2 = \sin \theta d\phi$. It is easy to see that

 $\omega_{12} = -\cos\theta \,\mathrm{d}\phi, \qquad \omega_{13} = \,\mathrm{d}\theta, \qquad \omega_{23} = \sin\theta \,\mathrm{d}\phi,$

from which you can check that the structure equations are satisfied. Moreover, using the θ_i , ω_{ij} we see from the proposition that,

$$\mathbf{I} = d\theta^2 + \sin^2 d\phi^2, \qquad \mathbf{I} = -(d\theta^2 + \sin^2 d\phi^2),$$

which are exactly the expressions obtained in 162A. This frame is plotted in Figure 3.

Since **x** is a local surface we have that the differential d**x** is a 1–1 linear map at each point: i.e. $d\mathbf{x}|_p : T_p U \rightarrow \{\text{tangent vectors to } \mathbf{x} \text{ at } p\}$ is a bijective linear map for each *p*. It follows that θ_1, θ_2 form a basis of the space of 1-forms at each point. This suggests the following:

Lemma 2.12. *There exist unique functions a, b, c s.t.,*

$$\omega_{13} = a\theta_1 + b\theta_2, \quad \omega_{23} = b\theta_1 + c\theta_2$$

Proof. That $\omega_{13}a\theta_1 + b\theta_2$, $\omega_{23} = \hat{b}\theta_1 + c\theta_2$ are linear combinations of θ_1, θ_2 is automatic. The symmetry equation then implies that

$$0 = \omega_{13} \wedge \theta_1 + \omega_{23} \wedge \theta_2 = (-b+b)\theta_1 \wedge \theta_2.$$

The above gives us that $\mathbb{I} = -a\theta_1^2 - 2b\theta_1\theta_2 - c\theta_1^2$.

Proposition 2.13. The Gauss and mean curvatures are given by,

$$K = ac - b^2$$
, $H = -\frac{1}{2}(a + c)$.

Proof. First we construct a dual basis to θ_1 , θ_2 . The map $z \mapsto d\mathbf{x}(z)$ is invertible at each point, so let v_1 , v_2 be the vector fields on U defined such that $d\mathbf{x}(v_i) = \mathbf{e}_i$ for i = 1, 2. Writing $d\mathbf{x} = \mathbf{e}_1\theta_1 + \mathbf{e}_2\theta_2$ implies that

$$\theta_i(v_j) = \delta_{ij} = \begin{cases} 1 & i = j, \\ 0 & i \neq j. \end{cases}$$

 v_1 , v_2 are orthonormal with respect to I.

Now we solve for the Gauss and mean curvatures with respect to the basis v_1 , v_2 at each point. The matrices of I, II with respect to this basis are the identity and $\begin{pmatrix} -a & -b \\ -b & -c \end{pmatrix}$ respectively. We find the principal curvatures (eigenvalues of II with respect to I) by solving,

$$0 = \det\left(\begin{pmatrix} -a & -b \\ -b & -c \end{pmatrix} - \lambda \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right) = (a+\lambda)(c+\lambda) + b^2 = \lambda^2 + (a+c)\lambda + ac - b^2.$$

Recalling that $(\lambda - k_1)(\lambda - k_2) = \lambda^2 - 2H\lambda + K$ gives the required expressions for *K* and *H*.

Since our exterior calculus doesn't mention co-ordinates at all, it is clear from this proposition that *K*, *H* are independent of any choice of co-ordinates on a surface.

With our expressions for ω_{13} , ω_{23} we have

$$\omega_{13} \wedge \omega_{23} = (a\theta_1 + b\theta_2)(b\theta_1 + c\theta_2) = (ac - b^2)\theta_1 \wedge \theta_2,$$

from which follows:

Theorem 2.14. The Gauss equation is equivalent to,

$$\mathrm{d}\omega_{12}=K\,\theta_1\wedge\theta_2.$$

The theorem often gives a faster method of calculating the Gauss curvature of a surface than using the linear algebra method from 162A. In particular, you should observe that we need only calculate 1-forms that are related to the *tangent* part of the moving frame: θ_1 , θ_2 decompose dx in terms of the tangent vectors \mathbf{e}_1 , \mathbf{e}_2 , while ω_{12} describes how that tangent vector fields \mathbf{e}_1 , \mathbf{e}_2 change with respect to each other. The unit normal \mathbf{e}_3 doesn't need to be considered, or calculated. Here are a couple of...

Examples. 1. With the unit sphere,

$$\mathrm{d}\omega_{12} = \sin\theta \,\mathrm{d}\theta \wedge \,\mathrm{d}\phi = \theta_1 \wedge \theta_2 \Rightarrow K = 1.$$

2. The unit cylinder parameterized by co-ordinates (ϕ, z) has the moving frame

$$\mathbf{e}_1 = \begin{pmatrix} -\sin\phi\\\cos\phi\\0 \end{pmatrix}, \quad \mathbf{e}_2 = \begin{pmatrix} 0\\0\\1 \end{pmatrix}, \quad \mathbf{e}_3 = \mathbf{e}_1 \times \mathbf{e}_2 \text{ (don't need),}$$

from which we see that $\omega_{12} = 0$, so that $d\omega_{12}$, and thus *K*, are both zero.

3. Consider the catenoid

$$\mathbf{x}(u,\phi) = \begin{pmatrix} \cosh u \cos \phi \\ \cosh u \sin \phi \\ u \end{pmatrix}$$

A moving frame here is

$$\mathbf{e}_1 = \frac{1}{\cosh u} \begin{pmatrix} \sinh u \cos \phi \\ \sinh u \sin \phi \\ 1 \end{pmatrix}, \qquad \mathbf{e}_2 = \begin{pmatrix} -\sin \phi \\ \cos \phi \\ 0 \end{pmatrix}, \qquad \mathbf{e}_3 = \frac{1}{\cosh u} \begin{pmatrix} -\cos \phi \\ -\sin \phi \\ \sinh u \end{pmatrix}.$$

Again, we needn't have calculated \mathbf{e}_3 . Thus $\omega_{12} = \mathbf{e}_1 \cdot d\mathbf{e}_2 = -\tanh u \, d\phi$. But here $d\mathbf{x} = \cosh u \, (\mathbf{e}_1 \, du + \mathbf{e}_2 \, d\phi)$, so that $\theta_1 = \cosh u \, du$ and $\theta_2 = \cosh u \, d\phi$. The Gauss equation says that,

$$\mathrm{d}\omega_{12} = -\operatorname{sech}^2 u \,\mathrm{d}u \wedge \mathrm{d}\phi = -\operatorname{sech}^4 u\theta_1 \wedge \theta_2 \Rightarrow K = -\frac{1}{\cosh^4 u}.$$

As an example of how exterior calculus is often used, we re-prove a theorem from 162A:

Theorem 2.15. *If every point on a surface is umbilic, then the surface is (part of a) round sphere.*

Proof. We have $II = \lambda I$, where λ might be a function. It follows that, with respect to an adapted frame, we have $a = c = -\lambda$ and b = 0. Indeed $\omega_{13} = a\theta_1$ and $\omega_{23} = a\theta_2$. Taking exterior derivatives and appealing to the Codazzi equations gives us

$$0 = d\omega_{13} + \omega_{12} \wedge \omega_{23} = da \wedge \theta_1 + a d\theta_1 + a\omega_{12} \wedge \theta_2$$

= $da \wedge \theta_1 - a\omega_{12} \wedge \theta_2 + a\omega_{12} \wedge \theta_2 = da \wedge \theta_1.$

Similarly $da \wedge \theta_2 = 0$. Since θ_1, θ_2 form a basis at each point, we have da = 0 and so *a* is constant. Now define $\mathbf{c} = \mathbf{x} - \frac{1}{a}\mathbf{e}_3$ and calculate

$$\mathbf{d}\mathbf{c} = \theta_1 \mathbf{e}_1 + \theta_2 \mathbf{e}_2 - \frac{1}{a}(\omega_{13}\mathbf{e}_1 + \omega_{23}\mathbf{e}_2) = 0.$$

x thus lies on the sphere of radius a^{-1} and center **c**.

3 Isometry, Gauss' Theorem Egregium, and Riemannian Geometry

In this section we consider isometries, both local and global and their effects of curvature. We finish by discussing Riemannian Geometry: how the imposition of an abstract first fundamental form induces curvature.

3.1 Invariance under Euclidean motions

There are two types of isometry to consider when it comes to surfaces.

Definition 3.1. Two surfaces $\mathbf{x}, \hat{\mathbf{x}}$ are *globally isometric* if they differ by an isometry of \mathbb{E}^3 . That is $\hat{\mathbf{x}} = A\mathbf{x} + \mathbf{b}$, where *A* is a constant orthogonal linear transformation³ and **b** is a constant vector. Two surfaces $\mathbf{x}, \hat{\mathbf{x}}$ are *locally isometric* if their first fundamental forms are identical; $\hat{\mathbf{l}} = \mathbf{I}$.

In the first case the isometry is *direct* if det A = 1 and *indirect* if det A = -1. When the isometry is direct we often say that \mathbf{x} , $\hat{\mathbf{x}}$ are related by a *rigid motion* or *Euclidean motion*. In the second case when $\hat{\mathbf{l}} = \mathbf{I}$ it is usual to say that \mathbf{x} , $\hat{\mathbf{x}}$ are simply *isometric*.

Theorem 3.2. Let $\mathbf{x} : U \to \mathbb{E}^3$ and $\hat{\mathbf{x}} : U \to \mathbb{E}^3$ be two surfaces related by a direct isometry,

$$\hat{\mathbf{x}}(u,v) = A\mathbf{x}(u,v) + \mathbf{b}.$$

Then the fundamental forms of the two surfaces are equal, and thus so are the two measures of curvature. If the isometry is indirect, then the Gauss curvature is unchanged, while the mean curvature changes sign.

Proof. Clearly $d\hat{\mathbf{x}} = A \, d\mathbf{x}$. Since A preserves dot products we have identical first fundamental forms. Matrices of positive determinant preserve orientation and those of negative determinant reverse it, so we have

$$\hat{\mathbf{U}} = \begin{cases} A\mathbf{U} & \text{direct isometry,} \\ -A\mathbf{U} & \text{indirect isometry.} \end{cases}$$

Thus $d\hat{\mathbf{U}} = \pm A \, d\mathbf{U}$. It follows that $\hat{\mathbf{I}} = \pm \mathbf{I}$ with a minus sign iff the isometry is indirect.

The converse is also true, though a little harder to prove. Two surfaces with equal fundamental forms can, for most purposes, be considered equivalent.

Lemma 3.3. Let $\mathbf{x} : U \to \mathbb{E}^3$ be a local surface with first fundamental form I. Let θ_1, θ_2 be 1-forms on U such that $I = \theta_1^2 + \theta_2^2$. Then there exists a unique adaptive frame such that $\theta_i = \mathbf{dx} \cdot \mathbf{e}_i$.

Proof. Let v_1 , v_2 be the dual vector fields to θ_1 , θ_2 . Then $\mathbf{e}_i = \mathbf{d}\mathbf{x}(v_i)$, i = 1, 2 and $\mathbf{e}_3 = \mathbf{e}_1 \times \mathbf{e}_2$ defines the frame.

Theorem 3.4 (Bonnet). Suppose that $\mathbf{x}, \hat{\mathbf{x}} : U \to \mathbb{E}^3$ are local surfaces with identical first and second fundamental forms. Then $\mathbf{x}, \hat{\mathbf{x}}$ differ by a Euclidean motion.

 $^{^{3}3 \}times 3$ orthogonal matrix with respect to any fixed basis of \mathbb{E}^{3} .

Proof. First we show that there exist moving frames for which the connection 1-forms and forms θ_i are identical for both maps. We use these frames to define an orthogonal matrix A, use the structure equations to show that A is constant, and then calculate that $\hat{\mathbf{x}} - A\mathbf{x}$ is constant.

With respect to a choice of adaptive frame $\mathbf{E} = (\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3)$, we have $\mathbf{I} = \theta_1^2 + \theta_2^2$. By Lemma 3.3 there exists a unique moving frame $\hat{\mathbf{E}} = (\hat{\mathbf{e}}_1, \hat{\mathbf{e}}_2, \hat{\mathbf{e}}_3)$ of $\hat{\mathbf{x}}$ with $\hat{\theta}_i = \theta_i$, i = 1, 2.

Writing the identical second fundamental forms in terms of coefficients $a, b, c, \hat{a}, \hat{b}, \hat{c}$ we have

 $\hat{\mathbf{I}} = -\hat{a}\hat{\theta}_{1}^{2} - 2\hat{b}\hat{\theta}_{1}\hat{\theta}_{2} - \hat{c}\hat{\theta}_{2}^{2} = -a\theta_{1}^{2} - 2b\theta_{1}\theta_{2} - c\theta_{2}^{2} = \mathbf{I}.$

Evaluating these on the pairs (v_1, v_1) , (v_1, v_2) , (v_2, v_2) respectively yields $\hat{a} = a$, $\hat{b} = b$, $\hat{c} = c$. It follows that $\hat{\omega}_{13} = \omega_{13}$ and $\hat{\omega}_{23} = \omega_{23}$.

Now consider the first structure equations for both surfaces: since $\hat{\theta}_i = \theta_i$ we have

$$\mathrm{d}\theta_1 + \alpha \wedge \theta_2 = 0 = \mathrm{d}\theta_2 - \alpha \wedge \theta_1,$$

for $\alpha = \omega_{12}$ and $\hat{\omega}_{12}$ respectively. It follows that $\beta := \hat{\omega}_{12} - \omega_{12}$ satisfies $\beta \wedge \theta_1 = 0 = \beta \wedge \theta_2$. Hence $\beta = 0$. We thus have that the moving frames **E**, $\hat{\mathbf{E}}$ have identical connection 1-forms $\hat{\omega} = \omega$.

Define a (possibly non-constant) orthogonal matrix A by $\hat{\mathbf{E}} = A\mathbf{E} = (A\mathbf{e}_1, A\mathbf{e}_2, A\mathbf{e}_3)$. Then

$$d\hat{\mathbf{E}} = (dA)\mathbf{E} + A\,d\mathbf{E} = (dA)\mathbf{E} + A\mathbf{E}\omega = (dA)\mathbf{E} + \hat{\mathbf{E}}\hat{\omega}.$$

It follows that dA = 0 and A is constant. Finally note that

$$d\hat{\mathbf{x}} = \hat{\theta}_1 \hat{\mathbf{e}}_1 + \hat{\theta}_2 \hat{\mathbf{e}}_2 = A(\theta_1 \mathbf{e}_1 + \theta_2 \mathbf{e}_2) = A \, d\mathbf{x} = d(A\mathbf{x}).$$

Thus $\hat{\mathbf{x}} = A\mathbf{x} + \mathbf{b}$ for some constant vector \mathbf{b} and orthogonal A (note that det A = 1 since both \mathbf{E} , $\hat{\mathbf{E}}$ are oriented frames).

We could have appealed to Theorem 2.9 to see that $\hat{\mathbf{x}}$ is determined from $\hat{\omega} = \omega$ and $\hat{\theta} = \theta$ by initial conditions.

3.2 Gauss' Theorem Egregium

We have already used differential forms and moving frames for a new method to compute the Gauss curvature. We can improve on this even further, so that you need not even compute the moving frame.

- 1. Find I.
- 2. Write $I = \theta_1^2 + \theta_2^2$. Do this by inspection or by finding an adaptive frame.
- 3. Solve the first structure equations for ω_{12} .
- 4. Use the Gauss equation to find *K*.

This process is, in most cases, much easier than trying to solve for the eigenvalues of I with I. Not only that, but you needn't calculate the moving frame if you can find θ_1 , θ_2 by inspection. A Lemma below shows that it is always possible to write a first fundamental form as $\theta_1^2 + \theta_2^2$.

Examples. 1. A surface of revolution. $\mathbf{x} = \begin{pmatrix} f(u) \cos \phi \\ f(u) \sin \phi \\ u \end{pmatrix}$.

(a) I = $(1 + f'(u)^2) du^2 + f(u)^2 d\phi^2$. (b) $\theta_1 = \sqrt{1 + f'(u)^2} du$, $\theta_2 = f(u) d\phi$. (c) $d\theta_1 = 0 = -\omega_{12} \wedge \theta_2$, $d\theta_2 = f'(u) du \wedge d\phi = \omega_{12} \wedge \theta_1$. Hence,

$$\omega_{12} = -\frac{f'(u)}{\sqrt{1+f'(u)^2}} \,\mathrm{d}\phi.$$

(d) Here we have

$$d\omega_{12} = -\frac{f''(u)}{\sqrt{1+f'(u)^2}} du \wedge d\phi = -\frac{f''(u)}{f(u)(1+f'(u)^2)^3} \theta_1 \wedge \theta_2,$$

so that $K = -\frac{f''}{f(1+f'^2)^3}.$
2. Let $\mathbf{x}(x,y) = \begin{pmatrix} x^2 \\ y \\ y^2 \end{pmatrix}$, then $\mathbf{I} = 4x^2 dx^2 + (1+4y^2) dy^2$. Choosing $\theta_1 = 2x dx$ and $\theta_2 = \sqrt{1+4y^2} dy$
we have $d\theta_1 = 0 = d\theta_2$. By the structure equations $\omega_{12} = 0$ and so $K = 0$.

Since the first fundamental form encodes what angle and length mean for inhabitants of the surface, the local geometry of isometric surfaces is the same for an ant living on said surfaces. By an ant inhabiting a surface we mean that the ant has no notion of what 'outside' the surface means, or normal to the surface: all his experience comes intrinsically from the surface. The following theorem says that the Gauss curvature of a surface may, in principle, be detected by an ant.

Theorem 3.5 (Gauss' Theorem Egregium). Isometric surfaces have the same Gauss curvature.

Proof. If $\hat{I} = I$ then, by Lemma 3.3 we may choose the same θ_1, θ_2 for each surface. By the above method we get the same *K* in each case.

Definition 3.6. A surface with $K \equiv 0$ is called *flat*.

Example. A cone can be opened out and laid flat as part of a plane, hence K = 0. The same is true for a cylinder. Algebraically, the cone below is sliced along the line $\phi = 0$ and laid flat as in Figure 4.

$$\begin{pmatrix} az\cos\phi\\az\sin\phi\\z \end{pmatrix} \mapsto (1+a^2) \begin{pmatrix} z\cos\left(\frac{a}{\sqrt{1+a^2}}\phi\right)\\z\sin\left(\frac{a}{\sqrt{1+a^2}}\phi\right) \end{pmatrix}.$$

Conversely to the fact that the Gauss curvature is intrinsic (detectible by our hypothetical ant), the mean curvature is not intrinsic. For example, even if our ant decides that his home is flat, he has no local way to tell whether he is crawling on a plane, a cone, a cylinder, or some other flat surface.



Figure 4: Unwrapping a cone

A corollary of the Theorem says that no part of a sphere is isometric to part of a plane. This has applications to map-making for it implies that the perfect map is impossible: any map of part of the Earth must distort distances in some way.

The Theorem implies that I gives us *K* without needing to know **I**. Since the usual formula $K = \frac{ln-m^2}{EG-F^2}$ involves **I**, there must exist a formula for *K* involving only *E*, *F*, *G* and their derivatives. This formula is very complicated and is given below.

Suppose that x_1, x_2 are co-ordinates on **x**. The Gauss curvature can be written in terms of the *Christoffel symbols* Γ_{ij}^k . Let the first fundamental form be written $I = E dx_1^2 + 2F dx_1 dx_2 + G dx_2^2 = g_{11} dx_1^2 + (g_{12} + g_{21}) dx_1 dx_2 + g_{22} dx_2^2$, where $g_{12} = g_{21}$ and define g^{ij} to be the *ij*-th entry of the inverse matrix $\left(\frac{g_{11}}{g_{21}}\frac{g_{12}}{g_{22}}\right)^{-1}$. If we define the Christoffel symbols by,

$$\Gamma^{i}_{jk} = rac{1}{2}\sum_{m=1}^{2}g^{im}\left(rac{\partial g_{mj}}{\partial x_{k}} + rac{\partial g_{mk}}{\partial x_{j}} - rac{\partial g_{jk}}{\partial x_{m}}
ight)$$

then the Gauss curvature is given by,

$$K = -\frac{1}{g_{11}} \left(\frac{\partial \Gamma_{12}^2}{\partial x_1} - \frac{\partial \Gamma_{11}^2}{\partial x_2} + \Gamma_{12}^1 \Gamma_{11}^2 - \Gamma_{11}^1 \Gamma_{12}^2 + \Gamma_{12}^2 \Gamma_{12}^2 - \Gamma_{11}^2 \Gamma_{22}^2 \right).$$

Working with curvatures in terms of Christoffel symbols is very popular in Physics. Good luck if that's where you're going...

3.3 Riemannian geometry

The idea of Riemann geometry is to consider a domain U and specify an abstract first fundamental form, or *metric*, without it necessarily having arisen from a map $\mathbf{x} : U \to \mathbb{E}^n$. By the above procedure, it makes sense to talk about (and calculate!) the *Gauss curvature of the metric*. In this situation there is no mean curvature or second fundamental form: being no unit normal vector to U, there is simply no notion of either H or \mathbb{I} . To do this in the abstract — without any notion of a moving frame or a map \mathbf{x} — we need a Lemma:

Lemma 3.7. Any first fundamental form can be written as $I = \theta_1^2 + \theta_2^2$ for some 1-forms θ_i .

Proof. Suppose that $I = a dx^2 + 2b dx dy + c dy^2$ written in co-ordinates *x*, *y*. Then

$$I = \left(\sqrt{a} \, dx + \frac{b}{\sqrt{a}} \, dy\right)^2 + \left(\sqrt{c - \frac{b^2}{a}} \, dy\right)^2.$$

It suffices to check that $c - b^2/a > 0$. This is equivalent to $ac - b^2 > 0$ which asserts the positivity of the determinant of the matrix of I in these co-ordinates. However first fundamental forms are positive definite, so the eigenvalues of this matrix are both positive and thus so is the determinant.

There are, in fact, an infinity of possible choices of θ_i , the above is just an example.

A basic example of Riemannian geometry is hyperbolic space.

Definition 3.8. Two-dimensional hyperbolic space is the upper half-plane equipped with the metric,

$$I = \frac{(dx)^2 + (dy)^2}{y^2}.$$

Theorem 3.9. *Hyperbolic space has* K = -1*.*

Proof. Take the obvious choice $\theta_1 = y^{-1} dx$, $\theta_2 = y^{-1} dy$ and follow the recipe. Thus if $\omega_{12} = a dx + b dy$, then

$$0 = d\theta_1 + \omega_{12} \wedge \theta_2 = (y^{-2} + ay^{-1}) dx \wedge dy,$$

$$0 = d\theta_2 - \omega_{12} \wedge \theta_1 = by^{-1} dx \wedge dy,$$

so that $\omega_{12} = -y^{-1} dx$. Then $d\omega_{12} = y^{-2} dy \wedge dx = -\theta_1 \wedge \theta_2$. Thus K = -1.

Hyperbolic space could be considered as the negative analogue of the sphere.

Why cut off at y = 0? Consider a curve z(t) = (0, 1 - t) where *t* runs from 0 to 1. The length of the curve is,

$$\int_0^t \sqrt{\mathbf{I}(z',z')} \, \mathrm{d}\tau = \int_0^t \frac{1}{y} \, \mathrm{d}\tau = \int_0^t \frac{1}{1-\tau} \, \mathrm{d}\tau$$
$$= -\ln(1-t) \to \infty \quad \text{as } t \to 1.$$

The *x*-axis is thus infinitely far away from all points in hyperbolic space. We will think about hyperbolic space in the next section on geodesics.

Finding metrics from a prescribed curvature

Since the curvature is determined entirely from the metric, we can ask what metrics give a particular curvature. This is useful in practical examples because we often want to design a metric that will have particular curvature properties. For example:

Find all the metrics of the form $I = f(r)^2 dr^2 + r^2 d\theta^2$ on \mathbb{R}^2 which have constant Gauss curvature.

Here we have $\theta_1 = f(r) dr$, $\theta_2 = r d\theta$. It is not hard to see that $\omega_{12} = -\frac{1}{f(r)} d\theta$, and that

$$K = \frac{f'(r)}{rf(r)^3}.$$

We want this to be constant, hence integrate,

$$\frac{f'(r)}{f(r)^3} = Kr \Longrightarrow \frac{1}{f(r)^2} = C - Kr^2,$$

where *C* is constant. Solving gives

$$f(r) = \frac{\pm 1}{\sqrt{C - Kr^2}}, \qquad I = \frac{1}{C - Kr^2} dr^2 + r^2 d\theta.$$

An example with a singularity

Consider the metric I = $(1 + r^{-2}) dr^2 + r^2 d\theta^2$ on \mathbb{R}^2 with the origin removed. Here we choose $\theta_1 = \sqrt{1 + r^{-2}} dr$ and $\theta_2 = r d\theta$. Then, writing $\omega_{12} = a dr + b d\theta$, we have

$$0 = d heta_1 + \omega_{12} \wedge heta_2 = ar \, \mathrm{d}r \wedge \mathrm{d} heta,$$

 $0 = d heta_2 - \omega_{12} \wedge heta_1 = (1 + b\sqrt{1 + r^{-2}}) \, \mathrm{d}r \wedge \mathrm{d} heta.$

Thus $\omega_{12} = -(1+r^{-2})^{-1/2} \,\mathrm{d} \theta$, from which we have,

$$\mathrm{d}\omega_{12} = (1/2)(-2r^{-3})(1+r^{-2})^{-3/2}\,\mathrm{d}r\wedge\mathrm{d} heta = -rac{r^{-4}}{(1+r^{-2})^2} heta_1\wedge heta_2.$$

Hence $K = -(1 + r^2)^{-2}$. While it looks like the metric has a singularity at r = 0, in fact the curvature has limit -1 there. In this example we can visualize the metric as being the first fundamental form of the surface $z = f(r, \theta) = \ln r$, which certainly has a singularity at r = 0. Thus singularities in surfaces do not necessarily correspond to singularities in the Gauss curvature.

An example with a singularity on a curve

Consider the region of the plane outside the unit circle equipped with the first fundamental form

$$I = (1 - r^{-1}) dr^2 + r^2 d\theta^2.$$

For large *r* this is close to the standard Euclidean metric, but as we get close to r = 1 the first term blows up. Here $\theta_1 = \sqrt{1 - r^{-1}} dr$ and $\theta_2 = r d\theta$. Writing $\omega_{12} = a dr + b d\theta$ and applying the structure equations,

$$0 = d\theta_1 + \omega_{12} \wedge \theta_2 = ar \, dr \wedge d\theta,$$

$$0 = d\theta_2 - \omega_{12} \wedge \theta_1 = (1 + b\sqrt{1 - r^{-1}}) \, dr \wedge d\theta,$$

so that $\omega_{12} = -(1 - r^{-1})^{-1/2} \,\mathrm{d}\theta$. Thus

$$\mathrm{d}\omega_{12} = rac{1}{2}(1-r^{-1})^{-3/2}r^{-2}\,\mathrm{d}r\wedge\mathrm{d} heta = rac{1}{2r(r-1)^2} heta_1\wedge heta_2,$$

so that $K = \frac{1}{2r(r-1)^2}$. For large r, the curvature is close to zero, but as $r \to 1$ the curvature goes to ∞ . This example cannot be written as the first fundamental form of a surface $z = f(r, \theta)$.

A very different curvature is obtained if you start instead with the metric

$$\mathbf{I} = (1 - r^{-1})^{-1} \,\mathrm{d}r^2 + r^2 \,\mathrm{d}\theta^2,$$

on r > 1. This time we can see that $K = -\frac{1}{2r^3}$.

Black holes and Relativity

The previous example can be viewed as a simpler, though far from identical, version of the following example, which is very much for the Physicists.

Spacetime The following example is a metric on 4-dimensional *spacetime*, written $\mathbb{R}^{1,3}$. For us this is identical to \mathbb{R}^4 but equipped with a first fundamental form I which has 3 orthogonal *spacelike* directions $\frac{\partial}{\partial x}$, $\frac{\partial}{\partial y}$, $\frac{\partial}{\partial z}$ and one *timelike* direction $\frac{\partial}{\partial t}$. A tangent vector $v \in T_p \mathbb{R}^{3,1}$ is

$$\begin{cases} \text{spacelike} & \text{I}(v,v) > 0, \\ \text{timelike} & \text{I}(v,v) < 0, \\ \text{lightlike} & \text{I}(v,v) = 0. \end{cases}$$

I is like a standard first fundamental form in all ways except that it is not positive definite; indeed $I(\frac{\partial}{\partial t}, \frac{\partial}{\partial t}) < 0$, etc. The *standard flat metric of spacetime* is

$$I = dx^2 + dy^2 + dz^2 - c^2 dt^2$$

where *c* is the speed of light.⁴ If you think about a tangent vector

$$v = v_x \frac{\partial}{\partial x} + v_y \frac{\partial}{\partial y} + v_z \frac{\partial}{\partial z} + v_t \frac{\partial}{\partial t},$$

then the speed of a particle moving with tangent vector *v* is the infinitessimal spacial change divided by the time change, i.e.

$$ext{speed} = rac{\sqrt{v_x^2 + v_y^2 + v_z^2}}{|v_t|}.$$

Evaluating gives us

$$\mathbf{I}(v,v) = (\mathbf{speed}^2 - c^2)v_t^2,$$

from which we see that

$$v$$
 is

$$\begin{cases}
\text{spacelike} \\
\text{timelike} \\
\text{lightlike}
\end{cases} \iff \text{speed is} \begin{cases}
> c, \\
< c, \\
= c.
\end{cases}$$

One of the principal ideas of relativity is that physical objects can only travel at speeds less than that of light: i.e. they must travel only in *timelike* directions.

⁴Alternatively I = $dr^2 + r^2 d\theta^2 + r^2 \sin^2 \theta d\phi^2 - c^2 dt^2$ in spherical polar co-ordinates.

The effect of mass on spacetime The above 'standard' spacetime metric is the background metric without the presence of any mass, or as a limiting case for deep space where gravity is negligible. It is the metric of spacetime used in special relativity. In general relativity, the effect of mass/energy is considered. It affects spacetime by changing the metric. Here is a model of how this might happen:

Consider a rotationally symmetric object of mass *m* and fix a co-ordinate system (r, θ, ϕ, t) of spacetime centered at the star (spherical polar co-ordinates together with time). The Schwarzschild *metric* is defined outside the object by

$$I := \left(1 - \frac{2Gm}{c^2r}\right)^{-1} dr^2 + r^2 d\theta^2 + r^2 \sin^2 \theta d\phi^2 - c^2 \left(1 - \frac{2Gm}{c^2r}\right) dt^2,$$

where *G* is the gravitational constant and *c* the speed of light.

If we remove the mass the the metric is simply the standard flat metric. Thusfar we have only considered the Gauss curvature for surfaces or 2-dimensional domains. The Gauss curvature is the 2-dimensional avatar of the more general Riemann curvature tensor. Although we have not introduced this object, its entries⁵ may be computed in an analogous way to how we calculate the Gauss curvature in 2-dimensions: i.e. from the 1-forms θ_i such that $I = \theta_1^2 + \theta_2^2 + \theta_3^2 - \theta_4^2$ we can use the structure equations to find the 6 connection 1-forms ω_{ij} , $1 \le i < j \le 4$, the curvature tensor being the components of the matrix-valued 2-form⁶ $\Omega := d\omega + \omega \wedge \omega$. This is the essence of the notion that mass/energy curves spacetime.

While we're not going to calculate the full curvature of the Schwarzschild metric, we can calculate certain sectional curvature: the Gauss curvature of a 2-dimensional subspace of $T_p \mathbb{R}^{3,1}$. In particular we can calculate the sectional curvature of the space $\left\langle \frac{\partial}{\partial r}, \frac{\partial}{\partial t} \right\rangle \leq T_p \mathbb{R}^{3,1}$, which is equivalent to calculating the Gauss curvature of the metric

$$\left(1-\frac{2Gm}{c^2r}\right)^{-1}\,\mathrm{d}r^2-c^2\left(1-\frac{2Gm}{c^2r}\right)\,\mathrm{d}t^2,$$

on \mathbb{R}^2 . This can be seen to be $K = \frac{2Gm}{c^2r^3}$. A homework question walks you through some of this. By analogy with the previous example, the sectional curvature of the (r, θ) -plane is $K = -\frac{Gm}{c^2 r^3}$.

The Schwarzschild metric is only valid outside a mass, thus at the surface of the Earth $r = 6 \times 10^6$ meters, the metric is approximately

$$\mathbf{I} = (1+10^{-9}) \,\mathrm{d}r^2 + r^2 \,\mathrm{d}\theta^2 + r^2 \sin^2\theta \,\mathrm{d}\phi^2 - (1-10^{-9})c^2 \,\mathrm{d}t^2,$$

differing from the standard flat metric by only a tiny tiny amount. Even at the surface of the sun the difference from the standard metric is only $4.25 \times 10^{-6} (dr^2 - c^2 dt^2)$, and the sectional curvature calculated above is only $K = 8.8 \times 10^{-24}$. This very slight difference from flat (K = 0) space is just detectable. In a famous 1919 experiment of was Eddington and Dyson, the light from stars very close to the sun was observed during a total eclipse and the bending of light was found to be within observational error of Einstein's prediction. This, combined with general relativity's prediction of the correct precession of Mercury's orbit, helped convince people that Einstein had a good theory. The process of light bending around massive objects is known as *gravitational lensing*.

The Schwarzschild metric was one of the first explicit non-trivial solutions to Einstein's equations of general relativity to be found. Analyzing the curvature of this metric allows Physicists to model

⁵In *n* dimensions it has $n^2(n^2 - 1)/12$ independent components. ⁶In 2-dimensions $\omega = \begin{pmatrix} 0 & \omega_{12} \\ -\omega_{12} & 0 \end{pmatrix}$, so that $\Omega = \begin{pmatrix} 0 & d\omega_{12} \\ -d\omega_{12} & 0 \end{pmatrix} = K \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \theta_1 \wedge \theta_2$.

what an object might experience as it fell towards a huge mass: does it stretch? does its volume change, etc.

At the time the solution was found, the physical significance of the seeming singularity at the *Schwarzschild radius* $r = R := \frac{2Gm}{c^2}$ was not understood. First of all, for the metric to be valid at such a radius requires all the mass of the object to be concentrated in the region r < R. Since $\frac{2G}{c^2} \approx 1.485 \times 10^{-27}$ meters per kilogram, the Schwarzschild radius is minuscule for non-stellar masses: for a human it's about 10^{-25} meters, roughly *one ten-billionth the size of a proton!*, while a Schwarzschild radius of 1 meter would have to contain 6.733×10^{26} kg ≈ 113 times the mass of the Earth. An object who's mass is contained entirely within its Schwarzschild radius is a *black hole*. Here are some of the predictions of the Schwarzschild black hole model.

- The region of spacetime given by r = R is termed the *event horizon*. As r gets very close to R (from above) all terms in the metric become dwarfed by the dr^2 term and the metric looks like it becomes singular. The curvature does not however (we've already seen that the sectional curvature of the (r, t)-plane is finite at the event horizon).
- Suppose you are an observer sitting a long way from the black hole: *t* is your measure of time. Now throw a ticking clock towards the black hole. An infinitessimal change dt in *t* corresponds (via the metric) to an infinitessimal change (1 - R/r) dt in the time on the clock. As $r \rightarrow R$, the time change on the sacrificial clock approaches zero: this is *time dilation*. It takes an infinite amount of time as measured by the observer a long way away for the clock to fall in, but according to the clock, it takes only a finite amount of time.
- A 2-dimensional visualization of the shape of space around a black hole is available if you fix time and restrict to a constant angle $\theta = \pi/2$ from the north pole. The metric is then $\left(1 \frac{2Gm}{c^2r}\right)^{-1} dr^2 + r^2 d\phi^2$. This is the first fundamental form of the surface $z = 2R\sqrt{1 R/r}$ in polar co-ordinates (see Figure 5).



Figure 5: Representation of a black hole at constant time

• At the Schwarzschild radius r = R, radial and timelike directions switch sign in the metric. This seems to make no sense until you remember that the metric describes spacetime from the point of view of an observer at rest. The switching of the signs says that it is impossible to be at rest inside the event horizon: there can be no stationary observers. No matter what is powering your spacecraft, it is impossible to resist the pull of the black hole, even light cannot stay still, let alone escape from within. Indeed, if you pretend for a moment that classical mechanics applies, the Schwarzschild radius is exactly the radius for which the escape velocity of a given mass is the speed of light.

• The center of the black hole at r = 0 is a genuine singularity, both of the metric, and of the curvature. At present there is no good physical understanding of what this might mean. Since nothing can be observed with the event horizon of a black hole, no-one knows whether there really is a singularity of spacetime at the center. There have recently been attempts (e.g. Chapline's Dark Energy star model) to meld quantum physics with relativity to explain what might happen inside the event horizon.

4 Geodesics, Parallel transport and covariant derivatives

The concept of a geodesic is very important for applications. It replaces the notion of straight line for surfaces and more complicated objects. For example, it is well-known, and we prove below, that a straight line is the shortest path between two points in Euclidean space. How would we go about finding the path of shortest length between two points on a surface?

4.1 Geodesics in Euclidean space

We want to find the shortest curve in \mathbb{E}^3 joining two given points. Let $\mathbf{x}(t)$ be unit speed such that $\mathbf{x}(a) = A$, $\mathbf{x}(b) = B$. Let ϵ be a small number, and $\mathbf{y} : [a, b] \to \mathbb{E}^3$ be a curve such that $\mathbf{y}(a) = \mathbf{y}(b) = 0$ and $\mathbf{x}' \cdot \mathbf{y} = 0$. Define

$$\mathbf{r}_{\boldsymbol{\epsilon}}(t) = \mathbf{x}(t) + \boldsymbol{\epsilon} \mathbf{y}(t).$$

Since **y** is orthogonal to the tangent direction of **x**, the point $\mathbf{r}_{\epsilon}(t)$ is obtained from $\mathbf{x}(t)$ by a normal movement for each *t*. Clearly $\mathbf{r}(a) = A$ and $\mathbf{r}(b) = B$ so that, for any choice of **y**, we have a family of curves \mathbf{r}_{ϵ} connecting *A* and *B* (see Figure 6).



Figure 6: Family of curves \mathbf{r}_{ϵ}

Definition 4.1. The curve $\mathbf{x}(t)$ has *stationary length* if the lengths $\ell(\epsilon)$ of the curves $\mathbf{r}_{\epsilon}(t)$ satisfy,

$$\left. \frac{\partial \ell}{\partial \epsilon} \right|_{\epsilon=0} = 0$$

for all choices of $\mathbf{y}(t)$ as defined above.

Theorem 4.2. The curve $\mathbf{x}(t)$ above has stationary length iff it is a straight line.

Proof. Throughout we ignore terms of order ϵ^2 or higher. Since we are eventually evaluating at $\epsilon = 0$, and, at most, are performing a single derivative with respect to ϵ , in the final analysis all such terms contribute zero to the final answer. Wherever terms of order ϵ^2 or higher have been deleted we use the \simeq sign.

Since the derivative of the curve \mathbf{r}_{ϵ} is given by $\mathbf{r}'_{\epsilon} = \mathbf{x}' + \epsilon \mathbf{y}'$, its speed is

$$v_{\epsilon}(t) = \sqrt{\mathbf{r}_{\epsilon}' \cdot \mathbf{r}_{\epsilon}'} = \sqrt{|\mathbf{x}'|^2 + 2\epsilon \mathbf{x}' \cdot \mathbf{y}' + \epsilon^2 |\mathbf{y}'|^2} \simeq (1 + 2\epsilon \mathbf{x}' \cdot \mathbf{y}')^{1/2} \simeq 1 + \epsilon \mathbf{x}' \cdot \mathbf{y}',$$

where the final \simeq uses the Taylor approximation. Then,

$$\begin{aligned} \frac{\partial \ell}{\partial \epsilon} \Big|_{\epsilon=0} &= \left. \frac{\mathrm{d}}{\mathrm{d} \epsilon} \Big|_{\epsilon=0} \int_{a}^{b} v_{\epsilon}(t) \, \mathrm{d} t = \int_{a}^{b} \mathbf{x}' \cdot \mathbf{y}' \, \mathrm{d} t \\ &= \int_{a}^{b} \frac{\mathrm{d}}{\mathrm{d} t} (\mathbf{x}' \cdot \mathbf{y}) \, \mathrm{d} t - \int_{a}^{b} \mathbf{x}'' \cdot \mathbf{y} \, \mathrm{d} t \\ &= -\int_{a}^{b} \mathbf{x}'' \cdot \mathbf{y} \, \mathrm{d} t, \end{aligned}$$

since $\mathbf{x}' \cdot \mathbf{y} = 0$. This expression vanishes for all \mathbf{y} iff \mathbf{x}'' is parallel to \mathbf{x}' . However \mathbf{x} has constant speed, hence $\mathbf{x}' \cdot \mathbf{x}'' = 0$. Thus $\mathbf{x}'' = 0$ and so \mathbf{x} is a straight line.

The above is a calculus of variations argument. Note that the proposition only shows a straight line to be of stationary length, not of minimum length. You can think about this as an infinite dimensional calculus problem. However, of all the curves of unit speed between the points A, B, x is the only one that has stationary length. The function ℓ : {smooth curves $A \rightsquigarrow B$ } $\rightarrow \mathbb{R}$ is smooth on an open set, so if there exists a minimum, then said curve must have stationary length. In fact, the straight line is the minimum length curve joining two point. We say that straight lines are *geodesics* in Euclidean space.

4.2 Geodesics in surfaces

We want to find a unit speed curve $\mathbf{x}(z(t))$ which is the shortest curve between two points in a surface **x**. Again we consider curves of the form

$$\mathbf{r}_{\epsilon}(t) = \mathbf{x}(z(t)) + \epsilon \mathbf{y}(t),$$

where $\mathbf{y}(a) = \mathbf{y}(b) = 0$ and $\mathbf{y} \cdot \mathbf{U} = 0$. \mathbf{r}_{ϵ} is thus a curve in \mathbb{E}^3 which is perturbed a small amount from $\mathbf{x}(z(t))$ in the tangent direction to \mathbf{x} at all points. Note that \mathbf{r}_{ϵ} is no longer a curve in the surface when $\epsilon \neq 0$.

Definition 4.3. $\mathbf{x}(z(t))$ has *stationary length* if, for all such \mathbf{y} ,

$$\left. \frac{\mathrm{d}\ell}{\mathrm{d}\epsilon} \right|_{\epsilon=0} = 0.$$

 $\mathbf{x}(z(t))$ is a *geodesic* if $\mathbf{x}'' = \frac{d^2}{dt^2}\mathbf{x}(z(t))$ is normal to the surface everywhere.

Theorem 4.4. A unit speed curve lying in a surface has stationary length iff it is a geodesic.

Proof. The argument begins in the same way as in Theorem 4.2 (recall throughout that $\mathbf{x}' = \frac{d}{dt}\mathbf{x}(z(t))$, etc.):

$$\begin{aligned} \frac{\partial \ell}{\partial \epsilon} \Big|_{\epsilon=0} &= \left. \frac{\mathrm{d}}{\mathrm{d} \epsilon} \right|_{\epsilon=0} \int_{a}^{b} v_{\epsilon}(t) \, \mathrm{d} t = \int_{a}^{b} \mathbf{x}' \cdot \mathbf{y}' \, \mathrm{d} t \\ &= \int_{a}^{b} \frac{\mathrm{d}}{\mathrm{d} t} (\mathbf{x}' \cdot \mathbf{y}) \, \mathrm{d} t - \int_{a}^{b} \mathbf{x}'' \cdot \mathbf{y} \, \mathrm{d} t \\ &= [\mathbf{x}' \cdot \mathbf{y}]_{a}^{b} - \int_{a}^{b} \mathbf{x}'' \cdot \mathbf{y} \, \mathrm{d} t = -\int_{a}^{b} \mathbf{x}'' \cdot \mathbf{y} \, \mathrm{d} t, \end{aligned}$$

where this time the first term vanishes due to $\mathbf{y}(a) = \mathbf{y}(b) = 0$. $\mathbf{x}(z(t))$ has stationary length iff this last term vanishes for all \mathbf{y} . But $\mathbf{y} \cdot \mathbf{U} = 0$, hence \mathbf{x}'' cannot have any component tangent to the surface. \mathbf{x}'' is thus normal.

A little more is true: the definition of geodesic forces it to have constant speed. Indeed if $\mathbf{x}(z(t))$ is a geodesic, then $\frac{d}{dt} |\mathbf{x}'|^2 = 2\mathbf{x}' \cdot \mathbf{x}'' = 0$, since \mathbf{x}'' is normal to the surface and \mathbf{x}' is tangent. It is just common practice to parameterize geodesics by unit speed.

Example. On the sphere, geodesics are great circles (radius equal to that of the sphere). These minimize length if we are less than half way round. The 'long' great circle between two points is a saddle of the length function. Here's how to calculate this naïvely: Suppose that $\mathbf{x}(z(t))$ is a geodesic. Since \mathbf{x} is a sphere, we may orient things so that that $\mathbf{x} = r\mathbf{U}$, where r is the radius of the sphere. The geodesic condition then reads

$$\mathbf{x}^{\prime\prime} = f\mathbf{x},\tag{(*)}$$

where *f* is some scalar function.⁷ Take dot products of (*) with **x**' to get $0 = \mathbf{x}'' \cdot \mathbf{x}' = \frac{1}{2} \frac{d}{dt} |\mathbf{x}'|^2$. Hence $\mathbf{x}(z(t))$ has constant speed *v*. Now take dot products of (*) with **x** to obtain

$$r^{2}f = \mathbf{x}'' \cdot \mathbf{x} = \frac{\mathrm{d}}{\mathrm{d}t}(\mathbf{x} \cdot \mathbf{x}') - |\mathbf{x}'|^{2} = -v^{2},$$

so that *f* is constant and indeed

$$\mathbf{x}'' = -\frac{v^2}{r^2}\mathbf{x}.$$

Finally, the vector $\mathbf{k} := \mathbf{x} \times \mathbf{x}'$ is constant ($\mathbf{k}' = \mathbf{x} \times \mathbf{x}'' = 0$), thus \mathbf{x} lies in the intersection of the constant plane \mathbf{k}^{\perp} and the sphere: the definition of a great circle.

 $^{^{7}}$ *f* **x** is *multiplication* of the value of *f* by **x**, evaluation which makes no sense.

Contrary to what happens in \mathbb{E}^3 , geodesics in surfaces do not necessarily minimize lengths between points, although they do locally. If a path is the shortest between two points, then it is necessarily a geodesic, but the converse is not true. For example, consider two points on a sphere. We've already seen that geodesics are great circles, but there are two ways to joint two points on a sphere by an arc of a great circle: the shorter of the two will be the shortest path. The longer geodesic 'going the long way round' is simply a stationary point of the length function and neither a maximum or a minimum.

Geodesics and cartography. Since the shortest path between two points is a geodesic, it is along these that planes tend to fly (if you discount wind effects, less distance = less fuel). If you think of the standard projections of the Earth (say the commonly used Mercator projection — see 162A) and you plot what happens to great circles, you see that they are not mapped to straight lines on the projection (great circles being mapped to straight lines only happens for the rather strange-looking gnomonic projection). This is why it looks like you're taking a longer curving path when you take a long flight. You are in fact taking the shortest path there is. Figure 7 shows a geodesic flight-path.⁸



Figure 7: The great circle from Irvine CA to Irvine Scotland

We now consider geodesics in terms of the structure equations.

Theorem 4.5. Let $\mathbf{x} : U \to \mathbb{E}^3$ be a surface for which a moving frame has been chosen.⁹ A unit speed curve $\mathbf{x}(z(t))$ lying in the surface is a geodesic iff it satisfies the following equations:

$$\begin{array}{l} \frac{d}{dt}(\theta_{1}(z')) + \omega_{12}(z')\theta_{2}(z') = 0\\ \frac{d}{dt}(\theta_{2}(z')) - \omega_{12}(z')\theta_{1}(z') = 0\\ (\theta_{1}(z'))^{2} + (\theta_{2}(z'))^{2} = 1 \end{array} \quad \text{Geode}$$

Geodesic equations, Energy equation.

⁸Courtesy of the Great Circle Mapper http://gc.kls2.com/

⁹Thus fixing the 1-forms $\theta_1, \ldots, \omega_{23}$.

Proof. The energy equation is just the unit speed condition I(z', z') = 1. Now a curve having \mathbf{x}'' normal to the surface is equivalent to $\mathbf{e}_1 \cdot \mathbf{x}'' = 0 = \mathbf{e}_2 \cdot \mathbf{x}''$. Expanding the first as $(\mathbf{e}_i \cdot \mathbf{x}')' - \mathbf{e}'_i \cdot \mathbf{x}'$ yields

$$0 = \mathbf{e}_1 \cdot \mathbf{x}'' = (\mathbf{e}_1 \cdot \mathbf{x}')' - \mathbf{e}'_1 \cdot \mathbf{x}' = (\mathbf{e}_1 \cdot \mathbf{d}\mathbf{x}(z'))' - \mathbf{d}\mathbf{e}_1(z') \cdot \mathbf{d}\mathbf{x}(z')$$
$$= \frac{\mathbf{d}}{\mathbf{d}t}(\theta_1(z')) + \omega_{12}(z')\theta_2(z').$$

 $\mathbf{e}_2 \cdot \mathbf{x}'' = 0$ similarly expands to the second geodesic equation.

Any parameterized curve $\mathbf{x}(z(t))$ satisfying just the geodesic equations (i.e. \mathbf{x}'' normal to the surface) automatically has constant speed: observe that the geodesic equations force

$$\frac{\mathrm{d}}{\mathrm{d}t}((\theta_1(z'))^2 + (\theta_2(z'))^2) = 0.$$

The energy equation simply normalizes this speed to be 1. In practice, 'geodesic' often simply refers to a non-parameterized curve as a subset of the surface which, if parameterized by arc-length, satisfies our definition.

Proposition 4.6. *Given a point* **p** *on a surface and a unit tangent vector* **v** *to the surface at that point, there exists a unique geodesic through* **p** *in the direction* **v***: equivalently* $\mathbf{x}(z(0)) = \mathbf{p}$ *and* $\mathbf{x}'(z(0)) = \mathbf{v}$.

Proof. This is a consequence of the usual theorem of existence of solutions to ODE's: the geodesic equations are a pair of second-order ODE's.

Given two points on a 'nice' surface, there exists a geodesic joining those points. Nice in this situation essentially means that you wouldn't want the geodesic to leave or touch the edge of the surface. Certainly any two points on a *handlebody* (a complete surface—closed with no edge and no self-intersections) may be joined by a geodesic. However, you'll need good luck to succeed in computing many of these explicitly!

4.3 The surface of revolution

As an example, we consider a standard surface of revolution,

$$\mathbf{x}(u,\phi) = \begin{pmatrix} f(u)\cos\phi\\f(u)\sin\phi\\u \end{pmatrix}.$$

We already know that $\theta_1 = \sqrt{1 + f_u^2} \, du$, $\theta_2 = f \, d\phi$ and $\omega_{12} = -\frac{f_u}{\sqrt{1 + f_u^2}} \, d\phi$. For all curves $\mathbf{x}(z(t))$ we have,

$$z' = u'\frac{\partial}{\partial u} + \phi'\frac{\partial}{\partial \phi}.$$

Thus,

$$\theta_1(z') = \sqrt{1 + f_u^2} u', \qquad \theta_2(z') = f \phi', \qquad \omega_{12}(z') = \frac{-f_u}{\sqrt{1 + f_u^2}} \phi'.$$

The geodesic equations and the energy equation thus become,

$$\frac{\mathrm{d}}{\mathrm{d}t}(\sqrt{1+f_{u}^{2}}u') - \frac{f_{u}\phi'^{2}}{\sqrt{1+f_{u}^{2}}} = 0,$$

$$\frac{\mathrm{d}}{\mathrm{d}t}(f\phi') + f_{u}u'\phi' = 0,$$

$$(1+f_{u}^{2})u'^{2} + f^{2}\phi'^{2} = 1.$$

Observe that the lines ϕ = constant (lines of longditude) are geodesics: the second equation is trivially satisfied while the 1st and the energy equation become the same first order ODE for *u* which, by the usual theorem from ODE's, has a solution.

The lines of latitude u = constant are geodesics if only if $ff_u \phi'^2 = 0$ which is iff $f_u = 0$ (the remaining equations allow one to solve for ϕ). Thus lines of latitude are only geodesics at critical values of the radius.

Notice that, since $f' = f_u u'$, the second geodesic equation is equivalent to $A = f^2 \phi' = \text{constant}$. Substituting back into the energy equation we have,

$$(1+f_u^2){u'}^2 + \frac{A^2}{f^2} = 1.$$

Hence $|A| \leq f$. This shows that a geodesic starting at a given point – which comes equipped with a value for A — is confined to regions of the surface where $f \geq A$. This means that geodesics can 'bounce off narrow necks' of surfaces of revolution.

It's not advisable to try to solve these equations explicitly for many surfaces. Even for the cone it is extremely messy The cone is one that we can manage. If you let f(u) = au, where *a* is constant then it can be seen that

$$u(t) = \pm \sqrt{\frac{A^2}{a^2} + \frac{(t+B)^2}{1+a^2}}, \quad \phi(t) = \frac{\sqrt{1+a^2}}{a} \tan^{-1} \left(a \frac{t+B}{A\sqrt{1+a^2}} \right) + C,$$

where *A*, *B*, *C* are constants. Normalizing B = 0 = C and fixing u > 0, we see that f(u(0)) = |A| and *u* increases as *t* moves in either direction, ϕ keeps increasing as *t* does, but never quite makes it to angle $\frac{\sqrt{1+a^2\pi}}{2a}$. If you take a shallow cone (a = 1 is the cone $z^2 = x^2 + y^2$), then the geodesics have no self intersections. However, once $a < \frac{1}{\sqrt{3}}$ the geodesics will begin to meet round the far side of the cone. Figure 8 illustrates a geodesic with $a = \frac{1}{3}$. Note that this is clearly not the shortest path between the two points on the geodesic at the bottom of the picture: indeed as *a* decreases and the cone becomes more sharp, the number of distinct geodesic paths between two points increases.

4.4 Riemannian geometry

Since the geodesic equations depend only on quantities derived from the first fundamental form I, we can apply them to abstract first fundamental forms as we did in the section on Riemannian geometry. Notice first that isometric surfaces **x**, **y** have the same geodesics in the sense that if $\mathbf{x}(z(t))$ is a geodesic, then so is $\mathbf{y}(z(t))$.

Geodesics are extremely important in Physics. For example light is always assumed to travel along the shortest path between two points. If your 'geometry' is that of refraction between two mediums for which the speed of light is different, you will obtain Snell's law. If your geometry is



Figure 8: Geodesic on a cone

some curved spacetime such as that of the Schwarzschild metric, then light is seen to curve. Indeed the concept of motion along a geodesic is the relativistic replacement of Newton's first law: instead of a body moving at constant speed along a straight line if unaffected by external forces, a body will move along a geodesic in spacetime. This can be difficult to visualize when you have curved spacetime, because the notion of speed is different in spacetime and 'normal' space: indeed a curve with zero acceleration in curved spacetime will, when translated into 3-dimensional space, not have constant speed. In flat spacetime however, geodesics are straight lines, and these translate to straight lines in our 3-dimensional reference frame. Indeed if a curve has zero acceleration in flat spacetime, and constant speed $-k^2$, then its speed in Euclidean space will be $\sqrt{c^2 - k^2/v_t^2}$, where v_t is the (necessarily constant) *t*-component of the tangent vector to the curve in spacetime.

First we consider the upper half plane model of hyperbolic space. The geodesic and energy equations become,

$$\left(\frac{x'}{y}\right)' - \frac{x'y'}{y^2} = 0,$$
$$\left(\frac{y'}{y}\right)' + \frac{x'^2}{y^2} = 0,$$
$$\frac{x'^2}{y^2} + \frac{y'^2}{y^2} = 1.$$

Notice that the first geodesic equation is equivalent to $A = \frac{x'}{y^2} = \text{constant}$. Taking A = 0 gives us $x' = 0 \Rightarrow \frac{y'}{y} = \pm 1 \Rightarrow y = Ce^{\pm t}$. Hence vertical straight lines are geodesics. Otherwise, first rearrange the energy equation,

$$1 + \frac{y^2}{x^2} = \frac{y^2}{x^2} = \frac{1}{A^2 y^2}.$$

Now observe that $\frac{dy}{dx} = \frac{y'}{x'}$, so that

$$1 + \left(\frac{\mathrm{d}y}{\mathrm{d}x}\right)^2 = \frac{1}{A^2 y^2} \Longrightarrow \left(\frac{\mathrm{d}y}{\mathrm{d}x}\right)^2 = \frac{1 - A^2 y^2}{A^2 y^2}.$$

Solving this we have

$$\int \frac{\pm Ay \, \mathrm{d}y}{\sqrt{1 - A^2 y^2}} = \int \, \mathrm{d}x \Longrightarrow \mp A^{-1} \sqrt{1 - A^2 y^2} = x - C,$$

where *C* is a constant. Squaring and rearranging yields

$$(x - C)^2 + y^2 = A^{-2},$$

thus circles centered on the *x*-axis are the other geodesics (see Figure 9).



Figure 9: Geodesics leaving a point in the hyperbolic upper half plane

Geometry in this picture is a little strange. For example if we wanted to talk about a triangle, the only sensible definition is that its edges be geodesics (since geodesic is what we mean by 'straight' in this geometry). Figure 10 gives an example.



Figure 10: A geodesic triangle

Geodesics are infinitely long in hyperbolic space. Given two points, there exists a unique geodesic joining them. The situation is like that of \mathbb{E}^2 except that Euclid's "parallel postulate" does hold. (Given a line (geodesic) and a point not on the line, there exist many geodesics through the point not meeting the original line.) Euclid's postulate was that there is only one such 'parallel' line. Hyperbolic space shows that the postulate cannot be proved from the other axioms of Euclidean geometry.

If you know a little complex analysis, specifically Möbius transforms, then it can be shown that the geodesics for Poincaré's disc model of hyperbolic space (given by the metric

$$\mathbf{I} = \frac{4}{(1-r^2)^2} (r^2 \, \mathrm{d}r^2 + r^2 \, \mathrm{d}\theta^2)$$

on the unit disc) are arcs of circles which intersect the edge of the disc at right angles. The *Circle Limit* artworks by M.C. Escher illustrate this geometry very nicely (Figure 11 is Circle Limit IV). If you look at a string of angels and demons connected head to toe you'll see that they're lined up along one of



Figure 11: M.C. Escher's Circle Limit IV

these geodesics. It's hard to measure exactly, but it seems reasonable to suppose that with respect to the hyperbolic metric, the individual figures all have the same length and area.

4.5 Covariant derivatives and geodesics

Geodesics are intimately related to the concept of covariant differentiation and parallel transport of vector fields. At the heart of this is the idea that, in a vector space, we have an inherent notion of transporting a vector along a curve (just slide it so that it points in the same direction). When thinking about transporting vectors along a curve in a surface we are left with a problem: just sliding a vector that is *tangent* to the surface along the curve so that it points 'in the same direction' will almost certainly result in the transformed vector not being tangent to the surface any longer. What is

the correct thing to do in this situation?

Definition 4.7. Suppose that $\gamma(t) = \mathbf{x}(z(t))$ is a parameterized curve in a surface $\mathbf{x} : U \to \mathbb{E}^3$. A *vector field* \mathbf{v} *along* γ is a smooth assignment $\mathbf{v}(t) = \mathbf{dx}(X(z(t)))$ of a vector tangent to the surface at $\gamma(t)$ (here X(z(t))) is a vector field X on U restricted to the parameterizing curve z(t)).

The assignment is smooth in the sense that the function $t \mapsto \mathbf{v}(t)$ is infinitely differentiable.

Definition 4.8. Let **v** be a vector field along a curve γ in a surface. The *covariant derivative* of **v** is the vector field $D_{\frac{d}{dt}} \mathbf{v} := \pi^T \mathbf{d}_{x'} \mathbf{v} = \pi^T \mathbf{d}_{z'} \mathbf{d} \mathbf{x}(X(z(t))) = \pi^T \frac{d}{dt} \mathbf{v}$, where π^T is orthogonal projection onto the tangent space to the surface at each point.

We differentiate the vector field \mathbf{v} with respect to the parameter t, and then project back onto the tangent space.

Example. If $\gamma(t)$ is a curve in a surface, then $\gamma'(t)$ is a vector field along γ . The covariant derivative of γ' is then $\pi^T \gamma''(t)$, often written $D_{\gamma'} \gamma'$.

Definition 4.9. A vector field **v** along a curve γ is *parallel* if its covariant derivative is zero everywhere.

By appealing to the above example we see that we have proved the following:

Theorem 4.10. A curve γ in a surface is a geodesic iff the vector field γ' is parallel along γ .

Let $\mathbf{v}(t) = \mathbf{d}\mathbf{x}(X(z(t)))$ be a vector field along a curve $\gamma(t) = \mathbf{x}(z(t))$. If we choose an adaptive frame \mathbf{e}_1 , \mathbf{e}_2 for the surface, then there exist functions a(t), b(t) such that

$$\mathbf{v}(t) = a(t)\mathbf{e}_1(z(t)) + b(t)\mathbf{e}_2(z(t)),$$

in which case the covariant derivative may be written

$$D_{\frac{d}{dt}}\mathbf{v} = \pi^T \frac{d}{dt}\mathbf{v}(t) = (a'(t) + b(t)\omega_{12}(z'(t)))\mathbf{e}_1(z(t)) + (b'(t) + a(t)\omega_{21}(z'(t)))\mathbf{e}_2(z(t)).$$

It follows that **v** is parallel along γ iff the coefficients *a*, *b* satisfy the differential equation

$$\begin{pmatrix} a'\\b' \end{pmatrix} = \begin{pmatrix} 0 & -\omega_{12}(z')\\\omega_{12}(z') & 0 \end{pmatrix} \begin{pmatrix} a\\b \end{pmatrix}.$$
 +

Note that $\omega_{12}(z'(t))$ is simply a function of t, and so we may define a function $g(t) = \int_{t_0}^t \omega_{12}(z'(t)) dt$. It is then easily checked that the solution to the above differential equation with initial condition prescribes at $t = t_0$ is

$$\begin{pmatrix} a(t) \\ b(t) \end{pmatrix} = \begin{pmatrix} \cos g(t) & -\sin g(t) \\ \sin g(t) & \cos g(t) \end{pmatrix} \begin{pmatrix} a(t_0) \\ b(t_0) \end{pmatrix}.$$
 \ddagger

We have thus proved the following theorem.

Theorem 4.11. Given a smooth curve γ in a surface \mathbf{x} , and an initial vector \mathbf{v}_0 tangent to the surface at $\gamma(t_0)$, there exists a unique parallel vector field \mathbf{v} along γ such that $\mathbf{v}(t_0) = \mathbf{v}_0$.

Definition 4.12. The vector field **v** is above termed the *parallel transport* of \mathbf{v}_0 along γ .

The concept of parallel transport is exactly what sliding a vector along a curve means inside a surface. Indeed it is clear that $|\mathbf{v}(t)|^2 = a^2 + b^2 = a_0^2 + b_0^2$ is constant so that the length of a parallel transported vector is constant. Moreover, it can be seen (see homework) that the parallel transport of a vector along a geodesic has constant angle with the tangent field to that geodesic.

- Examples. 1. If we jump dimensions for a moment and think about the covariant derivative in \mathbb{E}^3 , then D is simply 'differentiate then project onto the tangent space at each point'. But this tangent space is exactly \mathbb{E}^3 itself, so the projection is the identity map, and D is just normal differentiation. In this case, let γ be a curve in \mathbb{E}^3 , and \mathbf{v}_0 a vector at $\gamma(0)$. We may take our moving frame to be the standard basis \mathbf{e}_1 , \mathbf{e}_2 , \mathbf{e}_3 , the connection 1-forms ω_{ii} ($1 \le i, j \le 3$) of which are all zero. Comparing with the differential equation (†), we see that the coefficients of the parallel transport v of v₀ with respect to the standard basis are all constant. Thus $v = v_0$ and parallel transport really just moves the vector \mathbf{v}_0 keeping it parallel and of the same length.
 - 2. Now let $\mathbf{x} : U \to \mathbb{E}^3$ be the unit sphere $\mathbf{x}(\theta, \phi) = \begin{pmatrix} \sin \theta \cos \phi \\ \sin \theta \sin \phi \\ \cos \theta \end{pmatrix}$ and consider the parallel trans- $\sin (t + \pi/2) \cos t \end{pmatrix}$

port of the vector
$$\mathbf{v}_0 = \begin{pmatrix} 1/4 \\ 1/2 \end{pmatrix}$$
 at the point $\gamma(0)$ along the curve $\gamma(t) = \begin{pmatrix} \sin(t + \pi/2) \sin t \\ \cos(t + \pi/2) \end{pmatrix} = (\cos t \cos t)$

 $\begin{pmatrix} \cos t \cos t \\ \cos t \sin t \\ -\sin t \end{pmatrix}$. In terms of the moving frame on page 12 we have $d\mathbf{x} = \mathbf{e}_1 d\theta + \sin \theta \mathbf{e}_2 d\phi$ so that $\theta_1 = d\theta, \theta_2 = \sin \theta d\phi$ and $\omega_{12} = -\cos \theta d\phi$. Now $z(t) = (\theta(t), \phi(t)) = (t + \pi/2, t)$, so that $z'(t) = \frac{\partial}{\partial \theta} + \frac{\partial}{\partial \phi}$. Thus $\omega_{12}(z') = -\cos(t + \pi/2) = \sin t$. Hence $g(t) = \int_0^t \omega_{12}(z') dt = 1 - \cos t$. Noting that $a(0) = -\frac{1}{2}$ and b(0) = 1/4, we apply equation (‡) to see that the coefficients of the parallel transport of m along a(t) and parallel transport of \mathbf{v}_0 along $\gamma(t)$ are

$$a(t) = -\frac{1}{2}\cos(1-\cos t) + \frac{1}{4}\sin(1-\cos t), \qquad b(t) = -\frac{1}{2}\sin(1-\cos t) - \frac{1}{4}\cos(1-\cos t),$$

so that the parallel transport of \mathbf{v}_0 along γ is

$$\mathbf{v}(t) = a(t) \begin{pmatrix} -\sin t \cos t \\ -\sin t \sin t \\ -\cos t \end{pmatrix} + b(t) \begin{pmatrix} -\sin t \\ \cos t \\ 0 \end{pmatrix}.$$

Notice that the length of \mathbf{v} is constant. Figure 12 shows the curve and the transported vector field for $-\pi/2 < t < \pi/2$.

It is not hard to show (see homework) that by parallel transporting around closed curves on general curved surfaces, you can transform a tangent vector to any other. Figure 13 shows the parallel transport of the red tangent vector $(-\frac{1}{2},0,0)^T$ at the north pole down to the equator, round the equator by an angle ϕ_0 , and back to the north pole. $(-\frac{1}{2}, 0, 0)^T$ has become the blue vector $(-\frac{1}{2}\cos\phi_0,-\frac{1}{2}\sin\phi_0,0)^T.$



Figure 12: A parallel transported vector field



Figure 13: Transporting a vector field around a closed curve

4.6 Covariant derivatives more generally

The concept of covariant derivative is much more general than that along a curve. One may take any vector field on a surface, differentiate it and project back onto the tangent space at each point. However, when you taking the covariant derivative of a general vector field, the result is only a vector field once you specify a direction in which to differentiate. Thus:

Definition 4.13. Let $\mathbf{x} \subset \mathbb{E}^3$ be a parameterized surface and suppose that $\mathbf{v} = d\mathbf{x}(Y)$ is a vector field on *S* (i.e. *Y* is a vector field on *U*). Let *X* be a vector field on *U*. The *covariant derivative* of \mathbf{v} with respect to *X* is the vector field $D_X \mathbf{v}$, which may be written alternately

$$D_X \mathbf{v} = D_X \, \mathbf{d} \mathbf{x}(Y) := \pi^T \, \mathbf{d}_X \mathbf{v} = \pi^T \, \mathbf{d} \mathbf{v}(X) = \pi^T \, \mathbf{d}_X \, \mathbf{d}_Y \mathbf{x} = \pi^T (X[Y[\mathbf{x}]]).$$

As before, we differentiate and then project onto the tangent plane at each point.

In keeping with our program of moving all calculations to the parameterization space, we can define the covariant derivative of a vector field *Y* on *U* by *X* to be $\nabla_X Y$, where

$$d\mathbf{x}(\nabla_X Y) = D_X d\mathbf{x}(Y).$$

The operator *D* is often referred to as the *Levi-Civita connection* (or just the connection) of the surface \mathbf{x} , and ∇ as the Levi–Civita connection of the induced metric $\mathbf{I} = \mathbf{d}\mathbf{x} \cdot \mathbf{d}\mathbf{x}$ on *U*. The full curvature

tensor can be written in terms of ∇ : for any vector fields X, Y, Z on U, we have that¹⁰

$$R(X,Y)Z := (\nabla_X \nabla_Y - \nabla_Y \nabla_X - \nabla_{[X,Y]})Z = K(I(X,Z)Y - I(Y,Z)X),$$

where *K* is the Gauss curvature.

Why connection? The covariant derivative tells you how to parallel transport tangent vectors along curves. If we imagine two nearby points p, q on a surface S joined by a geodesic, then the covariant derivative defines an invertible linear map $T_pS \rightarrow T_qS$ where a tangent vector at p is parallel transported to a tangent vector at q. This *connects* the two tangent spaces. Indeed a smooth choice of connection is equivalent to a choice of covariant derivative. The Levi–Civita is just one of many connections, albeit one with very nice properties.

One of the best ways of thinking about *D* is that it is the natural restriction of the operator d to the tangent space at each point. Indeed, if we think about the equivalent of the moving frame equation $d\mathbf{E} = \mathbf{E}\omega$ (a 3 × 3 matrix equation), we can write

$$D(\mathbf{e}_1, \mathbf{e}_2) = (\mathbf{e}_1, \mathbf{e}_2) \begin{pmatrix} 0 & \omega_{12} \\ -\omega_{12} & 0 \end{pmatrix}.$$

For this reason it is common for to think of 'splitting up' d into pieces:

$$\omega = \begin{pmatrix} 0 & \omega_{12} & \omega_{13} \\ -\omega_{12} & 0 & \omega_{23} \\ \hline -\omega_{13} & -\omega_{23} & 0 \end{pmatrix} = \begin{pmatrix} [D] & -[\mathbf{I}]^T \\ \hline [\mathbf{I}] & 0 \end{pmatrix},$$

where [D], [I] can be thought of as the matrices of the connection D and the second fundamental form II. If you differentiate a vector field, D tells you how much points in a tangent direction, and II how much in a normal direction. This idea can be used to describe differentiating any (even non-tangent) vector field. In higher dimensions, when you have a k-dimensional surface in \mathbb{E}^n , the matrix [I] of the second fundamental form is $k \times (n - k)$, and the 0 in the lower right of ω becomes a $(n - k) \times (n - k)$ skew-symmetric matrix of 1-forms which describes how to differentiate normal vector fields.

Definition 4.14. A vector field $\mathbf{v} = d\mathbf{x}(Y)$ is *parallel* iff $D_X \mathbf{v} = 0$ for all vector fields X on U.

The existence of parallel vector fields is of great importance in applications. In higher dimensions the notion of a parallel tangent frame is very common: for example in the case of a surface in \mathbb{E}^3 , we would want \mathbf{e}_1 , \mathbf{e}_2 to be parallel vector fields. The generalization of the following theorem to higher dimensions is extremely useful.

Proposition 4.15. *There exists a parallel tangent frame iff* K = 0*.*

Proof. Suppose that \mathbf{e}_1 , \mathbf{e}_2 are parallel, then $D\mathbf{e}_1 = \pi^T d\mathbf{e}_1 = \mathbf{e}_2 \omega_{21} = 0$, hence ω_{12} vanishes and thus so does *K*.

Conversely, let K = 0. Then $d\omega_{12} = 0$ and so, at least locally, $\omega_{12} = df$ for some function f. Now consider the new tangent frame

$$(\hat{\mathbf{e}}_1, \hat{\mathbf{e}}_2) = (\mathbf{e}_1, \mathbf{e}_2) \begin{pmatrix} \cos f & -\sin f \\ \sin f & \cos f \end{pmatrix}.$$

Then $D\hat{\mathbf{e}}_1 = (-\sin f \, \mathrm{d}f + \sin f \omega_{12})\mathbf{e}_1 + (\cos f \, \mathrm{d}f - \cos f \omega_{12})\mathbf{e}_2 = 0$, hence $\hat{\mathbf{e}}_1$ is parallel. $\hat{\mathbf{e}}_2$ is parallel similarly, and we have a parallel frame.

¹⁰Here $[X, Y] := X \circ Y - Y \circ X$ is a vector field on *U* known as the *Lie bracket* of *X*, *Y* (see homework).

4.7 Geodesic curvature

We can think of the curvature of a curve in \mathbb{E}^2 as measuring how much the curve deviates from a straight line. Since a geodesic replaces the concept of straight line on a surface, we can look for a measure of how much a general curve deviates from a geodesic.

Definition 4.16. Let $\gamma(t)$ be a unit speed curve contained in an oriented surface. Then $\gamma'(t)$ is a unit length vector field along γ . Indeed the covariant derivative of γ' must be orthogonal both to γ' , and to the unit normal **U** of the surface. Hence

$$D_{\frac{\mathrm{d}}{\mathrm{d}t}}\gamma' = \kappa_g \mathbf{U} \times \gamma'$$

for some function κ_g called the *geodesic curvature of* γ .

Note that the sign of κ_g depends on the orientation of the surface (changing orientation switches the sign of **U**). It is clear that γ is a geodesic $\iff \kappa_g$ is identically zero.

The above can be reformulated using cross products so that we never see the covariant derivative. Indeed for any curve (unit-speed or otherwise), the formula becomes

$$\kappa_g = rac{\left(\gamma' imes \gamma''
ight) \cdot \mathbf{U}}{\left|\gamma'
ight|^3}.$$

This is reminiscent of the formula for curvature of a spacecurve (see homework for a proof).

The geodesic curvature of γ should be viewed as the curvature of the curve detectible to a dweller of the surface who knows nothing of Euclidean space and normal directions. Indeed it is the curvature of the curve that forms the best planar approximation to γ at each point, as the following theorem shows.

Theorem 4.17. Let $\gamma(t)$ be a unit speed curve in a surface **x** and let $\mathbf{U} = \mathbf{U}(0)$ and $\mathbf{T} = \mathbf{T}(0) = \gamma'(0)$ be the unit normal and tangent vectors to $\gamma(t)$ at t = 0.

- 1. The geodesic curvature $\kappa_g(0)$ of γ at zero is the plane curvature at t = 0 of the projection of γ onto the tangent plane to \mathbf{x} at $\gamma(0)$. I.e. of the curve $\hat{\gamma}(t) = \pi_0^T \gamma(t) = \gamma(t) (\gamma(t), \mathbf{U})\mathbf{U}$. Moreover $|\kappa_g| = |\pi^T \gamma''|$.
- 2. The normal curvature $\kappa_n(0)$ of $\gamma(t)$ at zero is the plane curvature at t = 0 of the projection of $\gamma(t)$ onto the space $Span(\mathbf{T}, \mathbf{U})$. I.e. of the curve $\tilde{\gamma}(t) = (\gamma(t), \mathbf{T})\mathbf{T} + (\gamma(t), \mathbf{U})\mathbf{U}$. Moreover $\kappa_n = (\gamma'', \mathbf{U})$.

3.
$$\kappa^2 = \kappa_g^2 + \kappa_n^2$$
.

Proof. 1. $\hat{\gamma}(t) = \gamma(t) - (\gamma(t), \mathbf{U})\mathbf{U}$, where (,) is the scalar product. Thus $\hat{\gamma}' = \gamma' - (\gamma', \mathbf{U})\mathbf{U}$ and $\hat{\gamma}'' = \gamma'' - (\gamma'', \mathbf{U})\mathbf{U}$. Evaluating at t = 0 yields

$$\hat{\gamma}'(0) = \gamma'(0) = \mathbf{T}, \qquad \hat{\gamma}''(0) = \gamma''(0) - (\gamma''(0), \mathbf{U})\mathbf{U}.$$

If we take the orientation of the tangent plane at t = 0 given by the unit normal, we have that $J\mathbf{T} = \mathbf{U} \times \mathbf{T}$, where *J* is 'rotate 90 degrees' (i.e. $\mathbf{T}, \mathbf{U} \times \mathbf{T}, \mathbf{U}$ form an oriented basis). The plane curvature of the curve $\hat{\gamma}$ at t = 0 is thus

$$\hat{\kappa}(0) = \hat{\gamma}''(0) \cdot J\mathbf{T} = (\gamma''(0) - (\gamma''(0), \mathbf{U})\mathbf{U}) \cdot (\mathbf{U} \times \mathbf{T}) = \gamma''(0) \cdot (\mathbf{U} \times \mathbf{T}) = (\mathbf{T} \times \gamma''(0)) \cdot \mathbf{U} = \kappa_g(0)$$

Moreover, since $\hat{\gamma}$ is unit speed at t = 0, we see that $|\kappa_g(0)| = |\hat{\gamma}''| = |\pi^T \gamma''(0)|$.

2. It is clear that $\tilde{\gamma}'(0) = \mathbf{T}$ and $\tilde{\gamma}''(0) = (\gamma''(0), \gamma'(0))\mathbf{U} + (\gamma''(0), \mathbf{U})\mathbf{U} = (\gamma''(0), \mathbf{U})\mathbf{U}$. Taking **U** to be the unit normal **N** for the curve $\tilde{\gamma}$ at t = 0, we clearly have $\tilde{\gamma}''(0) = (\gamma''(0), \mathbf{U})\mathbf{N}$, so that $\tilde{\kappa}(0) = (\gamma''(0), \mathbf{U})$.

Now suppose that $\gamma' = \cos \phi \mathbf{e}_1 + \sin \phi \mathbf{e}_2$, where $\mathbf{e}_1, \mathbf{e}_2$ are principal curvature directions on **x**. Thus

$$\tilde{\kappa}(0) = (\gamma''(0), \mathbf{U}) = -\cos\phi\omega_{13}(0) - \sin\phi\omega_{23}(0).$$

However, \mathbf{e}_1 , \mathbf{e}_2 are curvature directions, thus $\omega_{13}(0) = -k_1 \cos \phi$, and $\omega_{23}(0) = -k_2 \sin \phi$, so that

$$\tilde{\kappa}(0) = k_1 \cos^2 \phi + k_2 \sin^2 \phi = \kappa_n(0),$$

is the normal curvature¹¹ of γ at t = 0.

- 3. At each point, $\gamma'' = \pi^T \gamma'' + (\gamma'', \mathbf{U})\mathbf{U} = \kappa_g \mathbf{U} \times \gamma' + \kappa_n \mathbf{U}$ is an orthogonal decomposition, hence Pythagoras' $\implies \kappa^2 = \kappa_g^2 + \kappa_n^2$.
- **Examples.** 1. On the surface of a sphere of radius *r* it can be seen that curves of constant geodesic curvature are all circles: $\kappa_g = 0$ are great circles, $|\kappa_g| = a$ are circles of geodesic radius $r \tan^{-1}(r^{-1}a^{-1})$ (see the homework).
 - 2. Consider the curve γ on the surface of the cylinder given by wrapping round a sine wave

$$\gamma(t) = \begin{pmatrix} \cos t \\ \sin t \\ \sin t \end{pmatrix}.$$

We calculate all the curvatures.

$$\gamma'(t) = \begin{pmatrix} -\sin t \\ \cos t \\ \cos t \end{pmatrix}, \qquad \gamma''(t) = -\gamma(t), \qquad \gamma' \times \gamma'' = \begin{pmatrix} 0 \\ -1 \\ -1 \end{pmatrix}.$$

Since $\mathbf{U} = (\cos t, \sin t, 0)^T$ is the usual outward pointing normal of the cylinder, we have

$$\kappa_g = \frac{\left(\gamma' \times \gamma''\right) \cdot \mathbf{U}}{\left|\gamma'\right|^3} = \frac{-\sin t}{(1 + \cos^2 t)^{3/2}}, \qquad \kappa = \frac{\left|\gamma' \times \gamma''\right|}{\left|\gamma'\right|^3} = \frac{\sqrt{2}}{(1 + \cos^2 t)^{3/2}}$$

It follows that the normal curvature is given by

$$\kappa_n^2 = \kappa^2 - \kappa_g^2 = \frac{2 - \sin^2 t}{(1 + \cos^2 t)^3} = \frac{1}{(1 + \cos^2 t)^2}$$

Since the surface is curving away from the unit normal (the non-zero principal curvature is -1) we must have $\kappa_n \leq 0$, and so

$$\kappa_n = \frac{-1}{1 + \cos^2 t}.$$

¹¹We use Euler's formula from 162A

We can check this in an alternative way using Euler's formula. The angle θ between γ' and the vertical is given by

$$\cos \theta = \frac{\gamma'}{|\gamma'|} \cdot \begin{pmatrix} 0\\0\\1 \end{pmatrix} = \frac{\cos t}{\sqrt{1 + \cos^2 t}}.$$

Since $k_1 = 0$ and $k_2 = -1$, we have $\kappa_n = -\sin^2 \theta = -(1 - \cos^2 \theta) = -(1 + \cos^2 t)^{-1}$, as above.

5 Integration

We wish to develop integration in \mathbb{R}^n rather than \mathbb{E}^n so that the theory will be independent of Euclidean structure. Rather than integrate functions over open sets in \mathbb{R}^n , we integrate *n*-forms. The advantage is that our story becomes co-ordinate independent: when changing variables, *n*-forms transform through scaling by the Jacobian of the transformation.

5.1 Orientation

Definition 5.1. An *orientation* of \mathbb{R}^n is a choice of which *n*-forms are 'positive'. If x_1, \ldots, x_n are the standard co-ordinate functions on \mathbb{R}^n , then the *standard orientation* is given by taking asserting that all positive multiples of

$$dx_1 \wedge \cdots \wedge dx_n$$

are positive. Co-ordinates y_1, \ldots, y_n are said to be *oriented* if $dy_1 \wedge \cdots \wedge dy_n$ is positive.

From now on we will forget in the abstract that $x_1, ..., x_n$ are the standard co-ordinate functions, and just assume that these are oriented co-ordinates on some open set *U*.

Now let ω be an *n*-form defined on $U \subset \mathbb{R}^n$. Let x_1, \ldots, x_n be oriented co-ordinates on U. Then, since the set of *n*-forms on \mathbb{R}^n has dimension 1 at each point, there exists a smooth function $f : U \to \mathbb{R}$ such that

$$\omega = f(x_1,\ldots,x_n)\,\mathrm{d} x_1\wedge\cdots\wedge\mathrm{d} x_n.$$

Definition 5.2. The *integral of* ω over *U* is defined as

$$\int_U \omega = \int_U f(x_1,\ldots,x_n) \, \mathrm{d} x_1 \cdots \, \mathrm{d} x_n,$$

provided the integral on the right hand side exists.

Proposition 5.3. Let x_1, \ldots, x_n be oriented co-ordinates on $U \subset \mathbb{R}^n$. Let $y_1, \ldots, y_n : U \to \mathbb{R}$ be differentiable functions. Then

$$dy_1 \wedge \cdots \wedge dy_n = \det\left(\frac{\partial y_i}{\partial x_j}\right) dx_1 \wedge \cdots \wedge dx_n$$

The determinant is the usual Jacobian matrix:¹² i.e. the determinant of the matrix whose ij-th entry is $\frac{\partial y_i}{\partial x_j}$. It follows that the integral of an n-form is independent of the choice of oriented co-ordinates which we integrate with respect to.

¹²This matrix is often written $\frac{\partial(y_1,...,y_n)}{\partial(x_1,...,x_n)}$ in books.

Proof. Recall that $dy_i \left(\frac{\partial}{\partial x_j}\right) = \frac{\partial y_i}{\partial x_j}$. Since any two *n*-forms differ only by a multiple, and are multilinear maps, we need only evaluate on a family on *n* linearly independent vectors. Using the formula on page 4 we see that

$$dy_1 \wedge \cdots \wedge dy_n \left(\frac{\partial}{\partial x_1}, \dots, \frac{\partial}{\partial x_n}\right) = \det\left(\frac{\partial y_i}{\partial x_j}\right)$$

However, this is clearly the same result as obtained by evaluating the right hand side of the formula in the proposition on the same *n*-vectors.

Suppose that $\omega = g(y_1, ..., y_n) dy_1 \wedge \cdots \wedge dy_n$, and that $f(x_1, ..., x_n) = g(y_1, ..., y_n)$. If $y_1, ..., y_n$ are oriented co-ordinates, then the determinant above is positive. From multi-variable calculus, we know that

$$\int_{U} g(y_1,\ldots,y_n) \, \mathrm{d} y_1 \cdots \, \mathrm{d} y_n = \int_{U} f(x_1,\ldots,x_n) \left| \det \left(\frac{\partial y_i}{\partial x_j} \right) \right| \, \mathrm{d} x_1 \cdots \, \mathrm{d} x_n.$$

The left hand side is the integral of ω evaluated using the co-ordinates y_i , while the right hand side (since the determinant is positive without having to take absolute values) is the same integral evaluated using the co-ordinates x_i . The integral $\int_U \omega$ is thus independent of choice of oriented co-ordinates.

It is an obvious follow-on from the proof that if you change the orientation of the co-ordinates, the sign of $\int_U \omega$ changes. The Proposition says that change of variables is built into this definition of an integral. We simply have to be careful to change variables only to other *oriented* co-ordinates.

Examples. 1. Let $\omega = dx \wedge dy$, U = unit disk. Then

$$\int_U \omega = \int_U 1 \, \mathrm{d}x \, \mathrm{d}y = \pi$$

In polar co-ordinates, $dx \wedge dy = r dr \wedge d\theta$ and so

$$\int_U \omega = \int_0^{2\pi} \int_0^1 r \, \mathrm{d}r \, \mathrm{d}\phi = \pi.$$

2. Let $\omega = e^{-(x+y+z)} dx \wedge dy \wedge dz$, and U is the positive octant $\{(x,y,z) : x > 0, y > 0, z > 0\}$. Then

$$\int_{U} \omega = \int_0^\infty \int_0^\infty \int_0^\infty e^{-(x+y+z)} \,\mathrm{d}x \,\mathrm{d}y \,\mathrm{d}z = 1.$$

5.2 Integration over surfaces

We now know how to integrate 3-forms over open sets $U \subset \mathbb{R}^3$ and 1-forms over (segments of) curves in \mathbb{R}^3 . We now want to integrate 2-forms over surfaces in \mathbb{R}^3 . We will see that a Euclidean structure is not necessary in order to integrate.

Definition 5.4. Let $f : U \to \mathbb{R}^3$ and $g : \mathbb{R}^3 \to \mathbb{R}$ be smooth functions. Then dg is a 1-form on \mathbb{R}^3 . The *pull-back* of dg by f is the 1-form on U defined by

$$f^* \, \mathrm{d}g := \, \mathrm{d}(g \circ f).$$

Now let $\alpha = \sum_{i < j} a_{ij} dx_i \wedge dx_j$ be a general 2-form on \mathbb{R}^3 . The pull-back of α by f is the 2-form

$$f^*\alpha = \sum_{i < j} (a_{ij} \circ f) (f^* dx_i) \land (f^* dx_j) = \sum_{i < j} (a_{ij} \circ f) d(x_i \circ f) \land d(x_j \circ f).$$

There is a little more exterior calculus going on here. We have, as is common, brushed something about the exterior derivative under the carpet. We've said that if $g : U \to \mathbb{R}^m$ is a function, then, for each $p \in U$, $dg|_p : T_pU \to \mathbb{R}^m$ is a linear map. It is more correct to say that $dg|_p : T_pU \to T_{g(p)}\mathbb{R}^n$. I.e. $dg|_p$ maps tangent vectors to U at p to tangent vectors to \mathbb{R}^m at g(p). The map g sends points in U to points in \mathbb{R}^m , thus $dg|_p$ sends vectors based at p to vectors based at g(p). When n = 1, it is standard practice to write $T_p\mathbb{R} = \mathbb{R}$ and treat dg(v) as a number, rather than a 'tangent vector to \mathbb{R}' .

Consider the following labelling of bases and attendant co-ordinate systems required to describe a compound map $f \circ g$. For example, we choose a basis E_1, \ldots, E_m of \mathbb{R}^m and co-ordinates y_1, \ldots, y_m : $\mathbb{R}^m \to \mathbb{R}$ with respect to this basis (e.g. $y_3(2E_1 + 3E_3 - 5E_4) = 3$). The identification $\mathbb{R}^m \cong T_p \mathbb{R}^m$ is then $E_j \mapsto \frac{\partial}{\partial y_i}$.

Functions	$U \subset \mathbb{R}^l$ –	$\xrightarrow{g} \mathbb{R}^m$	$\xrightarrow{f} \mathbb{R}^n$
Differentials	T_pU —	$\xrightarrow{\mathrm{d}g} T_{g(p)} \mathbb{R}^m -$	$\xrightarrow{\mathrm{d}f} T_{f(g(p))} \mathbb{R}^n$
Basis vectors		E_j	\hat{E}_k
Co-ordinate functions	x_i	y_j	z_k
1-forms	dx_i	dy _i	$\mathrm{d}z_k$
Tangent vectors	$\frac{\partial}{\partial x_i}$	$\frac{\partial}{\partial y_j}$	$rac{\partial}{\partial z_k}$

In what follows, the tensor product notation $dx_i \otimes \frac{\partial}{\partial y_j}$ denotes the rank one linear map $T_p U \rightarrow T_{g(p)} \mathbb{R}^m$ which sends $\frac{\partial}{\partial x_i}$ to $\frac{\partial}{\partial y_j}$ and all other basis tangent vectors to zero.

With respect to the bases $\{E_j\}$ and $\{\hat{E}_k\}$, there exist functions $f_k : \mathbb{R}^m \to \mathbb{R}$ and $g_j : U \to \mathbb{R}$ such that

$$f = \sum_{k=1}^n f_k \hat{E}_k, \qquad g = \sum_{j=1}^m g_j E_j.$$

But then

$$\mathrm{d}f = \sum_{j=1}^m \sum_{k=1}^n \frac{\partial f_k}{\partial y_j} \,\mathrm{d}y_j \otimes \frac{\partial}{\partial z_k}, \qquad \mathrm{d}g = \sum_{i=1}^l \sum_{j=1}^m \frac{\partial g_j}{\partial x_i} \,\mathrm{d}x_i \otimes \frac{\partial}{\partial y_j}.$$

With respect to the bases $\frac{\partial}{\partial x_i}$, $\frac{\partial}{\partial y_j}$, $\frac{\partial}{\partial z_k}$, the matrices of the above linear maps are simply their Jacobians. It follows by the chain rule that

$$df \circ dg = \sum_{i=1}^{l} \sum_{j=1}^{m} \sum_{k=1}^{n} \frac{\partial f_{k}}{\partial y_{j}} \frac{\partial g_{j}}{\partial x_{i}} \left(dy_{j} \otimes \frac{\partial}{\partial z_{k}} \right) \circ \left(dx_{i} \otimes \frac{\partial}{\partial y_{j}} \right)$$
$$= \sum_{i=1}^{l} \sum_{k=1}^{n} \frac{\partial (f_{k} \circ g)}{\partial x_{i}} dx_{i} \otimes \frac{\partial}{\partial z_{k}} = d(f \circ g).$$

Applying this reasoning to the notion of a pull-back, we see that $f^* dx_i = d(x_i \circ f)$ may be written $dx_i \circ df$. Indeed, if α is any form, we may compactly write $f^*\alpha = \alpha \circ df$.

Definition 5.5. Let $f : U \to \mathbb{R}^3$ be an oriented local surface and let α be a 2-form on \mathbb{R}^3 . The *integral* of α over $\Sigma := f(U)$ is given by

$$\int_{\Sigma} \alpha = \int_{U} f^* \alpha.$$

The orientation of Σ induces an orientation on U in the following way. Suppose that u, v are co-ordinates on U. The unit normal vector **U** to Σ is a positive multiple of either $f_u \times f_v$ or $f_v \times f_u$. If the former, then we say that u, v are oriented co-ordinates on U and take $du \wedge dv$ to be positive. Changing the orientation of Σ changes the orientation on U.

Examples. 1. Let $\alpha = x_3 dx_1 \wedge dx_2 - x_1 dx_2 \wedge dx_3$ and

$$f(u,v) = \begin{pmatrix} uv \\ u^2 \\ v \end{pmatrix},$$

defined on $U = (0,1) \times (0,1)$ with orientation¹³ $\mathbf{U} = \frac{1}{|f_u \times f_v|} f_u \times f_v$. Then

$$\begin{aligned} x_1 \circ f &= uv, \qquad x_2 \circ f = u^2, \qquad x_3 \circ f = v, \\ f^* \, \mathrm{d}x_1 &= v \, \mathrm{d}u + u \, \mathrm{d}v, \qquad f^* \, \mathrm{d}x_2 &= 2u \, \mathrm{d}u, \qquad f^* \, \mathrm{d}x_3 &= \mathrm{d}v, \\ f^* (\, \mathrm{d}x_1 \wedge \mathrm{d}x_2) &= -2u^2 \, \mathrm{d}u \wedge \mathrm{d}v, \qquad f^* (\, \mathrm{d}x_2 \wedge \mathrm{d}x_3) &= 2u \, \mathrm{d}u \wedge \mathrm{d}v, \\ f^* \alpha &= -4u^2 v \, \mathrm{d}u \wedge \mathrm{d}v. \end{aligned}$$

Thus

$$\int_{f} \alpha = \int_{U} f^{*} \alpha = \int_{0}^{1} \int_{0}^{1} -4u^{2}v \, \mathrm{d}u \, \mathrm{d}v = -\frac{2}{3}.$$

2. Let $\alpha = x_1 x_3 dx_2 \wedge dx_3$ be a 2-form of \mathbb{R}^3 , $U = (0, 1) \times (0, 2\pi)$, and

$$f(u,\phi) = \begin{pmatrix} u\cos\phi\\ u\sin\phi\\ u \end{pmatrix}$$

be the cone, with orientation $\mathbf{U} = \frac{1}{|f_u \times f_v|} f_u \times f_v$. The composition $x_1 \circ f$ here is just $u \cos \phi$. Thus

$$f^*\alpha = u^2 \cos \phi \, \mathrm{d}(u \sin \phi) \wedge \mathrm{d}u = u^3 \cos^2 \phi \, \mathrm{d}\phi \wedge \mathrm{d}u,$$

and so

$$\int_f \alpha = \int_U u^3 \cos^2 \phi \, \mathrm{d}\phi \wedge \mathrm{d}u = -\int_0^{2\pi} \int_0^1 u^3 \cos^2 \phi \, \mathrm{d}u \, \mathrm{d}\phi = -\frac{\pi}{4}.$$

5.3 Stokes' theorem

Definition 5.6. For the purposes of what follows, a *k*-dimensional submanifold Σ of \mathbb{R}^3 with boundary $\partial \Sigma$ is;¹⁴

k = 1 An oriented curve with boundary the two ends and no self-intersections.

¹³I.e. we take the orientation such that $du \wedge dv$ is positive.

¹⁴We are being very loose with terminology here: a submanifold is a much more precisely defined object.

k = 2 An oriented surface bounded by a single curve $\partial \Sigma$.

k = 3 A domain (subset with no holes) $\Sigma \subset \mathbb{R}^3$ oriented as usual and bounded by a surface $\partial \Sigma$.

We can even put an orientation on a 0-dimensional manifold, a point, simply by attaching a choice of ± 1 .

- **Definition 5.7.** 1. Let Σ be an oriented curve in \mathbb{R}^3 . The *induced orientation* on $\partial \Sigma$ (the two endpoints of the curve) is to attach a -1 to starting end of the curve, and a +1 to the terminal end. Alternatively, let u_1 be an oriented co-ordinate on the curve: if $\frac{\partial}{\partial u_1}$ points from the curve towards the endpoint, then that endpoint is positively oriented.
 - 2. Let Σ be an oriented surface in \mathbb{R}^3 with boundary an oriented curve $\partial \Sigma$. The *induced orientation* on $\partial \Sigma$ is such that if u_1, u_2 are oriented co-ordinates on Σ with $u_1 \leq 0$ and $u_1 = 0$ on the boundary then u_2 is an oriented co-ordinate for $\partial \Sigma$.
 - 3. Now let Σ be an oriented domain in \mathbb{R}^3 , with boundary $\partial \Sigma$. The induced orientation on $\partial \Sigma$ is such that if u_1, u_2, u_3 are oriented co-ordinates on Σ , with $u_1 \leq 0$ and $u_1 = 0$ on the boundary, then u_2, u_3 are oriented co-ordinates for $\partial \Sigma$.

While this definition seems complicated, it really exists just so that we could deal with induced orientations in a dimension-independent way. What is really going on is this: on a surface Σ with co-ordinates as described in part 1, then, at the boundary curve $\partial \Sigma$, the tangent vector $\frac{\partial}{\partial u_1}$ is pointing out of the surface orthogonally to the boundary. Thus if u_1, u_2 are oriented co-ordinates on Σ , then $\frac{\partial}{\partial u_2}$ must point along the curve in such a way that $\left\{\frac{\partial}{\partial u_1}, \frac{\partial}{\partial u_2}, \mathbf{U}\right\}$ is an adapted frame for Σ . I.e. the induced orientation is found using the right normal, then the surface is on your left. Figure 14 shows the induced orientation on $\partial \Sigma$ when Σ is first a curve, then a surface.



Figure 14: Induced orientations on boundaries

The induced orientation on a domain Σ with bounding surface $\partial \Sigma$ is for the normal to be outward pointing: indeed this normal is $\frac{\partial}{\partial u_1}$ according to the definition.

Theorem 5.8 (Stokes). Let Σ be a k-dimensional submanifold of \mathbb{R}^n with boundary $\partial \Sigma$, given the induced orientation. Let α be a (k-1)-form defined on a neighborhood of Σ in \mathbb{R}^n . Then,

$$\int_{\Sigma} \mathrm{d}\alpha = \int_{\partial \Sigma} \alpha$$

Proof for a square in \mathbb{R}^2 . Let x_1, x_2 be co-ordinates on Σ scaled such that the boundary of Σ is the set

$$\partial \Sigma = \{(x_1, 0), (x_1, 1) : 0 \le x_1 \le 1\} \cup \{(0, x_2), (1, x_2) : 0 \le x_2 \le 1\}.$$

Let $\alpha = f(x_1, x_2) dx_1 + g(x_1, x_2) dx_2$ be a general 1-form on Σ . Taking $dx_1 \wedge dx_2$ as positive on Σ , the induced orientation on $\partial \Sigma$ is then counter-clockwise. We thus have

$$\int_{\partial \Sigma} \alpha = \int_0^1 (f(x_1, 0) - f(x_1, 1)) \, \mathrm{d}x_1 + \int_0^1 (g(1, x_2) - g(0, x_2)) \, \mathrm{d}x_2.$$

However

$$\begin{split} \int_{\Sigma} d\alpha &= \int_{\Sigma} (g_{x_1} - f_{x_2}) \, dx_1 \wedge dx_2 = \int_0^1 \int_0^1 (g_{x_1} - f_{x_2}) \, dx_1 \, dx_2 \\ &= \int_0^1 \int_0^1 g_{x_1} \, dx_1 \, dx_2 - \int_0^1 \int_0^1 f_{x_2} \, dx_2 \, dx_1 \\ &= \int_0^1 (g(1, x_2) - g(0, x_2)) \, dx_2 - \int_0^1 (f(x_1, 1) - f(x_1, 0)) \, dx_1 \\ &= \int_{\partial \Sigma} \alpha. \end{split}$$

Example. Let Σ be the triangular region $\{((x, y) \in \mathbb{R}^2 : 0 \le x \le 1, 0 \le y \le x\}$. Let $\alpha = 7x \, dx - 6xy \, dy$. If we take the orientation such that $dx \wedge dy$ is positive, then the induced orientation is to traverse $\partial \Sigma$ counter-clockwise. Now $\int_{\partial \Sigma} \alpha$ is calculated by restricting α along each edge of the triangle. On the first (horizontal) edge $\alpha = 7x \, dx$, on the second (vertical) edge $\alpha = -6y \, dy$, while on the third we have x = y and so $\alpha = (7x - 6x^2) \, dx$. Hence

$$\int_{\partial \Sigma} \alpha = \int_0^1 7x \, \mathrm{d}x + \int_0^1 -6y \, \mathrm{d}y + \int_1^0 (7x - 6x^2) \, \mathrm{d}x = \frac{7}{2} - 3 - \frac{7}{2} + 2 = -1.$$

Now $d\alpha = -6y dx \wedge dy$, so that

$$\int_{\Sigma} d\alpha = \int_0^1 \int_0^x -6y \, dy \, dx = \int_0^1 -3x^2 \, dx = -1.$$

The full proof of Stokes' theorem for a general region is beyond this course,¹⁵ so we'll content ourselves with observing that in \mathbb{E}^3 it reduces to one of three fundamental theorems.

- k = 1 The fundamental theorem of calculus: $\int_C df = f(b) f(a)$. Integration of a 0-form (function) over a point is equivalent to evaluation, the \pm signs on the right hand side coming from the induced orientation on the endpoints of the curve.
- k = 2 Here α is a 1-form. Using our identification of 1-forms and 2-forms with vectors as in Section 1.1 we recover what is often referred to as Stokes' theorem in vector calculus classes:

$$\int_{S} \nabla \times \mathbf{A} \cdot \mathbf{dS} = \oint_{C} \mathbf{A} \cdot \mathbf{dr}.$$

¹⁵Proving for cubic submanifolds is a straightforward generalization of the above argument, but for general submanifolds we need the concept of partitions of unity in order to patch together integrals on overlapping cubic regions.

The surface integral of the curl of a vector field is identified with the path integral of said vector field over the boundary curve.

Recall our identification of 1-forms with Euclidean vector fields:

$$\alpha = \alpha_1 \, \mathrm{d} x_1 + \alpha_2 \, \mathrm{d} x_2 + \alpha_3 \, \mathrm{d} x_3 \iff \mathbf{A} = \alpha_1 \mathbf{e}_1 + \alpha_2 \mathbf{e}_2 + \alpha_3 \mathbf{e}_3.$$

Thus if *C* is a curve parameterized by $\gamma(t)$ on a curve *I*, then

$$\int_C \alpha = \int_I \gamma^* \alpha = \int_I \alpha(\gamma'(t)) \, \mathrm{d}t$$

Similarly identifying

$$\gamma'(t) = \gamma_1 \frac{\partial}{\partial x_1} + \gamma_2 \frac{\partial}{\partial x_2} + \gamma_3 \frac{\partial}{\partial x_3} \iff \mathbf{dr} := \gamma_1 \mathbf{e}_1 + \gamma_2 \mathbf{e}_2 + \gamma_3 \mathbf{e}_3,$$

we see that

$$\oint_C \alpha = \int_I \sum_{i=1}^3 \alpha_i(\gamma(t)) \gamma_i(t) \, \mathrm{d}t = \oint_C \mathbf{A} \cdot \, \mathrm{d}\mathbf{r}.$$

Thus $\oint_C \alpha$ is the line integral of the vector field **A** round the curve. Now recall that $d\alpha \leftrightarrow \nabla \times \mathbf{A}$ (the curl of **A**). Suppose that *S* is parameterized by $f : U \to \mathbb{E}^3$. Then $f^* d\alpha = d\alpha \circ df$. We have

$$d\alpha = \beta_1 dx_2 \wedge dx_3 + \beta_2 dx_3 \wedge dx_1 + \beta_3 dx_1 \wedge dx_2 \iff \nabla \times \mathbf{A} = \beta_1 \mathbf{e}_1 + \beta_2 \mathbf{e}_2 + \beta_3 \mathbf{e}_3.$$

We thus have to identify the pull-back $f^* d\alpha$. For this, note that $d(x_1 \circ f)$ is simply the \mathbf{e}_1 component of the differential df. The pull-backs can thus be seen to be $f^* d\mathbf{x}_i = df \cdot \mathbf{e}_i$ for
each i = 1, 2, 3. From this we have that $f^* dx_1 \wedge dx_2$ is the \mathbf{e}_3 -component of the cross product of $f_u \times f_v du \wedge dv$ (and similarly). It follows that

$$f^* d\alpha = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix} \cdot (f_u \times f_v) du \wedge dv = (\nabla \times \mathbf{A}) \cdot (f_u \times f_v) du \wedge dv.$$

Hence $\int_{S} d\alpha = \int (\nabla \times \mathbf{A}) \cdot (f_u \times f_v) du dv = \int (\nabla \times \mathbf{A}) \cdot d\mathbf{S}$, where d**S** is the 'surface element' of *S*.

k = 3 In a similar way as for k = 2 we get the divergence theorem

$$\int_V \nabla \cdot \mathbf{A} \, \mathrm{d}V = \int_S \mathbf{A} \cdot \, \mathrm{d}\mathbf{S}.$$

Corollary 5.9 (Green's theorem in the plane). Let U be a domain in the plane with boundary curve C. The induced orientation on C is counter-clockwise due to the standard orientation on U. Stokes' theorem for the 1-form $\alpha = P(x,y) dx + Q(x,y) dy$ is,

$$\oint_C (P \, \mathrm{d} x + Q \, \mathrm{d} y) = \int_U \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} \right) \, \mathrm{d} x \wedge \mathrm{d} y.$$

Stokes' theorem in fact holds for more general objects than our definition of submanifold. For example a segment of a cylinder has two boundary circles, so that $\partial \Sigma$ consists of 2 curves.

6 Applications in Euclidean space

6.1 Integrating functions

From multi-variable calculus, we have that the integral of a function $f : \mathbb{E}^3 \to \mathbb{R}$ over a local surface $\mathbf{x} : U \to \mathbb{E}^3$ is given by,

$$\int_{\text{surface}} f = \int_{U} f(\mathbf{x}(u, v)) |\mathbf{x}_{u} \times \mathbf{x}_{v}| \, \mathrm{d}u \, \mathrm{d}v.$$

The $|\mathbf{x}_u \times \mathbf{x}_v| du dv$ factor is the area of an infinitessimal parallelogram on the surface. Taking f = 1 gives us the area of a piece of a local surface.

In *Euclidean space* we can integrate a function over a surface (rather than 2-forms over surfaces in \mathbb{R}^3). This is because the notion of distance in Euclidean space means that we also have a notion of area.

Proposition 6.1. Suppose we have a local surface in \mathbb{E}^3 referred to an adapted frame, and let u, v be oriented co-ordinates. Then

$$\theta_1 \wedge \theta_2 = |\mathbf{x}_u \times \mathbf{x}_v| \, \mathrm{d}u \wedge \mathrm{d}v.$$

Proof. Consider $d\mathbf{x} \stackrel{\wedge}{\times} d\mathbf{x}$ (a vector of 2-forms where we simultaneously take wedge products of 1-forms and cross products of vectors). On the one hand,

$$d\mathbf{x} \stackrel{\wedge}{\times} d\mathbf{x} = (\mathbf{x}_u \, du + \mathbf{x}_v \, dv) \stackrel{\wedge}{\times} (\mathbf{x}_u \, du + \mathbf{x}_v \, dv) = 2(\mathbf{x}_u \times \mathbf{x}_v) \, du \wedge dv$$
$$= 2|\mathbf{x}_u \times \mathbf{x}_v|\mathbf{e}_3 \, du \wedge dv.$$

Here we use the fact that u, v are oriented, so that $\mathbf{x}_u \times \mathbf{x}_v$ is a positive multiple of the unit normal \mathbf{e}_3 . However, on the other hand,

$$\mathbf{d}\mathbf{x}\overset{\wedge}{\times}\mathbf{d}\mathbf{x}=(\theta_1\mathbf{e}_1+\theta_2\mathbf{e}_1)\overset{\wedge}{\times}(\theta_1\mathbf{e}_1+\theta_2\mathbf{e}_1)=2\mathbf{e}_3\theta_1\wedge\theta_2,$$

giving the result.

The Proposition shows that the integral of a function over a surface is given by,

$$\int_{U} f(\mathbf{x}(u,v))\theta_1 \wedge \theta_2.$$

In particular, if $f \equiv 1$, we have the area of the surface. For this reason, the 2-form $\theta_1 \wedge \theta_2$ is often called the *area form* for the surface **x**.

Examples. 1. The sphere of radius *a* has $\theta_1 = a d\theta$, $\theta_2 = a \sin \theta d\phi$. If *z* is the standard vertical co-ordinate on \mathbb{E}^3 , we have,

(a)
$$\int_{S} z = \int_{U} z(\mathbf{x}(\theta, \phi))\theta_{1} \wedge \theta_{2} = \iint_{U} a \cos \theta a^{2} \sin \theta \, \mathrm{d}\theta \, \mathrm{d}\phi = 0$$

(b) Surface Area
$$= \int_{S} 1 = \iint_{U} a^{2} \sin^{2} \alpha \, \mathrm{d}\alpha \, \mathrm{d}\phi = 4\pi a^{2}.$$

2. In hyperbolic space the area form is $\frac{(dx)^2+(dy)^2}{y^2}$. To find the area of the infinite region *U* (shown in Figure 15) bounded by the geodesics $x = \pm 1$, $x^2 + y^2 = 1$ we evaluate,

$$\int_{U} \theta_1 \wedge \theta_2 = \int_{-1}^{1} \int_{\sqrt{1-x^2}}^{\infty} \frac{1}{y^2} \, \mathrm{d}x \, \mathrm{d}y = \int_{-1}^{1} \frac{1}{\sqrt{1-x^2}} \, \mathrm{d}x = \pi.$$

Conversely, the hyperbolic area between the geodesic $x^2 + y^2 = 1$ and the *x*-axis is infinite.



Figure 15: A region of hyperbolic space with area π

The same notation may be used to integrate functions over curves in Euclidean space. If we write $\int_{\gamma} f$, we mean $\int_{s_0}^{s_1} f(\gamma(s)) ds$, where $\gamma(s_0), \gamma(s_1)$ are the endpoints of γ , and $|\gamma'(s)| = 1$ (γ is parameterized by arc-length). Why do we need the Euclidean structure for this? Because without it we have no notion of *length* by which to parameterize and so we are forced to integrate only 1-forms.

6.2 Minimal surfaces

Now that we know how to integrate functions on surfaces and calculate surface area, we can approach the problem of finding minimal surfaces: i.e. those whose area is minimal for all surfaces with a given boundary. The argument is similar to the stationary distance approach to geodesics.

Let $\mathbf{x} : U \to \mathbb{E}^3$ be a surface, and consider the family of surfaces

$$\mathbf{x}_{\epsilon}(u,v) = \mathbf{x}(u,v) + \epsilon f(u,v)\mathbf{U}(u,v),$$

nearby **x**. Here f is a smooth function on U of compact support.¹⁶ The idea is that we start with **x** and perturb it by a small amount in the normal direction at each point in such a way that the boundary is unchanged.

Definition 6.2. A local surface has stationary area if for all such families,

$$\frac{\mathrm{d}}{\mathrm{d}\epsilon}\Big|_{\epsilon=0}\operatorname{Area}(\mathbf{x}_{\epsilon})=0.$$

Theorem 6.3. A surface has stationary area iff its mean curvature vanishes.

Proof. Take the exterior derivative of \mathbf{x}_{ϵ} .

$$\mathbf{d}\mathbf{x}_{\epsilon} = \mathbf{d}\mathbf{x} + \epsilon (\,\mathbf{d}f\,\mathbf{U} + f\,\mathbf{d}\mathbf{U}).$$

Just like in the geodesic arguments, we throw away all terms of order ϵ^2 and higher. We first calculate the first fundamental forms of the surfaces \mathbf{x}_{ϵ} .

$$\begin{split} \mathbf{I}_{\epsilon} &\simeq \, \mathbf{d} \mathbf{x} \cdot \mathbf{d} \mathbf{x} + 2\epsilon f \, \mathbf{d} \mathbf{U} \cdot \mathbf{d} \mathbf{x} = \mathbf{I} - 2\epsilon f \mathbf{I} \mathbf{I} \\ &\simeq \theta_1^2 + \theta_2^2 + 2\epsilon f(\omega_{13}\theta_1 + \omega_{23}\theta_2) \end{split}$$

¹⁶The only important point is that if f extends to the curve parameterizing the boundary of \mathbf{x} , then f would be zero there.

$$\simeq (\theta_1 + \epsilon f \omega_{13})^2 + (\theta_2 + \epsilon f \omega_{23})^2$$

We can therefore diagonalize I_{ϵ} up to first order in ϵ be choosing $\theta_{1,\epsilon} \simeq \theta_1 + \epsilon f \omega_{13}$ and $\theta_{2,\epsilon}^2 \simeq \theta_2 + \epsilon f \omega_{23}$. Hence,

$$\begin{aligned} \theta_{1,\epsilon} \wedge \theta_{2,\epsilon} &\simeq \theta_1 \wedge \theta_2 + \epsilon f(\omega_{13} \wedge \theta_2 + \theta_1 \wedge_{23}) \\ &\simeq \theta_1 \wedge \theta_2 + \epsilon f((a\theta_1 + b\theta_2) \wedge \theta_2 + \theta_1 \wedge (b\theta_1 + c\theta_2)) \\ &\simeq \theta_1 \wedge \theta_2 (1 + \epsilon f(a + c)) = (1 - 2\epsilon f H) \theta_1 \wedge \theta_2, \end{aligned}$$

where we've written $\omega_{13} = a\theta_1 + b\theta_2$ and $\omega_{23} = b\theta_1 + c\theta_2$. It follows that

$$\frac{\mathrm{d}}{\mathrm{d}\epsilon}\Big|_{\epsilon=0}\operatorname{Area}(\mathbf{x}_{\epsilon})=-2\int_{U}fH\theta_{1}\wedge\theta_{2},$$

which vanishes for all functions *f* iff $H \equiv 0$.

Thus minimal surfaces as defined in 162A are critical surfaces of the area functional. It certainly follows that any surface which has minimal area for a given perimeter must have mean curvature zero, although it is perfectly possible to have H = 0 for a surface which is not area minimizing.

6.3 Relations with Complex analysis

Let $\mathbb{R}^2 = \mathbb{C}$ the complex plane with complex co-ordinate z = x + iy and $\overline{z} = x - iy$. We can consider complex functions and complex differential forms. Exterior derivatives of real and imaginary parts can be taken. In particular we can define

$$dz = d(x + iy) = dx + i dy, \qquad d\overline{z} = dx - i dy.$$

dz and d \overline{z} can be used as a basis for complex 1-forms instead of dx, dy. Defining two complex vector fields,

$$\partial := \frac{\partial}{\partial z} = \frac{1}{2} \left(\frac{\partial}{\partial x} - i \frac{\partial}{\partial y} \right), \qquad \overline{\partial} = \frac{\partial}{\partial \overline{z}} = \frac{1}{2} \left(\frac{\partial}{\partial x} + i \frac{\partial}{\partial y} \right),$$

it is easy to see that any complex function satisfies,

$$\mathrm{d}f = \partial f\,\mathrm{d}z + \partial f\,\mathrm{d}\overline{z}.$$

Proposition 6.4. *f* is holomorphic iff $\overline{\partial} f$.

Proof. Let f = u + iv be the decomposition of f into real and imaginary parts. Then

$$\overline{\partial}f = \frac{1}{2}\left(\frac{\partial}{\partial x}(u+iv) + i\frac{\partial}{\partial y}(u+iv)\right) = \frac{1}{2}(u_x - v_y + i(v_x + u_y)).$$

Thus $\overline{\partial} f = 0 \iff f$ satisfies the *Cauchy–Riemann equations*.

48

The Cauchy–Riemann equations are exactly the condition that a complex function f(z) is differentiable (we say that f is holomorphic). Indeed we require

$$f'(z_0) = \lim_{z \to z_0} \frac{f(z) - f(z_0)}{z - z_0}$$

to be independent of the direction in which we approach z_0 . It is easy to see that independence is equivalent to the Cauchy–Riemann equations.

The proposition makes clear the statement that a holomorphic function is 'independent of \overline{z}' .

Complex 1-forms and 2-forms can be integrated be integrating real and imaginary parts respectively. One nice conclusion of Green's theorem in the plane comes form applying it to the 1-form $\alpha = f(z, \overline{z}) dz$. Then

$$\mathrm{d}\alpha = \mathrm{d}f \wedge \mathrm{d}z = (\partial f \, \mathrm{d}z + \overline{\partial}f \, \mathrm{d}\overline{z}) \wedge \mathrm{d}z = \overline{\partial}f \, \mathrm{d}\overline{z} \wedge \mathrm{d}z.$$

Green's theorem then says that

$$\int_C f(z,\overline{z}) \, \mathrm{d}z = \int_U \overline{\partial} f \, \mathrm{d}\overline{z} \wedge \mathrm{d}z$$

In particular if *f* is holomorphic then the right hand side is 0 in which case we have Cauchy's theorem, that the integral of a holomorphic function around a closed curve is zero.

7 The Gauss-Bonnet theorem

In this final section we think about the geometry of polygons, specifically geodesic triangles, on surfaces, and prove (using Stokes' theorem) a famous result that relates topological and differential-geometric data on a surface.

7.1 Polygons on surfaces

First we think about how the angle made by a curve with an adapted vector field changes along the curve.

Let $\mathbf{x}(z(t))$ be a unit speed curve in an oriented surface \mathbf{x} with adaptive frame field \mathbf{e}_1 , \mathbf{e}_2 , \mathbf{e}_3 (= **U**). In terms of the angle $\psi(t)$ the tangent vector $\mathbf{x}'(t) = d\mathbf{x}(z'(t))$ makes with \mathbf{e}_1 , we have

$$\mathbf{x}'(t) = \cos\psi(t)\mathbf{e}_1 + \sin\psi(t)\mathbf{e}_2.$$

Lemma 7.1. $\psi'(t) = \kappa_g + \omega_{12}(z').$

Proof. Recall that the geodesic curvature is defined in terms of the covariant derivative: $D_{\frac{d}{dt}}\mathbf{x}' = \kappa_{g}\mathbf{e}_{3} \times \mathbf{x}'$. Calculating we have:

$$D_{\frac{d}{dt}} \mathbf{x}' = \psi'(-\sin\psi\mathbf{e}_1 + \cos\psi\mathbf{e}_2) + \omega_{12}(z')(-\cos\psi\mathbf{e}_2 + \sin\psi\mathbf{e}_2)$$

= $(\psi' - \omega_{12}(z'))(-\sin\psi\mathbf{e}_1 + \cos\psi\mathbf{e}_2)),$
= $(\psi' - \omega_{12}(z'))\mathbf{e}_3 \times \mathbf{x}',$

hence $\kappa_g = \psi' - \omega_{12}(z')$.

Definition 7.2. Suppose that $\mathbf{x}(z(t))$ is a unit speed curve which runs from $t = t_0$ to t_1 . The *total rotation* of the curve with respect to the frame field is given by

$$\int_{t_0}^{t_1} (\omega_{12}(z') + \kappa_g) \, \mathrm{d}t = \psi(t_1) - \psi(t_0).$$

In particular if $\mathbf{x}(z(t))$ is a simple closed curve traversed counter-clockwise, then the above integral is 2π .

We now restrict to polygons whose edges are geodesics.

Definition 7.3. A *polygon* on a surface is a piecewise differentiable curve γ lying in a surface in such a way that γ has no self-intersections and the differentiable pieces of γ are all geodesics.

Figure 16 shows an example. If we orient the surface, then γ bounds a region of the surface *S* and has an induced orientation. We label the geodesic curves of γ as $\gamma_1, \ldots, \gamma_n$ in order of the orientation, and the internal angles A_1, \ldots, A_n .



Figure 16: A polygon on a cylinder

Theorem 7.4. Let γ be a polygon on an oriented surface bounding a region S. Then the internal angles satisfy

$$\sum_{i=1}^{n} A_i = (n-2)\pi + \int_{S} K.$$

Proof. Moving all the way around the polygon clearly rotates the tangent vector field through 2π radians. We know that the total rotation of the vector field along each curve γ_i is $\int_{\gamma_i} \omega_{12}$, while the rotation at each vertex is through an angle $\pi - A_i$. Adding up we have

$$2\pi = \sum_{\substack{i=1\\\text{turning at vertices}}}^{n} (\pi - A_i) + \sum_{\substack{i=1\\\text{turning over each edge}}}^{n} \int_{\gamma_i} \omega_{12}$$
$$= n\pi + \int_{\gamma} \omega_{12} - \sum_{i=1}^{n} A_i.$$

However, using Stokes' theorem,

$$\int_{\gamma} \omega_{12} = \int_{S} \mathrm{d}\omega_{12} = \int_{S} K \theta_{1} \wedge \theta_{2} = \int_{S} K,$$

which gives the result.

7.2 The geometry of a geodesic triangle

Before we move on, we can restrict to the simplest case, where γ is a geodesic triangle.

Definition 7.5. A *geodesic triangle* in a surface is three points joined to each other by (non-intersecting) geodesics.

The above theorem says that the angles in a geodesic triangle \triangle satisfy

$$A+B+C-\pi=\int_{\bigtriangleup}K.$$

Think about what this says for a minute. The familiar fact that the angles in a triangle add up to π is *false* on any surface which is not flat.

Example. If \triangle is a geodesic triangle with angles *A*, *B*, *C* and area *S* then,

- 1. in the plane $A + B + C = \pi$.
- 2. on the unit sphere $A + B + C = \pi + S$.
- 3. in hyperbolic space $A + B + C = \pi S$.

Notice that in hyperbolic space the area *S* of any geodesic triangle satisfies $S \le \pi$. Equality is iff the three angles *A*, *B*, *C* are all zero. This only happens if the geodesics meet on the *x*-axis. We are not saying that the total area of hyperbolic space is $\le \pi$ (in fact it is infinite), rather than the largest area you can fit inside a geodesic triangle is π . Somewhat counter-intuitively, all three edges of this geodesic triangle have infinite length. A similar, but less geometrically interesting result comes from observing that the fact that all angles in a triangle must be less than π forces the maximum area bounded by a geodesic triangle on a sphere of radius 1 to be 2π (the triangle in question is not an interesting triangle, rather it is a complete great circle with three marked points!).

There are many other things we could do in spherical or hyperbolic geometry. There are Sine and Cosine rules in spaces of constant curvature. Suppose we have a geodesic triangle in a space of constant Gauss curvature *K*, whose angles *A*, *B*, *C* are opposite the sides of lengths *a*, *b*, *c* respectively. Then:

$$\cos(c\sqrt{K}) = \cos(a\sqrt{K})\cos(b\sqrt{K}) + \sin(a\sqrt{K})\sin(b\sqrt{K})\cos C,$$
$$\frac{\sin(a\sqrt{K})}{\sin A} = \frac{\sin(b\sqrt{K})}{\sin B} = \frac{\sin(c\sqrt{K})}{\sin C}.$$

For the unit sphere we have K = 1 and the formulae simplify. For hyperbolic space with K = -1 the expressions may be rewritten using $\cos(ia) = \cosh(a)$ and $\sin(ia) = i \sinh(a)$, so that the sines and

cosines of lengths become hyperbolic sines and cosines with appropriate sign changes. To recover the cosine and sine rules in flat space, divide both sides by \sqrt{K} and take the limit as $K \rightarrow 0$.

The cosine rule allows us to recover the constant curvature Pythagoras' theorem: for any right angled triangle with hypotenuse of length *c* we have

$$\cos(c\sqrt{K}) = \cos(a\sqrt{K})\cos(b\sqrt{K}).$$

This doesn't look like the Pythagoras' theorem you're used to, but if you take Taylor approximations up to order two in *K* and substitute $c^4 \simeq (a^2 + b^2)^2$, we recover¹⁷

$$c^2 \simeq a^2 + b^2 - \frac{K}{3}a^2b^2.$$

You should expect the hypotenuse of a small right-angled triangle on a sphere to have shorter hypotenuse than in the plane. On a pseudosphere the hypotenuse should be longer. The approximation is good if the lengths of *a*, *b*, *c* are small when compared to the curvature, typically *a*, *b* < $1/\sqrt{|K|}$, or *a*, *b* < *r* for a sphere. Indeed if $a = b = \mu r$, then the ratio of the length of the above Pythagorean prediction to the correct hypotenuse is $c_P/c = \frac{\mu\sqrt{2-\mu^2/3}}{\cos^{-1}(\cos^2\mu)}$ which is within 2% of 1 for $\mu < 1$ and within 9% for $\mu = \pi/2$. This last is a geodesic triangle with edges going 1/4 of the way round the sphere, larger than you'd ever likely need in navigation.

7.3 The Gauss-Bonnet theorem

For the full version of this result, we will need a stronger version of the theorem of the rotation of a vector field around a geodesic polygon.

Theorem 7.6. Let *S* be a region on an oriented surface bounded by a piecewise differentiable closed curve γ . Suppose that γ has *n* corners and that its internal angles are A_1, \ldots, A_n . Then

$$\sum_{i=1}^n A_i = (n-2)\pi + \int_S K + \int_\gamma \kappa_g,$$

where the last integral¹⁸ is understood to be the sum over all the differentiable pieces of γ .

The proof is identical to that of Theorem 7.4 except that the rotation along each curve γ_i is as given in Definition 7.2, where the geodesic curvature is not (necessarily) zero.

Consider a surface *S* with a boundary ∂S .

Definition 7.7. A *dissection* of *S* is a cutting up of *S* into polygons (for us *geodesic* polygons). The number of faces *F* of the dissection is the number of polygons plus the number of regions with one edge as part of the boundary. Let *E* be the number of edges of the dissection (number of edges of polygons plus resulting boundary segments), and *V* the number of vertices of the dissection. The *Euler characteristic* of *S* is defined to be

$$\chi = F - E + V.$$



Figure 17: Two dissections of a disk

The surface *S* could be simply a region of a plane. Figure 17 shows two possible dissections of a disk. Both have $\chi = 1$.

Proposition 7.8. The Euler characteristic is independent of dissection and thus well-defined.

Proof. Start by observing that if you subdivide a polygon into triangles, then χ is unaltered (join up one vertex of a polygon with another and you get one extra edge, one extra face, and no extra vertices). This means that we need only consider triangular dissections. Triangles can now be subdivided in various ways, again leaving χ unchanged. Thus, given two dissections into triangles, we can subdivide each until we get a common subdissection. χ is thus independent of the choice of dissection.

There is something missing here. A general dissection does not have the edges being geodesics, so a complete proof would have to deal with this.

A region of the plane bounded by a simple closed curve (no intersections) has $\chi = 1$. The sphere (no boundary) has $\chi = 2$ as can be seen in many ways. The torus has $\chi = 0$ (for example take as a dissection to orthogonal circles of curvature to get $\chi = F - E + V = 1 - 2 + 1 = 0$).

To get some examples of Euler characteristics, we do a little surgery and consider how the Euler characteristic changes when you glue surfaces together.

Definition 7.9. Given two topological surfaces Σ_1 , Σ_2 , their *connected sum* $\Sigma_1 \# \Sigma_2$ is given by cutting a small hole in each surface and pasting them together.

By topological surface we mean that nothing matters except the topology: we may deform Σ_1, Σ_2 and glue them together in any way we like and the topology of $\Sigma_1 \# \Sigma_2$ is unchanged.

Note that the connected sum of any surface with a sphere effectively leaves the surface unchanged.

Theorem 7.10. $\chi_{\Sigma_1 \# \Sigma_2} = \chi_{\Sigma_1} + \chi_{\Sigma_2} - 2.$

Proof. Dissect Σ_1, Σ_2 into triangles. Remove one triangle on each surface to create a hole and stick these together. Let F, V, E are the total number of faces, vertices and edges of both original dissections added together, and $\hat{F}, \hat{V}, \hat{E}$ the faces, vertices and edges of the resulting dissection of $\Sigma_1 \# \Sigma_2$. Then $\hat{F} = F - 2$, $\hat{E} = E - 3$, $\hat{V} = V - 3$. The result follows.

¹⁷You can similarly take Taylor approximations of the cosine and sine rules to obtain the standard rules as approximations.

¹⁸Recall that the integral $\int_{\gamma} \kappa_g$ of a function over a curve is defined in Section 6.1.

This allows us to easily calculate χ for any handlebody formed by taking the connected sums of tori (since $\chi_{torus} = 0$, taking the connected sum with a torus decreases χ by 2). The *genus g* of a handlebody is the number of holes in the surface. It is clear from the theorem that $\chi = 2(1 - g)$.

Now that we have many examples of surfaces and their Euler characteristics we proceed to the theorem.

Theorem 7.11 (Gauss-Bonnet). Let Σ be a surface in \mathbb{E}^3 with differentiable boundary $\partial \Sigma$. Then

$$\int_{\Sigma} K + \int_{\partial \Sigma} \kappa_g = 2\pi \chi.$$

Proof. Dissect Σ into triangles, where all edges except those that are part of the boundary $\partial \Sigma$ are geodesics. Let the internal triangles be labeled $\triangle_1, \ldots, \triangle_{F-n}$, and let the triangular faces bordering the boundary be labeled $\triangle_{F-n+1}, \ldots, \triangle_F$. The boundary $\partial \Sigma$ is thus made up of *n* (curved) edges, and there are *n* triangles bordering the boundary.

First consider the geodesic internal edges (of which there are E - n). Each internal edge is common to two faces. Counting each internal edge twice, there are three edges per internal face (F - n of them) and two edges for each boundary face. Thus 2(E - n) = 3(F - n) + 2n and so, for such a dissection, $E = \frac{3F+n}{2}$. The Euler characteristic may therefore be written

$$\chi = F - \frac{3F + n}{2} + V = V - \frac{1}{2}(F + n).$$

Now, for each triangle \triangle_j (internal and external), $A_j + B_j + C_j - \pi = \int_{\triangle_j} K + \int_{\partial \triangle_j} \kappa_g$. Summing over *j* gives

$$\int_{\Sigma} K + \int_{\partial \Sigma} \kappa_g = \sum_{j=1}^{F} (A_j + B_j + C_j) - \pi F.$$

Adding up the angles around all the *V* – *n* internal vertices yields 2π for each, while for each of the *n* boundary vertices we get an angle of π . Hence

$$\int_{\Sigma} K + \int_{\partial \Sigma} \kappa_g = 2\pi (V - n) + \pi n - \pi F = 2\pi (V - (F + n)/2) = 2\pi \chi.$$

The theorem may be modified to include the case where the boundary $\partial \Sigma$ is piecewise differentiable with internal angles at its corners ϕ_i . It can also be simplified to the case where the boundary $\partial \Sigma$ is a closed geodesic, or is empty. In the latter case, we call Σ a *handlebody* (e.g. a sphere, torus, pretzel, etc...).

Corollary 7.12. Let Σ be a handlebody in \mathbb{E}^3 . Then

$$\int_{\Sigma} K = 2\pi \chi.$$

The Gauss–Bonnet theorem is particularly powerful because it relates the topological Euler characteristic to the differential gauss curvature. χ is a single number, which depends only on topology: twisting, squashing or stretching or otherwise altering a surface in any way that does not involve puncturing it, leaves χ unchanged. For example, the famous quip that a doughnut is topologically equivalent¹⁹ to a coffee mug is not just a play on the fact that mathematicians seem to require an abundance of both: they both have $\chi = 0$. A human (supposing of course that we are considered to be without holes...) has $\chi = 2$, seeing as we are just a highly deformed sphere. The Gauss curvature, by contrast, is an entirely local function. Deforming a surface will almost certainly change *K* enormously. The amazing fact is that the total curvature in the surface is unchanged.

In particular, for any deformed sphere, $\int_{\Sigma} K = 4\pi$. For a torus, $\int_{\Sigma} K = 0$. We can say, very non-precisely, that any handlebody with 2 or more holes must have more negative curvature than positive. In particular, any torus embedded in \mathbb{E}^3 must both have points where *K* is positive and negative (since *K* is continuous and not identically zero). Indeed we have proved that the only handlebodies whose curvature is everywhere positive are topological spheres. For more examples of applications see the homework.

The full version of the Gauss–Bonnet theorem reads that for a surface Σ with boundary $\partial \Sigma$ made up of *n* differentiable curves with internal angles A_1, \ldots, A_n we have

$$\int_{\Sigma} K + \int_{\partial \Sigma} \kappa_g + \sum_{i=1}^n (\pi - A_i) = 2\pi \chi.$$

Generializations of the Gauss–Bonnet theorem have enormous applications throughout Geometry and Mathematical Physics.

¹⁹Homeomorphic in technical language.