B3.2 Geometry of Surfaces

Richard Earl

Michaelmas 23

Synopsis (by individual lecture)

0. Background material.

CHAPTER 1

1. Introductory lecture.

CHAPTER 2

2. Topological surfaces. Examples. Euler number. Orientability.

3. Classification theorem. Canonical surfaces. Identifying surfaces.

CHAPTER 3

4. Smooth surfaces in $\mathbb{R}^3$. Examples. Tangent space. Gradient. Abstract smooth surfaces.

5. First fundamental form. Examples. The concept of a Riemannian 2-manifold. Isometries.

6. Second fundamental form. Weingarten map. Gaussian curvature.

7. Theorema Egregium.

CHAPTER 4

8. Geodesics.

CHAPTER 5

9. Local and Global Gauss-Bonnet Theorems and applications.

10. Poincaré-Hopf theorem. Morse theory.

CHAPTER 6

11. The hyperbolic plane, its isometries and geodesics.

12. Hyperbolic geometry and trigonometry.

13. Compact hyperbolic surfaces as surfaces of constant negative curvature.

CHAPTER 7

13-14. Riemann surfaces; examples, including the Riemann sphere, the quotient of the complex numbers by a lattice, and double coverings of the Riemann sphere.

15. Holomorphic maps of Riemann surfaces and the Riemann-Hurwitz formula.

16. Elliptic functions including Weierstrass $\wp$-function.

Recommended Texts

Elementary Differential Geometry, Andrew Pressley, Springer (2010)
Geometry of Surfaces, Graeme Segal, Mathematical Institute Notes (1986)
Geometry from a Differential Viewpoint, John McCleary, Cambridge (2012)
Surface Topology, Peter Firby, Cyril Gardiner, Woodhead (2001)
Complex Algebraic Curves, Frances Kirwan, Cambridge (1992)
Elementary Differential Geometry, Barrett O'Neill, Academic Press (2006)
Differential Geometry of Curves and Surfaces, Manfredo Do Carmo, Dover (2017)
Geometry of Surfaces, John Stillwell, Springer (1992)

Acknowledgements

The majority of these notes is original material, but they also make substantial use of Segal's Institute notes and Ritter's lecture notes.

0. BACKGROUND MATERIAL

There is a good deal of background material in this chapter. It is hoped, given the course's pre-requisites and recommedations, that most students will have met some, perhaps most, of this material but very few may have met it all. At no single point in this course will all this material be simultaneously necessary, but it will be helpful to either do preparatory reading on a topic ahead of the relevant lectures or revisit the material if you find yourself rustier than you expected.

As a guide:

- *arc length, curvature* and *surface area* will prove useful ahead of Chapter 3.
- *the real projective plane*'s topology will also appear in Chapters 1 and 2.
- the real and complex projective planes will appear in Chapter 7.
- *holomorphic branches* will prove useful in Chapter 7.
- *multivariable differentiability* will prove somewhat helpful in Chapter 3.
- *identification spaces* will be important ahead of Chapter 2.

0.1 Arc Length and Curvature

This is largely material from Prelims *Introductory Calculus.*

Definition 0.1 *A* ***smooth parameterized curve*** *in* $\mathbb{R}^3$ *is a map* $\gamma\colon I \to \mathbb{R}^3$ *from an open interval* $I \subseteq \mathbb{R}$ *such that*

- γ *is smooth, i.e.* γ *has derivatives of all orders;*
- $\gamma\colon I \to \gamma(I)$ *is a homeomorphism;*
- $\gamma'(t) \neq \mathbf{0}$ *for all* $t \in I$.

The requirement that γ be a homeomorphism onto its image is somewhat unusual here. Some authors will omit this requirement which allows the possibility of self-intersections, for the curve crossing itself. Defining a smooth parameterized curve as above means that the curve has no singular points and also mirrors the later definition of a smooth parameterized surface (see Definition 1.7).

A smooth parameterized curve γ is a curve in $\mathbb{R}^3$ with a preferred parameterization. The image of γ is also the image of other smooth parameterized curves. It's important to check that our definitions relating to curves and surfaces are independent of the choice of parameterization. For example, a simple application of the chain rule shows that the tangent line to a curve and arc length on a curve (as defined below) are independent of the choice of parameter. Arc length is an 'intrinsic' parameter for a curve.

Definition 0.2 *Let $\gamma\colon I \to \mathbb{R}^3$ be a smooth parameterized curve with $t_0 \in I$. Then the* ***arc length*** *$s(t)$ from $\gamma(t_0)$ to a point $\gamma(t)$ is defined by the integral*

$$s(t) = \int_{t_0}^{t} |\gamma'(u)| \, \mathrm{d}u.$$

As $\gamma'(t) \neq \mathbf{0}$ for all t then there is a well defined *tangent line* at each point of $\gamma(I)$.

Definition 0.3 *Let $\gamma\colon I \to \mathbb{R}^3$ be a smooth parameterized curve with $t_0 \in I$.*

(a) The ***tangent line*** *to γ at $\gamma(t_0)$ is the line containing the point $\gamma(t_0)$ and parallel to $\gamma'(t_0)$.*

(b) The ***unit tangent vector*** *$\mathbf{t}(s)$ is the tangent vector*

$$\mathbf{t}(s) = \frac{\mathrm{d}\gamma}{\mathrm{d}s},$$

when γ is parameterized by arc length s.

(c) The ***curvature*** *$\kappa(s)$ of γ at $\gamma(s)$ is defined to be*

$$\kappa(s) = \left|\frac{\mathrm{d}\mathbf{t}}{\mathrm{d}s}\right| = \left|\frac{\mathrm{d}^2\gamma}{\mathrm{d}s^2}\right|.$$

Example 0.4 *(**Logarithmic spiral**) Let $\gamma(t) = \left(ae^{bt}\cos t, ae^{bt}\sin t\right)$ for $t > 0$ and real constants $a > 0 > b$. Show that γ has finite arc length.*

Solution. The tangent vector $\gamma'(t)$ equals

$$(ae^{bt}(b\cos t - \sin t), ae^{bt}(b\sin t + \cos t)),$$

and has magnitude

$$ae^{bt}\sqrt{((b\cos t - \sin t)^2 + (b\sin t + \cos t)^2)} = ae^{bt}\sqrt{b^2+1}.$$

So the arc length from $\gamma(0) = (a, 0)$ to $\lim_{t\to\infty}\gamma(t) = (0, 0)$ equals

$$a\sqrt{1+b^2}\int_0^\infty e^{bu}\mathrm{d}u = a\sqrt{1+b^{-2}}.$$

■

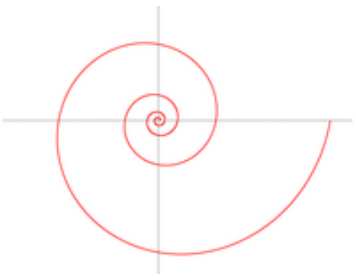


Figure 0.1 – The logarithmic spiral

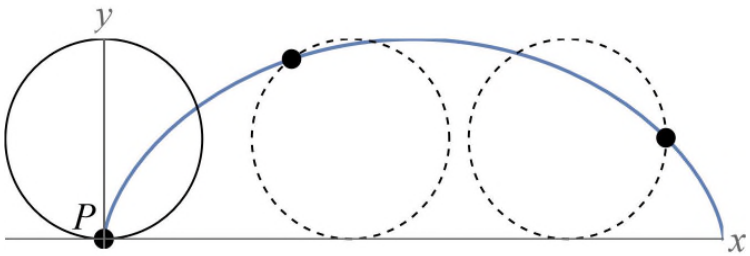


Figure 0.2 – The cycloid

Example 0.5 *The* ***tractrix*** *is the curve given by*

$$\gamma(t) = \left(-\cos t + \log \tan\left(\frac{t}{2}\right), \sin t \right), \qquad 0 < t < \frac{\pi}{2}.$$

Show that the length of the tangent line from a point $\gamma(t)$, to the point where the tangent meets the x-axis, is always 1 (see Figure 3.5).

Solution. Differentiating we find that $\gamma'(t)$ equals

$$\left(\frac{-\cos^2 t}{\sin t}, \cos t \right), \qquad 0 < t < \frac{\pi}{2}.$$

So the tangent from the curve at $\gamma(t)$ meets the x-axis at

$$\gamma(t) + (\cos t, -\sin t),$$

a point distance 1 away. ■

Example 0.6 *A circular disc of radius r in the xy-plane rolls without slipping along the x-axis. The locus described by a point of the circumference of the disc is called a* ***cycloid*** *(see Figure 0.2). Determine the arc length of a section of the cycloid which corresponds to a complete rotation of the disc.*

Solution. Assume that the disc begins with its centre at $(0, r)$. Consider the curve described by the point $(0, 0)$ as the disc rolls. After the disc has rolled distance $r\theta$ then the point $(0, 0)$ has moved on to

$$(x(\theta), y(\theta)) = (r(\theta - \sin\theta), r(1 - \cos\theta)).$$

Thus $(x')^2 + (y')^2 = r^2 \left[(1 - \cos\theta)^2 + \sin^2\theta \right] = 2r^2(1 - \cos\theta)$ and so

$$s \;=\; \sqrt{2}\, r \int_0^{2\pi} \sqrt{1 - \cos\theta}\, \mathrm{d}\theta \;=\; 2r \int_0^{2\pi} \left| \sin \frac{1}{2}\theta \right| \mathrm{d}\theta \;=\; 8r.$$

■

Example 0.7 *Show that the curvature of a curve is identically zero if and only if the curve is part of a line.*

Solution. For a curve that is part of a line, $\mathbf{t}$ is constant and so $\kappa = |\mathrm{d}\mathbf{t}/\mathrm{d}s| = 0$. Conversely if κ is identically zero, then $\ddot{\gamma}(s) = \mathbf{0}$ and hence $\gamma(s) = \mathbf{a}s + \mathbf{b}$ for constant vectors $\mathbf{a}, \mathbf{b}$. This is the parameterization of a line. ■

Example 0.8 *(a) Show that a circle of radius a has constant curvature $\kappa = a^{-1}$.*

(b) Conversely let γ be a curve in the xy-plane which has constant positive curvature κ. Show that γ is part of a circle. (There are non-planar curves with constant curvature, such as helices.)

Proof. (a) Without loss of generality we can take the circle's centre to be the origin in the xy-plane. A parameterization by arc length is

$$\gamma(s) = \left(a\cos\left(\frac{s}{a}\right), a\sin\left(\frac{s}{a}\right)\right).$$

Then

$$\kappa(s) = \ddot{\gamma}(s) = \left|\left(-\frac{1}{a}\cos\left(\frac{s}{a}\right), -\frac{1}{a}\sin\left(\frac{s}{a}\right)\right)\right| = \frac{1}{a}.$$

(b) Asumme now that the curvature κ is constant. We can write

$$\frac{\mathrm{d}\mathbf{t}}{\mathrm{d}s} = \kappa\mathbf{n},$$

where $\mathbf{n}$ is a unit vector in the same plane. As $\mathbf{t}$ is a unit vector, then $\mathbf{t}$ and $\mathbf{n}$ are perpendicular. As $\mathbf{n}$ is a unit vector then $\mathrm{d}\mathbf{n}/\mathrm{d}s$ is perpendicular to $\mathbf{n}$ and so parallel to $\mathbf{t}$. Further, differentiating $\mathbf{t}\cdot\mathbf{n} = 0$ gives

$$0 = \frac{\mathrm{d}\mathbf{t}}{\mathrm{d}s}\cdot\mathbf{n} + \frac{\mathrm{d}\mathbf{n}}{\mathrm{d}s}\cdot\mathbf{t} = \boldsymbol{\kappa} + \frac{\mathrm{d}\mathbf{n}}{\mathrm{d}s}\cdot\mathbf{t},$$

showing $\mathrm{d}\mathbf{n}/\mathrm{d}s = -\kappa\mathbf{t}$.

Now consider the vector

$$\mathbf{c} = \gamma + \frac{1}{\kappa}\mathbf{n}.$$

Note

$$\frac{\mathrm{d}\mathbf{c}}{\mathrm{d}s} = \mathbf{t} + \frac{1}{\kappa}(-\kappa\mathbf{t}) = \mathbf{0}.$$

So $\mathbf{c}$ is constant and $|\gamma - \mathbf{c}| = 1/\kappa$, showing γ is a circular arc, with centre $\mathbf{c}$ and radius κ^{-1}. ■

Example 0.9 *Let γ be a smooth curve in $\mathbb{R}^3$ parameterized by t, which need not be arc length. Show that*

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

Solution. The is left as Exercise 1 on Sheet 0. ■

0.2 Surface Area

This is largely material from Prelims *Geometry.*

Let $\mathbf{r}\colon U \to \mathbb{R}^3$ be a smooth parameterized surface with

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

and consider the small rectangle of the plane that is bounded by the co-ordinate lines $u = u_0$ and $u = u_0 + \delta u$ and $v = v_0$ and $v = v_0 + \delta v$. Then $\mathbf{r}$ maps this to a small region of the surface $\mathbf{r}(U)$ and we are interested in calculating the surface area of this small region, which is approximately that of a parallelogram. Note

$$\begin{aligned}
\mathbf{r}(u+\delta u, v) - \mathbf{r}(u,v) &\approx \frac{\partial\mathbf{r}}{\partial u}(u,v)\ \delta u,\\
\mathbf{r}(u, v+\delta v) - \mathbf{r}(u,v) &\approx \frac{\partial\mathbf{r}}{\partial v}(u,v)\ \delta v.
\end{aligned}$$

Recall that the area of a parallelogram with sides $\mathbf{a}$ and $\mathbf{b}$ is $|\mathbf{a} \wedge \mathbf{b}|$ where $\wedge$ denotes the vector product. So the element of surface area we are considering is approximately

$$\left| \frac{\partial \mathbf{r}}{\partial u}\,\delta u \wedge \frac{\partial \mathbf{r}}{\partial v}\,\delta v \right| = \left| \frac{\partial \mathbf{r}}{\partial u} \wedge \frac{\partial \mathbf{r}}{\partial v} \right| \delta u\,\delta v.$$

This motivates the following definitions.

Definition 0.10 *Let $\mathbf{r}\colon U \to \mathbb{R}^3$ be a smooth parameterized surface. Then the **surface area** (or simply **area**) of $\mathbf{r}(U)$ is defined to be*

$$\iint_U \left| \frac{\partial \mathbf{r}}{\partial u} \wedge \frac{\partial \mathbf{r}}{\partial v} \right| \mathrm{d}u\,\mathrm{d}v.$$

We will often write

$$\mathrm{d}S = \left| \frac{\partial \mathbf{r}}{\partial u} \wedge \frac{\partial \mathbf{r}}{\partial v} \right| \mathrm{d}u\,\mathrm{d}v$$

to denote an infinitesimal part of surface area.

Proposition 0.11 *The surface area of $\mathbf{r}(U)$ is independent of the choice of parameterization.*

Proof. Let $\Sigma = \mathbf{r}(U) = \mathbf{s}(W)$ be two different parameterizations of a surface X; take u, v as the co-ordinates on U and p, q as the co-ordinates on W. Let $f = (f_1, f_2) : U \to W$ be the co-ordinate change map; that is for any $(u, v) \in U$ we have

$$\mathbf{r}(u, v) = \mathbf{s}(f(u, v)) = \mathbf{s}(f_1(u, v), f_2(u, v)).$$

By the chain rule

$$\frac{\partial \mathbf{r}}{\partial u} = \frac{\partial \mathbf{s}}{\partial p}\frac{\partial f_1}{\partial u} + \frac{\partial \mathbf{s}}{\partial q}\frac{\partial f_2}{\partial u}, \qquad \frac{\partial \mathbf{r}}{\partial v} = \frac{\partial \mathbf{s}}{\partial p}\frac{\partial f_1}{\partial v} + \frac{\partial \mathbf{s}}{\partial q}\frac{\partial f_2}{\partial v},$$

giving

$$\begin{aligned} \frac{\partial \mathbf{r}}{\partial u} \wedge \frac{\partial \mathbf{r}}{\partial v} &= \frac{\partial \mathbf{s}}{\partial p}\frac{\partial f_1}{\partial u} \wedge \frac{\partial \mathbf{s}}{\partial q}\frac{\partial f_2}{\partial v} + \frac{\partial \mathbf{s}}{\partial q}\frac{\partial f_2}{\partial u} \wedge \frac{\partial \mathbf{s}}{\partial p}\frac{\partial f_1}{\partial v} \\ &= \left(\frac{\partial f_1}{\partial u}\frac{\partial f_2}{\partial v} - \frac{\partial f_1}{\partial v}\frac{\partial f_2}{\partial u} \right) \frac{\partial \mathbf{s}}{\partial p} \wedge \frac{\partial \mathbf{s}}{\partial q} \\ &= \frac{\partial (p, q)}{\partial (u, v)}\frac{\partial \mathbf{s}}{\partial p} \wedge \frac{\partial \mathbf{s}}{\partial q}. \end{aligned}$$

Finally

$$\begin{aligned} \iint_U \left| \frac{\partial \mathbf{r}}{\partial u} \wedge \frac{\partial \mathbf{r}}{\partial v} \right| \mathrm{d}u\,\mathrm{d}v &= \iint_U \left| \frac{\partial (p, q)}{\partial (u, v)}\frac{\partial \mathbf{s}}{\partial p} \wedge \frac{\partial \mathbf{s}}{\partial q} \right| \mathrm{d}u\,\mathrm{d}v \\ &= \iint_U \left| \frac{\partial \mathbf{s}}{\partial p} \wedge \frac{\partial \mathbf{s}}{\partial q} \right| \left| \frac{\partial (p, q)}{\partial (u, v)} \right| \mathrm{d}u\,\mathrm{d}v \\ &= \iint_W \left| \frac{\partial \mathbf{s}}{\partial p} \wedge \frac{\partial \mathbf{s}}{\partial q} \right| \mathrm{d}p\,\mathrm{d}q \end{aligned}$$

by the two-dimensional substitution rule (Apostol, *Mathematical Analysis* p.421). ■

Example 0.12 *Find the surface area of the cone*

$$x^2 + y^2 = z^2 \cot^2 \alpha \qquad 0 \leqslant z \leqslant h.$$

Solution. We can parameterize the cone as

$$\mathbf{r}(z, \theta) = (z \cot \alpha \cos \theta, z \cot \alpha \sin \theta, z), \qquad 0 < \theta < 2\pi,\ 0 < z < h.$$

We have

$$\mathbf{r}_z = (\cot \alpha \cos \theta, \cot \alpha \sin \theta, 1), \qquad \mathbf{r}_\theta = (-z \cot \alpha \sin \theta, z \cot \alpha \cos \theta, 0),$$

giving

$$\mathbf{r}_z \wedge \mathbf{r}_\theta = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ \cot \alpha \cos \theta & \cot \alpha \sin \theta & 1 \\ -z \cot \alpha \sin \theta & z \cot \alpha \cos \theta & 0 \end{vmatrix} = \begin{pmatrix} -z \cot \alpha \cos \theta \\ -z \cot \alpha \sin \theta \\ z \cot^2 \alpha \end{pmatrix}.$$

Thus the cone has surface area

$$\begin{aligned} & \int_{\theta=0}^{2\pi} \int_{z=0}^{h} \sqrt{z^2 \cot^2 \alpha \cos^2 \theta + z^2 \cot^2 \alpha \sin^2 \theta + z^2 \cot^4 \alpha}\, \mathrm{d}z\, \mathrm{d}\theta \\ = & \int_{\theta=0}^{2\pi} \int_{z=0}^{h} z \cot \alpha \sqrt{1 + \cot^2 \alpha}\, \mathrm{d}z\, \mathrm{d}\theta \\ = & \ 2\pi \int_{z=0}^{h} z \cot \alpha \csc \alpha\, \mathrm{d}z \\ = & \ 2\pi \times \frac{\cos \alpha}{\sin^2 \alpha} \times \left[\frac{z^2}{2} \right]_0^h \\ = & \ \frac{\pi h^2 \cos \alpha}{\sin^2 \alpha}. \end{aligned}$$

Note that as $\alpha \to 0$ this area tends to infinity as the cone transforms into the plane and the area tends to zero as $\alpha \to \pi/2$. ■

Proposition 0.13 *(**Surface area of a graph**) Let $z = f(x, y)$ denote the graph of a function f defined on a subset S of the xy-plane. Show that the graph has surface area*

$$\iint_S \sqrt{1 + (f_x)^2 + (f_y)^2}\, \mathrm{d}x\, \mathrm{d}y.$$

Proof. We can parameterize the surface as

$$\mathbf{r}(x, y) = (x, y, f(x, y)) \qquad (x, y) \in S.$$

Then

$$\mathbf{r}_x \wedge \mathbf{r}_y = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ 1 & 0 & f_x \\ 0 & 1 & f_y \end{vmatrix} = (-f_x, -f_y, 1).$$

Hence the graph has surface area

$$\iint_S |\mathbf{r}_x \wedge \mathbf{r}_y| \, \mathrm{d}x \, \mathrm{d}y = \iint_S \sqrt{1 + (f_x)^2 + (f_y)^2} \, \mathrm{d}x \, \mathrm{d}y.$$

■

Example 0.14 *Find the area of the paraboloid $z = x^2 + y^2$ that lies below the plane $z = 4$.*

Solution. By Proposition 0.13 the desired area equals

$$A = \iint_R \sqrt{1 + (2x)^2 + (2y)^2} \, \mathrm{d}A$$

where R is the disc $x^2 + y^2 \leqslant 4$ in the xy-plane. We can parameterize R using polar co-ordinates

$$x = r\cos\theta, \qquad y = r\sin\theta, \qquad 0 < r < 2, \quad 0 < \theta < 2\pi,$$

and then we have that

$$\begin{aligned} A &= \int_{\theta=0}^{2\pi} \int_{r=0}^{2} \sqrt{1 + (2r\cos\theta)^2 + (2r\sin\theta)^2} \, r \, \mathrm{d}r \, \mathrm{d}\theta \\ &= \int_{\theta=0}^{2\pi} \int_{r=0}^{2} \sqrt{1 + 4r^2} \, r \, \mathrm{d}r \, \mathrm{d}\theta \\ &= 2\pi \int_{r=0}^{2} \sqrt{1 + 4r^2} \, r \, \mathrm{d}r \\ &= 2\pi \times \frac{1}{8} \times \frac{2}{3} \times \left[\left(1 + 4r^2\right)^{3/2} \right]_{r=0}^{2} \\ &= \frac{\pi}{6} \left[17^{3/2} - 1 \right]. \end{aligned}$$

■

Proposition 0.15 ***(Surfaces of revolution)*** *A surface S is formed by rotating the graph of*

$$y = f(x) \qquad a < x < b,$$

about the x-axis. Here $f(x) > 0$ for all x. The surface area of S equals

$$\mathit{Area}(S) = 2\pi \int_{x=a}^{x=b} f(x) \frac{\mathrm{d}s}{\mathrm{d}x} \, \mathrm{d}x.$$

Proof. Using the parameterization

$$\mathbf{r}(x, \theta) = (x, f(x)\cos\theta, f(x)\sin\theta) \qquad -\pi < \theta < \pi, \; a < x < b$$

we have

$$\mathbf{r}_x \wedge \mathbf{r}_\theta = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ 1 & f'(x)\cos\theta & f'(x)\sin\theta \\ 0 & -f(x)\sin\theta & f(x)\cos\theta \end{vmatrix} = \begin{pmatrix} f'(x)f(x) \\ -f(x)\cos\theta \\ -f(x)\sin\theta \end{pmatrix}.$$

So

$$|\mathbf{r}_x \wedge \mathbf{r}_\theta|^2 = f(x)^2 f'(x)^2 + f(x)^2 = f(x)^2(1 + f'(x)^2) = f(x)^2 \left(\frac{\mathrm{d}s}{\mathrm{d}x}\right)^2.$$

The result follows. ■

Example 0.16 *Rederive the area of the paraboloid $z = x^2 + y^2$ that lies below the plane $z = 4$, by thinking of the paraboloid as a surface of revolution.*

Solution. We can consider the paraboloid as a rotation of the curve $x = \sqrt{z}$ about the z-axis where $0 < z < 4$. We then have

$$\left(\frac{\mathrm{d}s}{\mathrm{d}z}\right)^2 = 1 + \left(\frac{\mathrm{d}x}{\mathrm{d}z}\right)^2 = 1 + \left(\frac{1}{2\sqrt{z}}\right)^2 = 1 + \frac{1}{4z}.$$

Hence

$$\begin{aligned} A &= 2\pi \int_{z=0}^{4} x \frac{\mathrm{d}s}{\mathrm{d}z}\,\mathrm{d}z \\ &= 2\pi \int_{z=0}^{4} \sqrt{z}\sqrt{1 + \frac{1}{4z}}\,\mathrm{d}z \\ &= 2\pi \int_{z=0}^{4} \sqrt{z + \frac{1}{4}}\,\mathrm{d}z \\ &= 2\pi \left[\frac{2}{3}\left(z + \frac{1}{4}\right)^{3/2}\right]_0^4 \\ &= \frac{4\pi}{3}\left[\left(\frac{17}{4}\right)^{3/2} - \left(\frac{1}{4}\right)^{3/2}\right] \\ &= \frac{\pi}{6}\left[17^{3/2} - 1\right]. \end{aligned}$$

■

Proposition 0.17 *Isometries preserve area.*

Proof. An isometry is a bijection between surfaces which preserves the lengths of curves. Say that $\mathbf{r}\colon U \to \mathbb{R}^3$ is a parameterization of a smooth surface $X = \mathbf{r}(U)$ and $f\colon \mathbf{r}(U) \to Y$ is an isometry from X to another smooth surface Y. Then the map

$$\mathbf{s} = f \circ \mathbf{r} : U \to Y$$

is a parameterization of Y also using co-ordinates from U.

Consider a curve

$$\gamma(t) = \mathbf{r}(u(t), v(t)) \qquad a \leqslant t \leqslant b$$

in X. By the chain rule

$$\gamma' = u'\mathbf{r}_u + v'\mathbf{r}_v$$

and

$$|\gamma'|^2 = E(u')^2 + 2Fu'v' + G(v')^2$$

where

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

The length of γ equals is

$$\mathcal{L}(\gamma) = \int_{t=a}^{t=b} |\gamma'(t)| \, \mathrm{d}t = \int_{t=a}^{t=b} \sqrt{E(u')^2 + 2Fu'v' + G(v')^2} \, \mathrm{d}t.$$

In a similar fashion the length of the curve $f(\gamma)$ equals

$$\mathcal{L}(f(\gamma)) = \int_{t=a}^{t=b} \sqrt{\tilde{E}(u')^2 + 2\tilde{F}u'v' + \tilde{G}(v')^2} \, \mathrm{d}t$$

where

$$\tilde{E} = \mathbf{s}_u \cdot \mathbf{s}_u, \qquad \tilde{F} = \mathbf{s}_u \cdot \mathbf{s}_v, \qquad \tilde{G} = \mathbf{s}_v \cdot \mathbf{s}_v.$$

As f is an isometry then

$$\int_{t=a}^{t=b} \sqrt{E(u')^2 + 2Fu'v' + G(v')^2} \, \mathrm{d}t = \int_{t=a}^{t=b} \sqrt{\tilde{E}(u')^2 + 2\tilde{F}u'v' + \tilde{G}(v')^2} \, \mathrm{d}t.$$

This is true for all b, so it must follow that

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

for all values of t and all functions u, v. By choosing $u = t, v = 0$, we find $E = \tilde{E}$ and we also obtain $G = \tilde{G}$ by setting $u = 0, v = t$. It follows then that $F = \tilde{F}$ as well.

Now the area of a subset $\mathbf{r}(V)$ of X is given by

$$\iint_V |\mathbf{r}_u \wedge \mathbf{r}_v| \, \mathrm{d}u \, \mathrm{d}v.$$

However, by the quadruple scalar product

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

As

$$|\mathbf{s}_u \wedge \mathbf{s}_v| = \sqrt{\tilde{E}\tilde{G} - \tilde{F}^2} = \sqrt{EG - F^2} = |\mathbf{r}_u \wedge \mathbf{r}_v|$$

then the area of $f(\mathbf{r}(V))$ equals

$$\iint_V |\mathbf{s}_u \wedge \mathbf{s}_v| \, \mathrm{d}u \, \mathrm{d}v = \iint_V |\mathbf{r}_u \wedge \mathbf{r}_v| \, \mathrm{d}u \, \mathrm{d}v$$

and we see that isometries preserve areas. ■

Remark 0.18 *As angles between curves can similarly be written in terms of E, F, G, then isometries also preserve angles.*

0.3 The Real Projective Plane

This is material from Part A *Projective Geometry.*

The following theorem, *Bézout's theorem,* is not actually part of this course, but it is a clean and general result which readily motivates the worth of projective geometry. For those interested, the theorem is part of B3.3 *Algebraic Curves.*

Bézout's theorem is a first significant result in *algebraic geometry*, which is unsurprisingly interested in geometric objects that can be described using the language of algebra, and proved using the theorems of algebra.

So, for example, curves defined by polynomials such as $x^2 + xy + y^2 = 1$ are of interest to an algebraic geometer whereas the curve with equation $y = e^x$ would not be. Bézout's theorem addresses a natural first question: how many times do two curves, defined by polynomials of degrees m and n, intersect?

If we begin with $m = n = 1$ then we are talking about two lines. These typically meet in a point but we recognize that this wouldn't be the case if the lines are parallel. If $m = 1$ and $n = 2$, so that we're considering a line and, say, a parabola, then there can be as many as two intersections. We appreciate that there may be no intersections – with $y = 0$ and $y = x^2 + 1$ – but that can be circumvented by working with complex numbers, and we can see that the answer might be just one – with $y = 0$ and $y = x^2$ – but we could think of this as a double contact or repeated root in some sense. But we are still left with cases like $y = x$ and $(y - x)^2 = 1$ which appear to have no intersection, or $y = 0$ and $y^2 = x$ which has one 'single contact' intersection. Think about the $m = n = 2$ case and you'll find the number of intersections can be $0, 1, 2, 3, 4$.

Perhaps, then, the best we can do is to say that the two curves meet in at most mn points. Even the use of complex numbers and appreciation of multiple contacts cannot completely resolve the issue. It turns out, though, that all we are missing is the notion of *points at infinity.* Once we properly introduce the notion of parallel lines meeting at a point at infinity then Bézout's theorem states that the two curves have mn intersections, counting multiple contacts, using complex numbers, and including points at infinity.

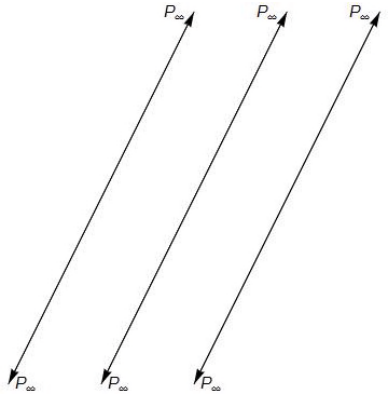


Figure 0.3 – parallel lines meeting at infinity

So given two parallel lines, we will agree that they meet at some idealized point at infinity (Figure 0.3). As lines should only meet once, this point at infinity lies in both directions. Given a third parallel line, it will meet each of these two lines in a point at infinity, and so in fact at the same point at infinity. So to each family of parallel lines there is a single point at infinity.

Put another way there is a point at infinity for each gradient m, that is the lines $y = mx + c$ all meet in the same point at infinity. And we need to remember to allow $m = \infty$ as a possible gradient, relating to the family of parallel vertical lines. These points at infinity make the *line at infinity.*

Note though that these 'points at infinity' aren't special in any way, or rather we've only made them special by our choice of where to put our affine xy-axes. The family of parallel lines passing through a point at infinity, properly judged from infinity, would look the same as the family of lines passing through the origin.

If we return to our earlier examples when Bézout's theorem appeared not to hold:

- $y = 0$, $y^2 = x$. The parabola and line meet a second time at the point at infinity at the 'end' of the x-axis.
- $y = x$, $(y - x)^2 = 1$. The two lines $y = x \pm 1$ both meet $y = x$ at a point at infinity in the same way that $y = 0$ and $y^2 = x^2$ meet at the origin.

We need, then, a rigorous, formal way of introducing these points at infinity if we are to prove geometric results involving them. For fixed m the lines $y = mx + c$ all meet at a point at infinity. This point at infinity is where the points (x, mx) move to as $x \to \pm\infty$. So it's the ratio of x and y that is important here. Somehow we want to include all the points (x, y) of the standard affine plane $\mathbb{R}^2$ and a line at infinity including the points $(\infty, m\infty)$ where $m \in \mathbb{R} \cup \{\infty\}$.

We cannot make easy meaning of $(\infty, m\infty)$ but if we recognize this ∞ as the consequence of some erroneous division by zero, then we can describe our 'extended' plane with the introduction of **homogeneous co-ordinates.**

Definition 0.19 *Given real x_0, x_1, x_2, not all zero, then we write $[x_0 \colon x_1 \colon x_2]$ for the equivalence class of $(x_0, x_1, x_2) \in \mathbb{R}^3 \backslash \{\mathbf{0}\}$ under the equivalence relation*

$$(x_0, x_1, x_2) \sim (\lambda x_0, \lambda x_1, \lambda x_2) \qquad \text{where } \lambda \neq 0.$$

How does this help us with the previous discussion? Well if $x_0 \neq 0$ then we may divide by x_0 (i.e. set $\lambda = 1/x_0$) to see that such equivalence classes can be represented as $[1 \colon x \colon y]$ where $x = x_1/x_0$ and $y = x_2/x_0$. These are 'most' of the equivalence classes and $[1 \colon x \colon y]$ can be identified with the point $(x, y) \in \mathbb{R}^2$. And the remaining equivalence classes, when $x_0 = 0$ are $[0 \colon 1 \colon m]$ when $x_1 \neq 0$ which corresponds to the point at infinity $(\infty, m\infty)$, and finally $[0 \colon 0 \colon 1]$ which corresponds to '$m = \infty$' the point at infinity of the vertical lines.

Whilst here, and remembering that $x = x_1/x_0$ and $y = x_2/x_0$, we can see that the affine lines $y = mx + c$ would become

$$x_2 = mx_1 + cx_0$$

and that each passes through the point at infinity $[0 \colon 1 \colon m]$. Further the parabola $y^2 = x$ would become $x_2^2 = x_0 x_1$. The variables x_1/x_0 and x_2/x_0 are known as **inhomogeneous co-ordinates.**

So the earlier 'problematic' examples we see now that

- $y = 0$, $y^2 = x$ homogeneously become $x_2 = 0$ and $x_2^2 = x_0 x_1$ so each passes through the point at infinity $[0 \colon 1 \colon 0]$.

- $y = x$, $(y-x)^2 = 1$ homogeneously become $x_2 = x_1$ and $(x_2 - x_1)^2 = x_0^2$ so each passes includes the points at infinity at $[0\colon 1\colon 1]$. Indeed these two curves in a like manner to how $y = 0$ meets with $y^2 = x^2$ at the origin.

0.4 Holomorphic branches

This is material from A2 *Metric Spaces and Complex analysis* we know:

Proposition 0.20 *Let $z \in \mathbb{C}\backslash(-\infty, 0]$.*

(a) Then z can be written as $z = re^{i\theta}$ where $r > 0, \theta \in (-\pi, \pi)$ in a unique fashion.

(b) The function

$$\sqrt{z} = \sqrt{r}e^{i\theta/2}$$

is a holomorphic function on the cut plane $\mathbb{C}\backslash(-\infty, 0]$ with a sign discontinuity over the cut.

Remark 0.21 *If we were to take points z_+ and z_-, respectively just above and below the cut $(-\infty, 0]$ then we would have*

$$z_+ = re^{i\theta_+} \quad \text{where } \theta_+ \approx \pi; \qquad z_- = re^{i\theta_-} \quad \text{where } \theta_- \approx -\pi.$$

So with $\sqrt{z}$ as defined above we see

$$\sqrt{z_+} \approx \sqrt{r}e^{i\pi/2} = i\sqrt{r}; \qquad \sqrt{z_-} \approx \sqrt{r}e^{-i\pi/2} = -i\sqrt{r}.$$

We see this time that there is a sign change as we cross the cut.

The only other holomorphic function on $\mathbb{C}\backslash(-\infty, 0]$ which satisfies $w^2 = z$ is $w = -\sqrt{z}$ and these two functions, $\sqrt{z}$ and $-\sqrt{z}$ are the two holomorphic branches of $\sqrt{z}$ on this cut plane. We see that as we cross the cut we move from one branch's values to the other's values.

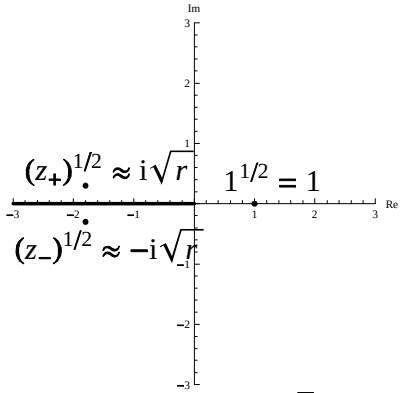


Figure 0.4a: $\sqrt{z}$

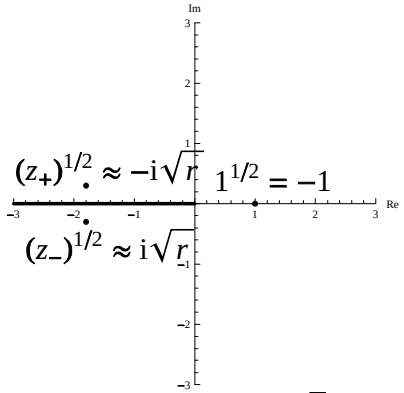


Figure 04.b: $-\sqrt{z}$

Example 0.22 *For z in the cut plane $\mathbb{C}\backslash(-\infty, 1]$ we will let*

$$\theta_1 \quad \textit{denote the value of } \arg(z+1) \quad \textit{in the range} \quad (-\pi, \pi),$$
$$\theta_2 \quad \textit{denote the value of } \arg(z-1) \quad \textit{in the range} \quad (-\pi, \pi),$$

as in the diagram below.

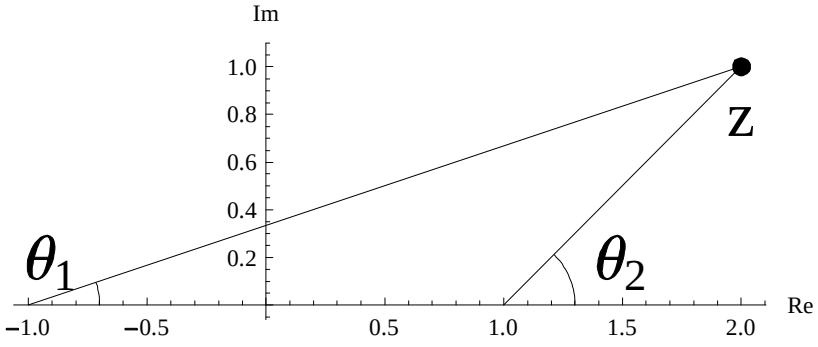


Figure 0.5

So we have

$$(z+1)(z-1) = |z+1|\, e^{i\theta_1}\, |z-1|\, e^{i\theta_2}$$

and

$$w = \sqrt{|z+1|\,|z-1|}e^{i(\theta_1+\theta_2)/2}$$

is a holomorphic function on $\mathbb{C}\backslash(-\infty, 1]$ *which satisfies*

$$w^2 = z^2 - 1.$$

What about the continuity, or otherwise, of w *over the cut? Firstly let* r *be a real number in the range* $-1 < r < 1$ *and let* r_+ *and* r_- *be complex numbers just above and just below* r *in the complex plane. Then*

$$\begin{aligned} &\textit{for } r_+ \textit{ we have } \theta_1 \approx 0 \textit{ and } \theta_2 \approx \pi; \\ &\textit{for } r_+ \textit{ we have } \theta_1 \approx 0 \textit{ and } \theta_2 \approx -\pi. \end{aligned}$$

So

$$\begin{aligned} w_+ &\approx \sqrt{1-r^2}e^{i(0+\pi)/2} = i\sqrt{1-r^2}; \\ w_- &\approx \sqrt{1-r^2}e^{i(0-\pi)/2} = -i\sqrt{1-r^2}. \end{aligned}$$

So we see that we have a sign discontinuity across $(-1, 1)$.

However if we take r *be a real number in the range* $r < -1$ *and let* r_+ *and* r_- *be complex numbers just above and just below* r *in the complex plane. Then*

$$\begin{aligned} &\textit{for } r_+ \textit{ we have } \theta_1 \approx \pi \textit{ and } \theta_2 \approx \pi; \\ &\textit{for } r_+ \textit{ we have } \theta_1 \approx -\pi \textit{ and } \theta_2 \approx -\pi. \end{aligned}$$

So

$$\begin{aligned} w_+ &\approx \sqrt{r^2-1}e^{i(\pi+\pi)/2} = -\sqrt{r^2-1}; \\ w_- &\approx \sqrt{r^2-1}e^{i(-\pi-\pi)/2} = -\sqrt{r^2-1}. \end{aligned}$$

We see that w *is actually continuous across* $(-\infty, -1)$ *and we can in fact extend* w *to a holomorphic function on all of* $\mathbb{C}\backslash[-1, 1]$.

Note the behaviour of w *near the points* -1 *and* 1. *If* $z \approx -1$ *then* $w \approx \sqrt{2}i\sqrt{z+1}$ *where* $\sqrt{z+1}$ *is a standard branch of* $\sqrt{z+1}$ *on the cut plane* $\mathbb{C}\backslash[-1, \infty)$. *If* $z \approx 1$ *then* $w \approx \sqrt{2}\sqrt{z-1}$ *where* $\sqrt{z-1}$ *is a standard branch of* $\sqrt{z-1}$ *on the cut plane* $\mathbb{C}\backslash(-\infty, 1]$.

Remark 0.23 *To properly consider the* ***multifunction*** $\sqrt{z^2-1}$ *(or any similar multi-valued function) it helps to consider its* ***Riemann surface****. In this case the (affine) Riemann surface is the set of points*

$$\Sigma = \left\{(w,z) \in \mathbb{C}^2 : w^2 = z^2 - 1\right\}.$$

Firstly consider the situation in $\mathbb{R}^2$*. The curve* $y^2 = x^2 - 1$ *is a hyperbola. Above* $(1,\infty)$ *and* $(-\infty,-1)$ *sit branches* $y = \pm\sqrt{x^2-1}$ *and these two branches meet at* $(\pm 1, 0)$*. So most of the curve is in one or other of the sets*

$$C_+ = \left\{\left(x, \sqrt{x^2-1}\right) \mid |x| > 1\right\}; \qquad C_- = \left\{\left(x, -\sqrt{x^2-1}\right) \mid |x| > 1\right\}.$$

In fact $C_+ \cup C_-$ *excludes only the branch points* $(\pm 1, 0)$ *and we also see that as we cross the branch points we move from* C_+ *to* C_- *(or vice versa).*

In the complex case, for $z \notin [-1,1]$ *there are two values* $\pm w$*. For* $z = \pm 1$ *the only value of* w *is* 0*. The points* (z,w) *and* $(z,-w)$ *have already been described as two different branches of* $\sqrt{z^2-1}$ *but we need to take some care to see how these branches fit together as subset of* Σ*. If we set as above*

$$\Sigma_+ = \{(z,w) \mid z \notin [-1,1]\} \qquad \text{and} \qquad \Sigma_- = \{(z,-w) \mid z \notin [-1,1]\}.$$

Then $\Sigma_+ \cup \Sigma_-$ *is most of* Σ *missing only those points associated with* $z \in [-1,1]$*. We can note, as with previous branches, that as* z *crosses the cut* $[-1,1]$ *then* (z,w) *moves continuously to the other branch* Σ_- *and likewise* $(z,-w)$ *moves continuously to the other branch* Σ_+*.*

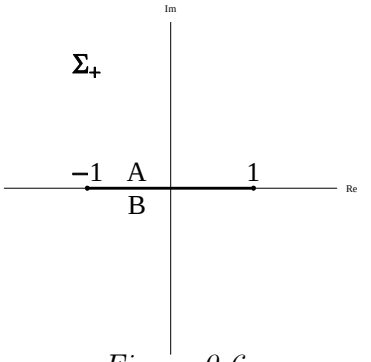


Figure 0.6a

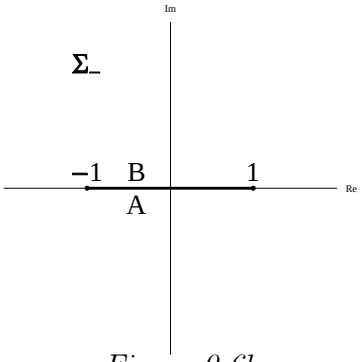


Figure 0.6b

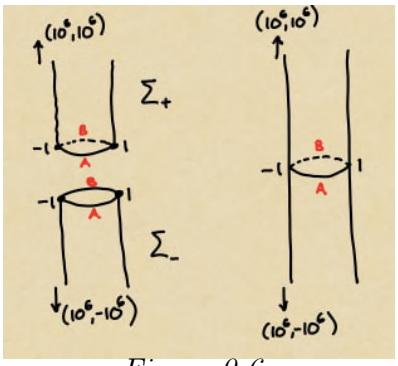


Figure 0.6c

So Σ_+ *and* Σ_- *fit together on* Σ *by gluing either side of* $[-1,1]$ *as labelled in Figures 0.6a/b. We can then see that topologically* Σ *is a cylinder in* $\mathbb{C}^2$ *(Figure 0.6c).*

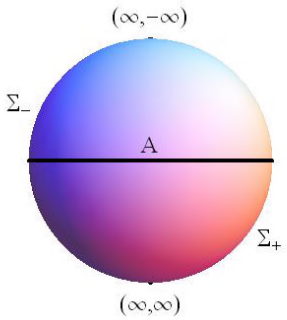


Figure 0.7

Whilst there is only one ∞ in the extended complex plane, note that, as z becomes large, then (z, w) and $(z, -w)$ are diverging points in $\mathbb{C}^2$. So we should introduce two points at infinity to Σ which at either end of the cylinder to reflect this behaviour. Topologically, with these points included, Σ is a sphere (in what is called complex projective space).

More rigorously, considering instead Σ as a subset of the complex projective plane $\mathbb{CP}^2$, the projectivized version of Σ is

$$\Sigma = \left\{ [z_0 \colon z_1 \colon z_2] \in \mathbb{CP}^2 : z_1^2 = z_2^2 - z_0^2 \right\}.$$

The line at infinity has equation $z_0 = 0$ and so the two points at infinity are $[0 \colon 1 \colon 1]$ and $[0 \colon 1 \colon -1]$.

0.5 Differentiability in $\mathbb{R}^n$

This is material from Part A *Multidimenstional Analysis and Geometry.*

Definition 0.24 *Let $f \colon \mathbb{R}^n \to \mathbb{R}^m$ be a smooth map, (i.e. all partial derivatives of f of all orders exist everywhere.) Let $\mathbf{p}, \mathbf{v} \in \mathbb{R}^n$ and let $\gamma \colon (-\varepsilon, \varepsilon) \to \mathbb{R}^n$ be a smooth curve in $\mathbb{R}^n$ such that*

$$\gamma(0) = \mathbf{p} \quad \textit{and} \quad \gamma'(0) = \mathbf{v}.$$

*Then $f \circ \gamma$ is a smooth curve in $\mathbb{R}^m$. The **differential** of f at p is the linear map $\mathrm{d}f_p \colon \mathbb{R}^n \to \mathbb{R}^m$ defined by*

$$\mathrm{d}f_p(\mathbf{v}) = \mathrm{d}f_p(\gamma'(0)) = (f \circ \gamma)'(0).$$

Proposition 0.25 *$\mathrm{d}f_p(\mathbf{v})$ is independent of the choice of curve γ.*

Proof. For ease of notation we shall consider the case when $m = n = 2$. Write $f = \binom{f_1}{f_2}$ and $\gamma = \binom{\gamma_1}{\gamma_2}$. Then

$$\begin{aligned}
(f \circ \gamma)'(0) &= \begin{pmatrix} (f_1 \circ \gamma)'(0) \\ (f_2 \circ \gamma)'(0) \end{pmatrix} \\
&= \begin{pmatrix} \frac{\partial f_1}{\partial x}\gamma_1'(0) + \frac{\partial f_1}{\partial y}\gamma_2'(0) \\ \frac{\partial f_2}{\partial x}\gamma_1'(0) + \frac{\partial f_2}{\partial y}\gamma_2'(0) \end{pmatrix} \\
&= \begin{pmatrix} \frac{\partial f_1}{\partial x} & \frac{\partial f_1}{\partial y} \\ \frac{\partial f_2}{\partial x} & \frac{\partial f_2}{\partial y} \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix}.
\end{aligned}$$

As the partial derivatives in the above matrix depend only on the function f and the point p then $\mathrm{d}f_p$ (which we see has the Jacobian as its matrix) is independent of the choice of γ. ■

For those meeting multivariable differentials for the first time, this definition contrasts markedly with the usual notion of a differential $\mathrm{d}f/\mathrm{d}x$. Clearly when $m = n = 1$ then the two definitions agree, but the general differential cannot simply be visualized as a gradient. Rather $\mathrm{d}f_p$ is a first, linear approximation of the function f at p. Here are two examples to help motivate this appreciation.

Example 0.26 *By Taylor's theorem for a smooth function $f = (u, v) : \mathbb{R}^2 \to \mathbb{R}^2$ we have*

$$\begin{aligned} f\begin{pmatrix} x+h \\ y+k \end{pmatrix} &= \begin{pmatrix} u(x+h, y+k) \\ v(x+y, y+k) \end{pmatrix} \\ &= \begin{pmatrix} u(x,y) + hu_x(x,y) + ku_y(x,y) + \cdots \\ v(x,y) + hv_x(x,y) + kv_y(x,y) + \cdots \end{pmatrix} \\ &= \begin{pmatrix} u(x,y) \\ v(x,y) \end{pmatrix} + \begin{pmatrix} u_x & u_y \\ v_x & v_y \end{pmatrix} \begin{pmatrix} h \\ k \end{pmatrix} + O\left(|(h,k)|^2\right). \end{aligned}$$

This result generalizes naturally to the general m, n case.

Example 0.27 *For a holomorphic function $f\colon \mathbb{C} \to \mathbb{C}$ and $p \in \mathbb{C}$, then*

$$\mathrm{d}f_p = \begin{pmatrix} u_x & u_y \\ v_x & v_y \end{pmatrix},$$

where $u = \operatorname{Re} f$ and $v = \operatorname{Im} f$. By the Cauchy-Riemann equations

$$\mathrm{d}f_p = \begin{pmatrix} u_x & -v_x \\ v_x & u_x \end{pmatrix} = \lambda \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix},$$

where $\lambda = \sqrt{u_x^2 + v_x^2} = |f'(p)|$ and $\theta = \arg z$. This shows that when $f'(p) \neq 0$, then f is approximately enlarging by $|f'(p)|$ and rotating by $\arg z$.

This can be more easily seen using Taylor's theorem for a holomorphic function in one complex variable. We then have

$$f(p+h) = f(p) + f'(p)h + O\left(|h|^2\right).$$

At the zeroth degree of approximation then p maps to $f(p)$. When we consider nearby points $p+h$ to p, then the first degree approximation is the map to $f(p) + f'(p)h$. The effect of multiplying by $f'(p)$ is a scaling by $|f'(p)|$ and rotation by $\arg f'(p)$.

On occasion we will also find the following result useful.

Theorem 0.28 ***(Inverse function theorem)*** *Let $f\colon \mathbb{R}^n \to \mathbb{R}^n$ be a smooth map defined near p. If $\mathrm{d}f_p$ is invertible then f is a local diffeomorphism. That is there is a smooth map g defined near $f(p)$ such that $g(f(x)) = x$ and $f(g(y)) = y$ for x near p and y near $f(p)$.*

Note that when $f = u + iv$ is holomorphic the determinant $|\mathrm{d}f_p| = u_x^2 + v_x^2 = |f'(p)|^2$ and so f will be a local diffeomorphism if and only if it is conformal at p.

0.6 Identification Spaces

This is material from Part A *Topology.*

Definition 0.29 *Let $(X, \mathcal{T})$ be a topological space and $f\colon X \to Y$ be a map onto a set Y. Then the **quotient topology** on Y is the collection*

$$\tau = \left\{U \subseteq Y \mid f^{-1}(U) \in \mathcal{T}\right\},$$

and (Y,τ) is called a ***quotient space.***

As pre-image respects unions and intersections then τ is closed under arbitrary unions and finite intersections. Further $f^{-1}(\varnothing)=\varnothing\in\mathcal{T}$ and $f^{-1}(Y)=X\in\mathcal{T}$. Thus τ is a topology.

By definition, $f\colon (X,\mathcal{T})\to(Y,\tau)$ is continuous. Indeed τ is the finest topology on Y such that f is continuous.

Definition 0.30 *Given an equivalence relation $\sim$ on a topological space $(X,\mathcal{T})$ then there is a natural surjective map*

$$\pi\colon X\to X/\!\sim \qquad \text{given by} \qquad x\mapsto[x]$$

which sends an element x to its equivalence class $[x]$. In this case $(X/\!\sim,\tau)$ is referred to as an ***identification space****.*

Example 0.31 *The quotient space of any compact (resp. connected) space is compact (resp. connected). This is because the continuous image of a compact (resp. connected) space is compact (resp. connected).*

Example 0.32 *Define $\sim$ on $\mathbb{R}$ by $x\sim y$ if and only if $x-y\in\mathbb{Z}$. Show that $\mathbb{R}/\!\sim$, which is also written $\mathbb{R}/\mathbb{Z}$, is homeomorphic to the circle S^1.*

Solution. The bijection $\mathbb{R}/\mathbb{Z}\to S^1$ defined by $[x]\mapsto e^{2\pi ix}$ is a homeomorphism. It is an easy check that basic open subsets in the circle correspond to open subsets of $\mathbb{R}$ which are unions of equivalence classes. ■

Example 0.33 *Define $\sim$ on $\mathbb{R}$ by $x\sim y$ if and only if $x-y\in\mathbb{Q}$. Show that $\mathbb{R}/\!\sim$, which is also written $\mathbb{R}/\mathbb{Q}$, has the trivial topology.*

Solution. Let U be a non-empty open set in $\mathbb{R}/\mathbb{Q}$. Then $U+\mathbb{Q}$ is open in $\mathbb{R}$ and is a union of equivalence classes. But as a non-empty open subset of $\mathbb{R}$ contains a representative of each equivalence class we have $U+\mathbb{Q}=\mathbb{R}$ and hence $U=\mathbb{R}/\mathbb{Q}$. ■

Example 0.34 *Define $\sim$ on $\mathbb{C}$ by $z_1\sim z_2$ if and only if there exists $\lambda>0$ such that $z_1=\lambda z_2$. Show that $\mathbb{C}/\!\sim$ is not Hausdorff.*

Solution. In a Hausdorff space singleton points are closed. But in $\mathbb{C}/\!\sim$ the only closed point is $[0]$. Note the closure of $[1]$ is $[1]\cup[0]$. ■

1. INTRODUCTION

Definition 1.1 *A **topological surface**, or **topological 2-manifold**, is a Hausdorff topological space S such that for every $p \in S$ there is an open set $U \subseteq S$ and a homeomorphism $\varphi\colon U \to V$ where V is an open subset of $\mathbb{R}^2$. Such a surface is referred to as an **abstract** topological surface, the term 'abstract' refers to the fact that the surface is not situated (or 'embedded') in some Euclidean space.*

*The map φ is called a **chart** or **patch** and a collection $\{\varphi_i\colon U_i \to V_i\}$ such that*

$$\bigcup_i U_i = S$$

*is known as an **atlas**.*

*The surface S is called **closed** if it is compact.*

Remark 1.2 *In Definition 1.1 we have defined an 'abstract' topological surface. The surface has not been situated in any Euclidean space; the surface's topology is part of the definition, rather than being inherited as a subspace of some ambient space. This may contrast with most previous examples you have of surfaces, especially compared with parameterized surfaces discussed in Prelims Geometry.*

*As a consequence of Whitney's embedding theorem, every (separable) topological surface can be embedded in $\mathbb{R}^3$ or $\mathbb{R}^4$, so the benefit of the above definition may be even less clear. Here an **embedding** is a continuous, injective map which is a homeomorphism between the surface and its image. However these embeddings can often be complicated functions, in which case it's easier to work with an abstract definition. For example the Klein bottle, which we will introduce soon, cannot be embedded in $\mathbb{R}^3$; the hyperbolic plane, which is topologically just $\mathbb{R}^2$, cannot be isometrically embedded in $\mathbb{R}^3$.*

With an atlas we can therefore parameterize the surface S. At this point the atlas provides no further structure to S, which already has a Hausdorff topology. However these parameters provide a useful means with which to define functions on S. But $\mathbb{R}^2$ has (or can have) further structures – smooth, metric, orientable, complex – and we will in due course see how we can use atlases to consistently transfer these structures to surfaces.

Example 1.3 ***(Atlas for the sphere)*** *Let $S^2 = \{(x,y,z) \in \mathbb{R}^3 \mid x^2+y^2+z^2 = 1\}$. As $\mathbb{R}^3$ is Hausdorff then so is S^2, and as S^2 is closed and bounded then it is compact.*

The following six maps form an atlas for S^2.

$$\begin{aligned}
U_1 &= \left\{(x,y,z) \in S^2 \mid z > 0\right\}, \quad \varphi_1(x,y,z) = (x,y);\\
U_2 &= \left\{(x,y,z) \in S^2 \mid z < 0\right\}, \quad \varphi_2(x,y,z) = (x,y);\\
U_3 &= \left\{(x,y,z) \in S^2 \mid x > 0\right\}, \quad \varphi_3(x,y,z) = (y,z);\\
U_4 &= \left\{(x,y,z) \in S^2 \mid x < 0\right\}, \quad \varphi_4(x,y,z) = (y,z);\\
U_5 &= \left\{(x,y,z) \in S^2 \mid y > 0\right\}, \quad \varphi_5(x,y,z) = (x,z);\\
U_6 &= \left\{(x,y,z) \in S^2 \mid y < 0\right\}, \quad \varphi_6(x,y,z) = (x,z).
\end{aligned}$$

In each case $V_i = \varphi_i(U_i)$ is the open unit disc in $\mathbb{R}^2$. As $x^2+y^2+z^2=1$ for any $(x,y,z) \in S^2$ then at least one of the co-ordinates is non-zero, meaning every point of S lies in at least one patch.

We have thus shown S^2 to be a topological surface. Note an atlas for S^2 cannot consist of a single chart $\varphi\colon S^2 \to V$ as S^2 is compact and V is not, but it's not hard to find an atlas consisting of two charts.

Example 1.4 *(**Bug-eyed plane**) The following example shows the necessity of the requirement that S be Hausdorff. Consider $S = X/\sim$ where $X = \mathbb{R}^2 \times \{\pm 1\}$ and every point $(x,y,-1)$ is identified with $(x,y,1)$ except when $x=y=0$. The space S is then not Hausdorff as the two origins $(0,0,\pm 1)$ cannot be separated but the two charts $\varphi_{\pm 1}(x,y,\pm 1) = (x,y)$ form an atlas for S.*

Proposition 1.5 *Let S be a topological surface with atlas $\{\varphi_i\colon U_i \to V_i\}$. Let $f\colon S \to T$ be a map to a topological space T. Then f is continuous if and only if each $f \circ \varphi_i^{-1}\colon V_i \to T$ is continuous.*

Proof. If f is continuous then $f \circ \varphi_i^{-1}$ is the composition of two continuous maps and therefore continuous. Conversely suppose all these maps are continuous and take $p \in S$. As we have an atlas then $p \in U_i$ for some i and then $f = f \circ \varphi_i^{-1} \circ \varphi_i$ is continuous at p. ■

Example 1.6 *The **real projective plane** $\mathbb{P} = S^2/\{\pm 1\}$ is the space formed by identifying antipodal points of the sphere. Find an atlas for $\mathbb{P}$.*

Solution. Each equivalence class of points in $\mathbb{P} = S^2/\{\pm 1\}$ has a representative in one (or more) of the domains U_1, U_3, U_5 previously used in Example 1.3 to cover the sphere. Given a point where $z \neq 0$, we can assume in fact that $z > 0$ without loss of generality. Then the maps

$$\psi_1(x,y,z) = (x,y), \quad \psi_3(x,y,z) = (y,z), \quad \psi_5(x,y,z) = (x,z),$$

form an atlas for $\mathbb{P}$. ■

In the Prelims Geometry course, the definition of a parameterized surface was as follows.

Definition 1.7 *A **smooth parameterized surface** is a map,*

$$\mathbf{r}\colon U \to \mathbb{R}^3 \qquad (u,v) \mapsto (x(u,v), y(u,v), z(u,v))$$

from an open subset $U \subseteq \mathbb{R}^2$ to $\mathbb{R}^3$ such that

- *$\mathbf{r}$ is smooth i.e. x, y, z have continuous partial derivatives of all orders,*
- *$\mathbf{r}\colon U \to \mathbf{r}(U)$ is a homeomorphism,*
- *(**smoothness condition**) at each point of $\mathbf{r}(U)$ the vectors*

$$\mathbf{r}_u = \frac{\partial \mathbf{r}}{\partial u} \quad \text{and} \quad \mathbf{r}_v = \frac{\partial \mathbf{r}}{\partial v}$$

are linearly independent.

Comparing this with our earlier definition of a topological surface, we note that $\mathbf{r}^{-1}\colon \mathbf{r}(U) \to U$ is a chart forming an atlas by itself. So parameterized surfaces in $\mathbb{R}^3$ are examples of topological surfaces. However the adjective *smooth* suggests that we have more structure now than a topological surface generally has. The independence of the vectors $\mathbf{r}_u$ and $\mathbf{r}_v$ means that the surface has a well-defined *tangent plane* and *normal* at each point. But it's currently unclear how we might generalize this notion to a topological surface S that is not situated in Euclidean space. Around each point $p \in S$ we can assign co-ordinates via a chart $\varphi\colon U \to V$ and so it might seem reasonable to say that a function $f\colon S \to \mathbb{R}$ is smooth at p if

$$f \circ \varphi^{-1}\colon V \to \mathbb{R}$$

is smooth. Recall that V is an open subset of $\mathbb{R}^2$ so this would just mean that $f \circ \varphi^{-1}$ has partial derivatives of all orders. The catch is that, when p is in the domain of more than one chart, f might be deemed to be smooth at p using one chart and not smooth using another chart. We need to ensure we have consistency across the surface.

Definition 1.8 *Given an atlas $\{\varphi_i\colon U_i \to V_i\}$ for a topological surface S, if $U_i \cap U_j \neq \varnothing$ then*

$$\varphi_i \circ \varphi_j^{-1}\colon \varphi_j\left(U_i \cap U_j\right) \to \varphi_i\left(U_i \cap U_j\right)$$

is known as a ***transition map****.*

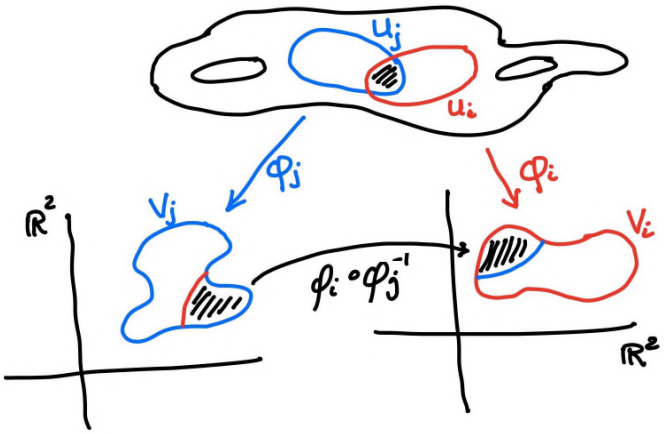


Figure 1.1 – a transition map

For a topological surface the transition maps are always homeomorphisms. So if $f \circ \varphi_i^{-1}$ is continuous then $f \circ \varphi_j^{-1}$ automatically is too. But we need to further require the transition maps to be smooth, to have a consistent notion of smoothness across a surface. Hence we define:

Definition 1.9 *A* ***differentiable surface****, or* ***differentiable 2-manifold****, is a topological surface S with an atlas $\{\varphi_i\colon U_i \to V_i\}$ such that all the transition maps*

$$\varphi_i \circ \varphi_j^{-1}\colon \varphi_j\left(U_i \cap U_j\right) \to \varphi_i\left(U_i \cap U_j\right)$$

are smooth. Such an atlas is called a ***differentiable structure*** *on S.*

Definition 1.10 *(a) Let S be a differentiable surface with atlas $\{\varphi_i \colon U_i \to V_i\}$. We define $f \colon S \to \mathbb{R}$ to be* ***smooth*** *at $p \in U_i$ if*

$$f \circ \varphi_i^{-1} \colon V_i \to \mathbb{R}$$

is smooth. A quick check shows there is no possibility of inconsistency.

(b) Let Σ be a second differentiable surface with atlas $\{\psi_i \colon A_i \to B_i\}$ and let $f \colon S \to \Sigma$ be a map between the surfaces. Let $p \in S$, so that $p \in U_i$ for some i, and then $f(p) \in B_j$ for some j. We define f to be ***smooth*** *at p if*

$$\psi_j \circ f \circ \varphi_i^{-1}$$

is smooth at $\varphi_i(p)$. As the transition maps are smooth there is again no chance of inconsistency.

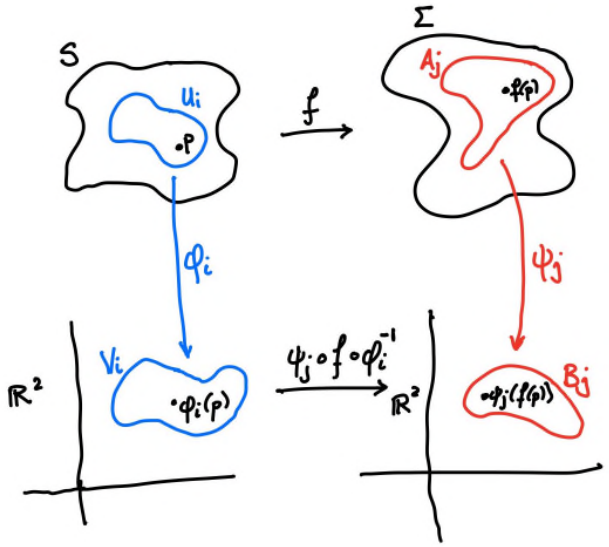


Figure 1.2 – a map between surfaces

Remark 1.11 *(**Consistency and other structures**) Note that a transition map ofr a differentiable surface is bijective, is smooth, and its inverse – another transition map – is also smooth. That is to say that the transition maps are* ***diffeomorphisms****.*

We can now see how the previous definitions can be generalized to higher dimensions to define topological manifolds and smooth manifolds. Despite surfaces being the focus of much study in the eighteenth and nineteenth centuries – by Euler, Lagrange, Monge, Gauss, Riemann, Möbius, et al. – a formal definition of surfaces (and manifolds) did not arise until the 1930s, variously due to Whitehead, Whitney and Veblen.

The transition maps are the key to assigning structures to a surface beyond the purely topological. Requiring the transition maps to be smooth means we can consistently define a smooth structure on the whole surface. But $\mathbb{R}^2$ naturally has other structures:

- *metric structure – we would then need the transition maps to be isometries;*
- *orientability – we would then need the transition maps to be orientation-preserving;*

- *complex structure – we can identify $\mathbb{R}^2$ with $\mathbb{C}$ and would then need the transition maps to be biholomorphic (that is, conformally equivalent).*

Note that a single patch of surface can be assigned any of these structures. However for a general topological surface, it may not be possible to endow a surface globally with certain structures precisely because of its topology. The real projective plane cannot be consistently oriented; the sphere cannot be given a metric structure with everywhere 'negative curvature'. When we meet Riemann surfaces later we will see there are a great deal of differences between complex structures and real smooth ones. In higher dimensions, these problems are yet more complicated and subtle.

We say a little now about how a metric structure can be assigned to a co-ordinate patch of a surface. We will revisit these ideas in detail in Chapter 3. We have already noted that $\mathbf{r}_u(p)$ and $\mathbf{r}_v(p)$ are independent tangent vectors of a point p in a co-ordinate patch $\mathbf{r}(U) \subseteq \mathbb{R}^3$. Given a curve $\gamma(t) = \mathbf{r}(u(t), v(t))$ where $a \leqslant t \leqslant b$ then, by the chain rule,

$$\dot{\gamma}(t) = \dot{u}\mathbf{r}_u + \dot{v}\mathbf{r}_v$$

and

$$|\dot{\gamma}(t)|^2 = E\dot{u}^2 + 2F\dot{u}\dot{v} + G\dot{v}^2$$

where

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

Definition 1.12 *The quadratic form $I_p \colon T_p \to \mathbb{R}$,*

$$\alpha\mathbf{r}_u + \beta\mathbf{r}_v \mapsto |\alpha\mathbf{r}_u + \beta\mathbf{r}_v|^2 = E\alpha^2 + 2F\alpha\beta + G\beta^2$$

*on the tangent space $T_p = \langle \mathbf{r}_u, \mathbf{r}_v \rangle$ is known as the **first fundamental form**. Any property of a surface that can be expressed in terms of the first fundamental form is said to be **intrinsic**.*

The first fundamental form expresses how the co-ordinate domain has been curved on to the surface. All metric properties of the surface can be expressed in terms of the first fundamental form. We will need to consider quite what we mean by tangent spaces when we have an abstract surface, rather than one situated in $\mathbb{R}^3$, but we will deal with that in Chapter 3. In the meantime note that lengths and areas can be expressed in terms of the first fundamental form; importantly these definitions apply whatever Euclidean space the surface is situated in.

The **length** of the above curve γ equals

$$\mathcal{L}(\gamma) = \int_a^b |\dot{\gamma}(t)| \, \mathrm{d}t = \int_a^b \sqrt{E\dot{u}^2 + 2F\dot{u}\dot{v} + G\dot{v}^2} \, \mathrm{d}t.$$

We have also previously defined the **area** of $\mathbf{r}(U)$ by

$$\mathcal{A} = \iint_U |\mathbf{r}_u \wedge \mathbf{r}_v| \, \mathrm{d}u \, \mathrm{d}v.$$

One issue with this definition is that the vector product $\wedge$ is defined in $\mathbb{R}^3$ but not generally in higher dimensions. However, the scalar quadruple product gives

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

Hence we can instead define the area of $\mathbf{r}(U)$ as

$$\mathcal{A} = \iint_U \sqrt{EG - F^2}\, \mathrm{d}u\, \mathrm{d}v,$$

a definition which is well-defined whatever $\mathbb{R}^n$ the surface is situated in.

Let's conclude this introduction by considering the transition maps for the atlases we previously defined for the sphere and real projective plane.

Example 1.13 *(The sphere reprised.) Consider the two charts*

$$\begin{aligned} U_1 &= \left\{(x,y,z) \in S^2 \mid z > 0\right\}, \quad \varphi_1(x,y,z) = (x,y); \\ U_3 &= \left\{(x,y,z) \in S^2 \mid x > 0\right\}, \quad \varphi_3(x,y,z) = (y,z). \end{aligned}$$

So $U_1 \cap U_3 = \{(x,y,z) \in S^2 \mid x, z > 0\}$ *is an open quarter of the sphere and*

$$\begin{aligned} \varphi_1(U_1 \cap U_3) &= \left\{(x,y) \in \mathbb{R}^2 \mid x^2 + y^2 < 1, x > 0\right\} \\ \varphi_3(U_1 \cap U_3) &= \left\{(x,y) \in \mathbb{R}^2 \mid x^2 + y^2 < 1, y > 0\right\} \end{aligned}$$

and

$$(u(x,y), v(x,y)) = \varphi_1 \circ \varphi_3^{-1}(x,y) = \left(\sqrt{1 - x^2 - y^2}, x\right).$$

Note that the Jacobian of this map equals

$$\begin{vmatrix} u_x & u_y \\ v_x & v_y \end{vmatrix} = \begin{vmatrix} \frac{-x}{\sqrt{1-x^2-y^2}} & \frac{-y}{\sqrt{1-x^2-y^2}} \\ 1 & 0 \end{vmatrix} = \frac{y}{\sqrt{1 - x^2 - y^2}} > 0.$$

That this is non-zero means that the transition map is smooth. That it is positive means that the transition map is orientation preserving. As this is true of the other transition maps too, then we have given the sphere the structure of an oriented differentiable surface.

Example 1.14 *(The real projective plane revisited.) Recall the charts*

$$\begin{aligned} \psi_1(x,y,z) &= (x,y), \quad \textit{where } z > 0 \textit{ by assumption WLOG;} \\ \psi_5(x,y,z) &= (x,z), \quad \textit{where } y > 0 \textit{ by assumption WLOG.} \end{aligned}$$

So $U_1 \cap U_5$ *consists of those* $[x\colon y\colon z]$ *where* $y \neq 0 \neq z$ *and* $x^2 + y^2 + z^2 = 1$. *Then*

$$\begin{aligned} \psi_1(U_1 \cap U_5) &= \left\{(x,y) \in \mathbb{R}^2 \mid x^2 + y^2 < 1, y \neq 0\right\}, \\ \psi_5(U_1 \cap U_5) &= \left\{(x,z) \in \mathbb{R}^2 \mid x^2 + z^2 < 1, z \neq 0\right\}, \end{aligned}$$

so in fact $\psi_1(U_1 \cap U_5) = \psi_5(U_1 \cap U_5)$. *Then*

$$(u(x,y), v(x,y)) = \psi_1 \circ \psi_5^{-1}(x,y) = (x,y),$$

as

$$(x,y) \overset{\psi_5^{-1}}{\mapsto} \left[x\colon \sqrt{1-x^2-y^2}\colon y\right] \overset{\psi_1}{\mapsto} \begin{cases} \left(x, \sqrt{1-x^2-y^2}\right) & \text{when } y > 0; \\ \left(-x, -\sqrt{1-x^2-y^2}\right) & \text{when } y < 0. \end{cases}$$

The Jacobian of this map when $y > 0$ *equals*

$$\begin{vmatrix} u_x & u_y \\ v_x & v_y \end{vmatrix} = \begin{vmatrix} 1 & \frac{-x}{\sqrt{1-x^2-y^2}} \\ 0 & \frac{-y}{\sqrt{1-x^2-y^2}} \end{vmatrix} = \frac{-y}{\sqrt{1-x^2-y^2}} < 0.$$

That this is non-zero means that the transition map is smooth. That it is negative means that the transition map is orientation-reversing. As the transition maps are all smooth then we have endowed $\mathbb{P}$ *with a differentiable structure. As this particular transition map is orientation-reversing then we have not endowed* $\mathbb{P}$ *with an oriented structure. It is then a somewhat harder matter to show that* ***no*** *oriented atlas exists.*

2. TOPOLOGICAL SURFACES

Recall from the introductory lecture the definition of a topological surface.

A **topological surface**, or **topological 2-manifold**, is a Hausdorff topological space S such that for every $p \in S$ there is an open set $U \subseteq S$ and a homeomorphism $\varphi\colon U \to V$ where V is an open subset of $\mathbb{R}^2$.

A surface S is called **closed** if it is compact. In this chapter we discuss the *classification of closed topological surfaces up to homeomorphism*. So two topological surfaces are to be considered the same if they are homeomorphic; the 'classification' then means providing a comprehensive list of the different homeomorphism classes with no omissions and no duplications.

This material was discussed at some length in the A5 topology course. The closed surfaces there were created as *identification spaces* (or quotient spaces) from closed polygons. Two examples are given below.

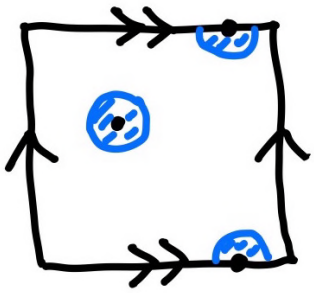


Figure 2.1 – torus

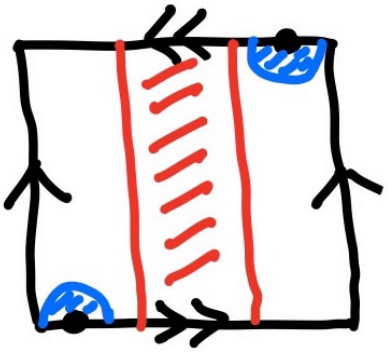


Figure 2.2 – Klein bottle

In Figure 2.1 a torus is formed by pairwise identifying the edges of the square $[0,1]^2$ as described by the arrows. So $(0,y)$ and $(1,y)$ are identified for $0 \leqslant y \leqslant 1$ (the single arrows) and $(x,0)$ and $(x,1)$ are identified for $0 \leqslant x \leqslant 1$ (the double arrows). The square is compact and so the resulting identification space also is. Around each interior point of the square we can associate an open disc U; points on the square's boundary can be associated with a disc split as two semi-discs as sketched in Figure 2.1.

Similarly, in Figure 2.2 a Klein bottle is formed by pairwise identifying the edges of the square $[0,1]^2$ as described by the arrows. So $(0,y)$ and $(1,y)$ are identified for $0 \leqslant y \leqslant 1$ (the single arrows) and $(x,0)$ and $(1-x,1)$ are identified for $0 \leqslant x \leqslant 1$ (the double arrows). The square is compact and so the resulting identification space also is. Around each interior point of the square we can associate an open disc U; points on the square's boundary can be associated with a disc split as two semi-discs as sketched in Figure 2.2. Note in the case of a boundary point on the bottom/top edges the semi-discs are not directly opposite one another because of the reverse identification. Note further, because of this reversed identification, the shaded rectangle is in fact a Möbius strip, rather than a cylinder.

Further examples include the

- **torus with $g \geqslant 0$ holes** or, equally, sphere with g handles (Figure 1.3);
- **sphere with $k \geqslant 1$ cross-caps** (Figure 1.4).

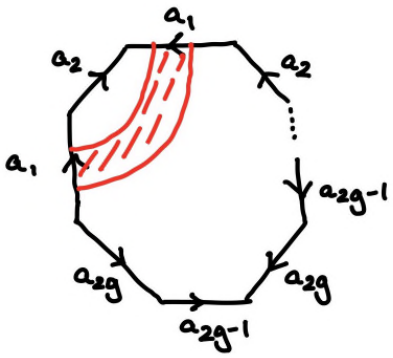


Figure 2,3 – torus with g holes

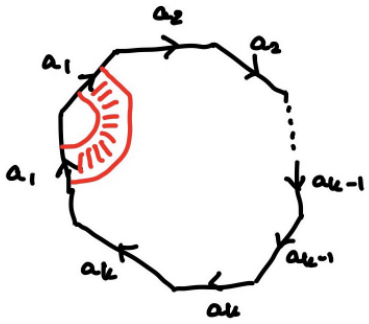


Figure 2.4 – sphere with k cross-caps

The torus with zero holes is the sphere. The torus with $g \geqslant 1$ can be formed by pairwise identifying the edges of a $4g$-gon as shown in Figure 2.3. Note that, in each case, the shaded region connecting identified edges is a cylinder. Consequently the torus with g holes is an orientable surface. This canonical identification is can be denoted

$$a_1 a_2 a_1^{-1} a_2^{-1} a_3 a_4 a_3^{-1} a_4^{-1} \cdots a_{2g-1} a_{2g} a_{2g-1}^{-1} a_{2g}^{-1}.$$

Each string $a_i a_{i+1} a_i^{-1} a_{i+1}^{-1}$ represents a further hole or handle being attached to the surface. See Proposition 2.9.

The sphere with $k \geqslant 1$ cross-caps can be formed by pairwise identifying the edges of a $2k$-gon as shown in Figure 2.4. Note that, in each case, the shaded bar connecting identified edges is a Möbius strip. Consequently the sphere with $k \geqslant 1$ cross-caps is an non-orientable surface. This canonical identification is denoted

$$a_1 a_1 a_2 a_2 \cdots a_k a_k.$$

A cross-cap is formed in the sphere by making a cut and identified the cut's two sides in reverse orientation. This is the equivalent of sewing a Möbius strip into the sphere, which is what each string $a_i a_i$ represents. See Proposition 2.9.

Example 2.1 *The Klein bottle $\mathbb{K}$ is homeomorphic to the sphere with 2 cross-caps.*

Solution. These two versions of the Klein bottle can be transformed into one another as shown below.

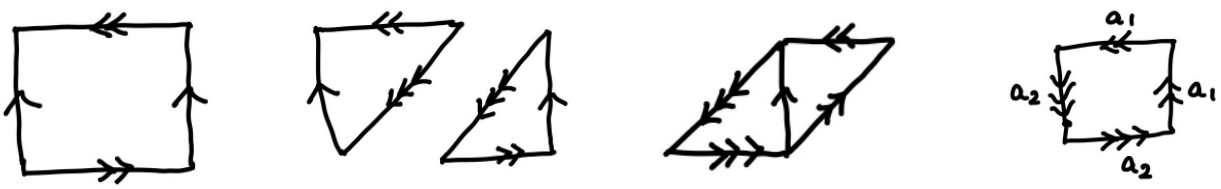


Figure 2.5 – equivalent Klein bottles

This transforms $\mathbb{K}$ from the surface in Figure 2.2 to the surface $a_1a_1a_2a_2$, the sphere with two cross-caps. ■

In the A5 topology course it was rigorously shown that:

- Every closed topological surface is homeomorphic to one of (a) a torus with $g \geqslant 0$ holes or (b) a sphere with $k \geqslant 1$ cross-caps.

This is half the classification theorem. The above is a comprehensive list of all closed topological surfaces up to homeomorphisms. There are no omissions but there may yet be duplications. We need one or more topological invariants which can be used to distinguish between the homeomorphism classes listed above. The two invariants we shall use are *orientability* and the *Euler characteristic.*

We already introduced the notion of orientability in the introductory lecture; a differentiable surface was orientable if it had an atlas with orientation-preserving transition map. We shall use, in this chapter, an equivalent criterion for orientability. This second definition of orientability is due to Klein.

Proposition 2.2 *A differentiable surface is non-orientable if and only if it contains a Möbius strip.*

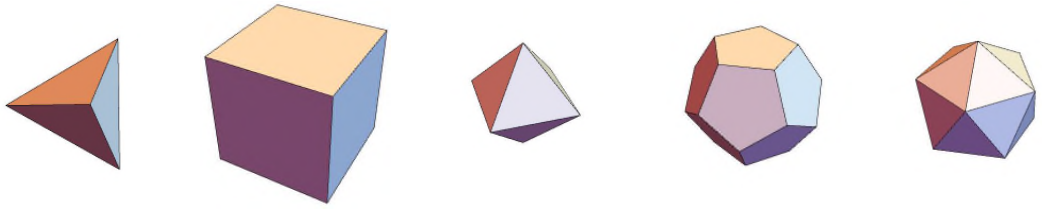
Proof. Say that a surface includes a Möbius strip. Then we can take an orientation-reversing curve along the Möbius strip and consider the co-ordinate patches it passes through (which can be taken to be finite by compactness). Each transition map between patches cannot be orientation-preserving or else the curve would not be orientation-reversing.

Conversely, suppose that the surface contains no Möbius strip and so no orientation-reversing curve. Make a choice of orientation at a fixed point. Any other point can be connected by a path to the fixed point and the chosen orientation can be extended consistently to the second point. Thus the surface is orientable. ■

So the Klein bottle, and more generally, the spheres with k cross-caps are therefore non-orientable. As the tori with n holes can be embedded in $\mathbb{R}^3$ then they are orientable; we can consistently associate an outward-pointing normal on the entirety of such a surface. Thus orientability separates out the closed surfaces into two families, but we need a further invariant to separate the orientable surfaces from one another and likewise separate out the non-orientable surfaces. This invariant is the *Euler characteristic.*

You may well be aware that for the Platonic solids $V - E + F = 2$ where V, E, F respectively denote the number of vertices, edges and faces on the solid.

surface	V	E	F
tetrahedron	4	6	4
cube	8	12	6
octahedron	6	12	8
dodecahedron	20	30	12
icosahedron	12	30	20



tetrahedron cube octahedron dodecahedron icosahedron

Figure 2.6 – the Platonic solids

Indeed this relation is true for any polyhedron of the same *shape* (such as any pyramid or a cuboid). That is $V - E + F$ will equal 2 for any polyhedron that is homeomorphic to a sphere. So this number 2 is known as the Euler characteristic of the sphere.

Remark 2.3 *Euler arrived at his $V - E + F = 2$ formula for convex polyhedra in 1750 (in a letter to Goldbach) and this is arguably one of the first topological results. It is, in fact, equivalent to a result of Descartes' from 1639 but Euler's formulation of the result was more obviously topological in nature. The formula had been noted as early as 1537 by Francesco Maurolico. In 1811 Cauchy gave a semi-rigorous proof of the formula, though it would not be considered watertight by modern standards.*

We need to be a little careful in how we assign vertices, edges and faces to the surface. For example, were we to assign no vertices and no edges to a sphere and treat the entire surface as a face then we would arrive at an Euler characteristic of $0 - 0 + 1 = 1 \neq 2$, so presumably this should not be permitted. Likewise a single edge as an equator, no vertices and two hemispherical faces gives $0 - 1 + 2 = 1 \neq 2$ and should again not be admissable. The important point is that our vertices, edges and faces make a *subdivision* of the surface.

Definition 2.4 *Let X be a closed topological surface.*

*(a) An **edge** on X is the image of a continuous map $f\colon [0,1] \to X$ which is 1-1 except possibly that $f(0) = f(1)$.*

*(b) A **subdivision** of X is a finite set of edges, together with a finite set of points of X, called **vertices** (singular: vertex), such that*

(i) each edge begins and ends in a vertex and passes through no other vertices;

(ii) two edges intersect, at most, at their ends;

(iii) if Γ is the union of the edges then each connected component of $X \backslash \Gamma$ is homeomorphic to $\mathbb{R}^2$.

*(c) The closure of a connected component of $X \backslash \Gamma$ is known as a **face**.*

With the earlier inadmissible examples: we cannot use the entire surface of the sphere as a face as it is not homeomorphic to $\mathbb{R}^2$ invalidating (iii) – if we included a single solitary vertex on the sphere we would then have a valid subdivision; for the second example the edge does not begin and end in a vertex invalidating (i) – if we included a vertex on the edge then we would have a valid subdivision.

For those that did A5 we note the following:

Example 2.5 *If a topological surface is the realisation $|K|$ of a simplicial complex K then the simplicial complex is a valid subdivision of $|K|$ with the 0-simplices as vertices, the 1-simplices as edges and the 2-simplices as faces.*

The important result – which we shall not prove in this course – is the following:

Theorem 2.6 *Let X be a topological surface. Then the number*

$$\chi(X) = V - E + F$$

*is the same for any subdivision, where V, E, F are respectively the number of vertices, edges and faces in the subdivision. The number $\chi(X)$ is known as the **Euler characteristic** of X, and also sometimes as its **Euler number** or its **Euler-Poincaré characteristic**.*

Consequently the Euler characteristic is a topological invariant of the surface – that is, it is preserved by homeomorphisms.

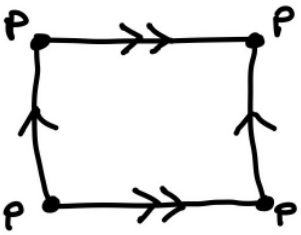


Figure 2.7 – torus

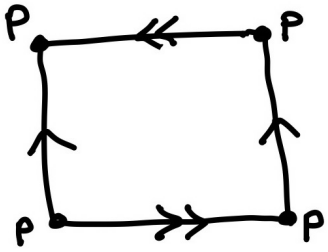


Figure 2.8 – Klein bottle

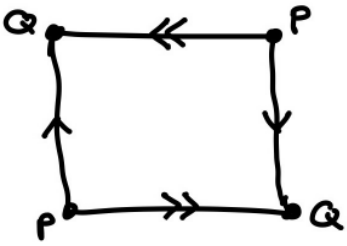


Figure 2.9 – projective plane

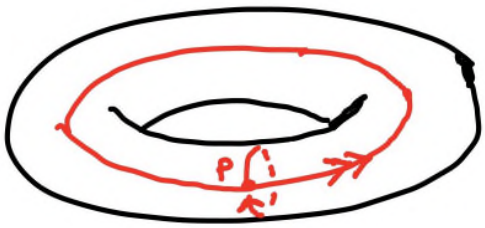


Figure 2.10 – torus with subdivision

Example 2.7 *Find the Euler characteristics of (a) the torus, (b) the Klein bottle, (c) the projective plane.*

Solution. Each of these surfaces begin with a square face, bounded by four edges and four vertices. The important point is how many vertices and edges remain after the identifications are made. In each case there is just one face, the square itself.

(a) For the torus $\mathbb{T}$ (Figure 2.7) the four edges are pairwise identified to leave two edges – the single arrows and the double arrows. Following the identifications around the four vertices are all identified to become a single vertex P. How these edges and vertices would look on a torus is drawn in Figure 2.10. This means that the Euler characteristic of the torus is

$$\chi(\mathbb{T}) = 1 - 2 + 1 = 0.$$

(b) For the Klein bottle $\mathbb{K}$ (Figure 2.8) the four edges are pairwise identified to leave two edges – the single arrows and the double arrows. Following the identifications around the four vertices are all identified to become a single vertex P. This means that the Euler characteristic of the Klein bottle is

$$\chi(\mathbb{K}) = 1 - 2 + 1 = 0.$$

So $\mathbb{T}$ and $\mathbb{K}$ have the same Euler characteristic despite not being homeomorphic – $\mathbb{T}$ is orientable, whilst $\mathbb{K}$ is not.

(c) For the projective plane $\mathbb{P}$ (Figure 2.8) the four edges are pairwise identified to leave two edges – the single arrows and the double arrows. Following the identifications around the four vertices become identified a two vertices, P and Q. This means that the Euler characteristic of the Klein bottle is

$$\chi(\mathbb{P}) = 2 - 2 + 1 = 1.$$

It's apparent from the identification that $\mathbb{P}$ is the sphere with 1 cross-cap. Just treat the single and double arrows as one edge and we see that $\mathbb{P}$ is the surface a_1a_1. ■

Example 2.8 *Find the Euler characteristic of the surface created from the three polygons below. Is the surface orientable?*

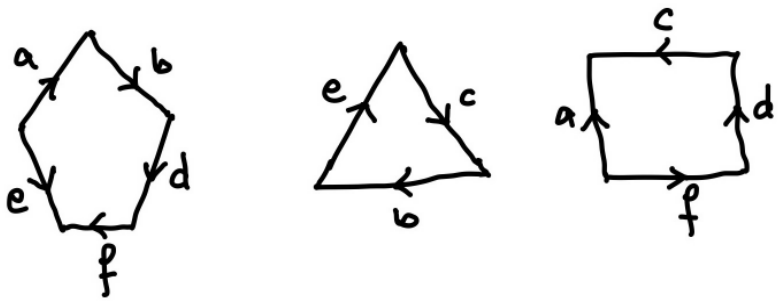


Figure 2.11 – a more complicated example

Solution. The surface, as drawn, comes with a natural subdivision. There are 3 faces – the pentagon, triangle and square – and 6 edges, namely a, b, c, d, e, f. It's not immediately clear how the original 12 vertices identify though.

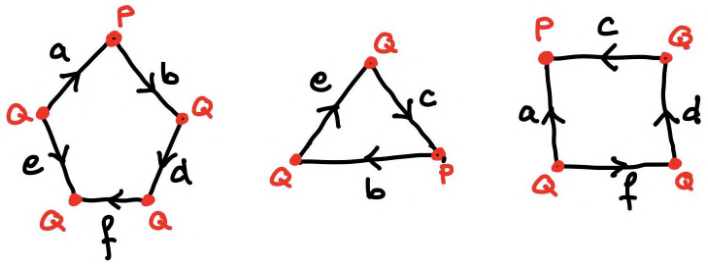


Figure 2.12 – counting the vertices

If we label the vertex at the top of the pentagon as P then, by following around the identifications, we can see what other vertices it is identified with. P is at the front end of a and so we follow around the identifications

$$\text{front of } a \;\to\; \text{front of } c \to \text{back of } b \to \text{front of } a$$

and we are back where we started. So the three vertices labelled P in Figure 2.12 are identified together. Labelling another vertex Q we can follow around the identifications and see in this case that the remaining 9 vertices are identified with Q. Thus there are 2 vertices once identified and we find

$$\chi = 2 - 6 + 3 = -1.$$

We cannot immediately see whether there is a Möbius strip within the surface as each edge is identified with an edge on a different face. However if we bring the pentagon and triangle together as in Figure 2.13

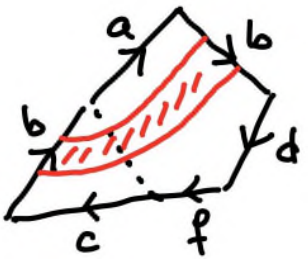


Figure 2.13 – non-orientability

we now see that the senses of the two b-edges are the same or equivalently the shaded region is a Möbius strip. Thus the surface is non-orientable. ■

Proposition 2.9 *(a) Adding a handle to a surface reduces the Euler characteristic by* 2.
(b) Adding a cross-kap to a surface reduces the Euler characteristic by 1.

Proof. (a) As shown in Figure 2.14 a handle can be added to a surface and subdivided with two further edges. The vertex shown is already part of the original surface's subdivision. As E increases by 2 then $V - E + F$ reduces by 2.

(b) As shown in Figure 2.15 a cross-cap can be added to a surface and subdivided using two new edges and a new vertex. The unlabelled vertex shown is already part of the original surface's subdivision. As E increases by 2 and V by 1 then $V - E + F$ reduces by 1 overall. ■

Corollary 2.10 *(a) The Euler characteristic of the torus with $g \geqslant 0$ holes equals $2 - 2g$.*
(b) The Euler characteristic of the sphere with $k \geqslant 1$ cross-caps equals $2 - k$.

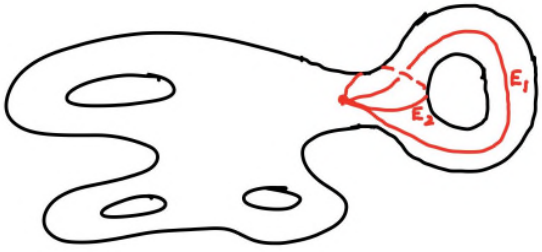


Figure 2.14 – adding a handle

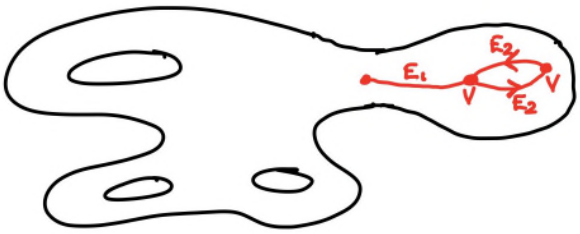


Figure 2.15 – adding a cross-cap

We are now in a position to state the classification theorem as we see that, between them, orientability and the Euler characteristic are enough to distinguish the homeomorphism classes.

Theorem 2.11 *(**Classification Theorem for Closed Surfaces**) Let X be a closed topological surface. Then X is homeomorphic to precisely one of the following.*

*(a) If X is orientable, then X is homeomorphic to a torus with $g \geqslant 0$ holes. g is called the **genus** of X.*

(b) If X is non-orientable, then X is homeomorphic to a sphere with $k \geqslant 1$ cross-caps.

Proof. From the A5 result we know that X is homeomorphic to one of these surfaces. None of the surfaces in list (a) is homeomorphic to a surface in list (b) by orientability. Further the Euler characteristics of the surfaces in list (a) are distinct, thus separating them topologically. And the same can be said of the surfaces in list (b). ■

Remark 2.12 *It is worth noting that the early topologists who 'proved' the classification theorem did not have available in their time the rigorous definitions necessary to prove their results to modern standards. In 1861 Möbius gave an early sketch proof of the classification for orientable surfaces, and Von Dyck gave a sketch proof for all closed surfaces in 1888. But without any formal definition of what a surface is, these proofs can at best be considered incomplete. Somewhat differently expressed rigorous versions of the classification theorem would be proved by Dehn and Heegaard in 1907 and by Brahana in 1921.*

Remark 2.13 *The above classification theorem relates to closed topological surfaces up to homeomorphism; we could easily consider instead closed differentiable surfaces up to diffeomorphism and the classification theorem would essentially read the same. The situation is similar in 3 dimensions but there are topological 4-manifolds which admit no differentiable structure and others which admit many; indeed there are 'exotic' versions of $\mathbb{R}^4$ which are homeomorphic to the standard $\mathbb{R}^4$ but not diffeomorphic to it.*

When it comes to 'complex structures' on surfaces the situation is very different and considerably more subtle. Riemann surfaces are necessarily orientable so no complex structure can be given to a sphere with k cross-caps. Only one structure, up to biholomorphism, can be put on the sphere but uncountably many can be put on a torus.

Remark 2.14 *We now see that the surface created in Example 2.8 is a sphere with 3 cross-caps. Indeed having worked out that the Euler characteristic equals -1 we did not need to determine the orientability as this is the only surface, up to homeomorphism, with this Euler characteristic.*

Remark 2.15 *Euler noted his formula for polyhedra that are topologically a sphere around 1750. The French-Swiss mathematician, Simon Lhuilier, noted in 1812 that $V-E+F = 2(1-g)$ when a polyhedron has g holes – this number g is called the polyhedron's genus.*

A modern demonstration of the topological invariance of the Euler characteristic usually appears in an algebraic topology course – see the Part C course of that name. In fact, the Euler characteristic is a homotopy invariant – homotopy equivalence is a more general notion than that of being homeomorphic. The Euler characteristic of a surface is the alternating sum of its Betti numbers.

$$\chi = b_0 - b_1 + b_2.$$

For an n-dimensional manifold, Betti numbers $b_0, b_1, \ldots, b_n$ can be defined which are the ranks of the manifold's homology groups which are topological invariants by definition. For the torus with g holes we have

$$b_0 = 1, \qquad b_1 = 2g, \qquad b_2 = 1,$$

giving $\chi = 2 - 2g$. That $b_0 = 1$ signifies the surface to be connected and that $b_2 = 1$ signifies that it has an 'inside' or is orientable. b_1 equalling $2g$ represents the loops that go through or go around each of the g holes. For the sphere with k cross-caps,

$$b_0 = 1, \qquad b_1 = k - 1, \qquad b_2 = 0,$$

giving $\chi = 2 - k$. That $b_0 = 1$ signifies connectedness and $b_2 = 0$ signifies non-orientability. b_1 equalling $k - 1$ represents $k - 1$ that loops are (in some technical sense) independent. Much of this early work was due to Poincaré around the end of the nineteenth century and the start of the twentieth and consequently the Euler characteristic is commonly referred to as the Euler-Poincaré characteristic.

More complicated surfaces can be created from simpler ones using the *connected sum*.

Definition 2.16 *Given two closed topological surfaces X_1 and X_2, their connected sum $X_1 \# X_2$ is created by removing two small discs, one from each surface, and identifying the circumferences of the two discs.*

Note that $X_1 \# X_2$ is orientable if and only if X_1 and X_2 are both orientable. The Euler characteristic of the connected sum can be quickly determined – as below – and we can then see that the torus $\mathbb{T}$ and projective plane $\mathbb{P}$ can be used as the building blocks for general closed topological surfaces.

Theorem 2.17 *Let X_1 and X_2 be closed topological surfaces. Then*

$$\chi(X_1 \# X_2) = \chi(X_1) + \chi(X_2) - 2.$$

Corollary 2.18 *(a) For $g \geqslant 0$, $\chi(\mathbb{T}^{\#g}) = 2 - 2g$.*
(b) For $k \geqslant 1$, $\chi(\mathbb{P}^{\#k}) = 2 - k$.

Proof. Say that X_i has subdivisions with V_i, E_i, F_i vertices, edges and faces and suppose that one of the faces in each subdivision is a triangle. When those two triangles are removed, and their boundaries identified, then 6 vertices become 3, 6 edges become 3 and 2 faces are lost. Thus

$$V_\# = V_1 + V_2 - 3, \qquad E_\# = E_1 + E_2 - 3, \qquad F_\# = F_1 + F_2 - 2$$

so that

$$\begin{aligned} \chi(X_1 \# X_2) &= (V_1 + V_2 - 3) - (E_1 + E_2 - 3) + (F_1 + F_2 - 2) \\ &= (V_1 - E_1 + F_1) + (V_2 - E_2 + F_2) - 2 \\ &= \chi(X_1) + \chi(X_2) - 2. \end{aligned}$$

The corollaries then follow by induction noting

$$\begin{aligned}\chi(\mathbb{T}^{\#g}) &= \chi(\mathbb{T}^{\#g-1}\#\mathbb{T}) = \chi(\mathbb{T}^{\#g-1}) + 0 - 2 = \chi(\mathbb{T}^{\#g-1}) - 2;\\ \chi(\mathbb{P}^{\#k}) &= \chi(\mathbb{P}^{\#k-1}\#\mathbb{P}) = \chi(\mathbb{P}^{\#k-1}) + 1 - 2 = \chi(\mathbb{P}^{\#k-1}) - 1,\end{aligned}$$

with the initial steps verified by

$$\begin{aligned}\chi(\mathbb{T}^{\#0}) &= \chi\left(S^2\right) = 2 = 2 - 2 \times 0;\\ \chi(\mathbb{P}^{\#1}) &= \chi(\mathbb{P}) = 1 = 2 - 1.\end{aligned}$$

The corollaries are essentially alternative proofs of Proposition 2.9. ■

We shall see later,in Chapter 5, with the Gauss-Bonnet theorem, the Poincaré-Hopf theorem and in elements of Morse theory, that the Euler characteristic is a topological obstruction to the global analysis and geometry of a surface.

3. SMOOTH AND GEOMETRIC SURFACES

In the introductory lecture we recalled the definition of a parameterized surface and introduced the notion of a differentiable structure on a surface. Here we will mainly be discussing the *local* geometric structure of surfaces, so it will be sufficient to focus on parameterized surfaces, though we will wish to make sure our definitions are not dependent on the choice of parameterization.

Definition 3.1 *Let $\mathbf{r}(U)$ be a smooth parameterized surface in $\mathbb{R}^3$ and let $p = \mathbf{r}(u_0, v_0)$. The **tangent plane** to $\mathbf{r}(U)$ at p is the plane through p that is parallel to*

$$\mathbf{r}_u(u_0, v_0) \quad \text{and} \quad \mathbf{r}_v(u_0, v_0).$$

*The **tangent space** $T_p(\mathbf{r}(U))$ is the vector space spanned by the above two vectors and any element of $T_p(\mathbf{r}(U))$ is called a **tangent vector.** It is easy to check that the tangent space at p consists of all the tangent vectors to all curves in $\mathbf{r}(U)$ which pass through p.*

Note that a parameterized surface is a surface in $\mathbb{R}^3$ with a preferred choice of co-ordinates from a particular chart $\mathbf{r}^{-1}$. But $\mathbf{r}(U)$ can also be associated with other charts, technically giving a different parameterized surface but we would hope that any questions asked of X (simply as a subspace of $\mathbb{R}^3$) such as, 'what is the area of X?' and 'what is the length of a curve in X?', will yield the same answers, irrespective of what chart we use. This will be an important consideration in all future definitions, namely that any new definitions are chart independent.

Proposition 3.2 *The tangent space is independent of the choice of parameterization.*

Proof. Let $\mathbf{r}(U) = \mathbf{s}(X)$ be two parameterizations

$$(u, v) \mapsto \mathbf{r}(u, v), \qquad (x, y) \mapsto \mathbf{s}(x, y).$$

If we have $\mathbf{r}(u, v) = \mathbf{s}(x, y)$ then by the chain rule

$$\mathbf{r}_u = x_u \mathbf{s}_x + y_u \mathbf{s}_y, \qquad \mathbf{r}_v = x_v \mathbf{s}_x + y_v \mathbf{s}_y.$$

Applying the vector product, we find

$$\mathbf{r}_u \wedge \mathbf{r}_v = (x_u y_v - x_v y_u)\, \mathbf{s}_x \wedge \mathbf{s}_y = \frac{\partial(x, y)}{\partial(u, v)} \mathbf{s}_x \wedge \mathbf{s}_y$$

are parallel. This is the normal direction to the tangent space which we see is also independent of the choice of parameterization. ■

Definition 3.3 *Let $\mathbf{r}(U)$ be a smooth parameterized surface in $\mathbb{R}^3$. A **normal vector** to $\mathbf{r}(U)$ at the point p is any (non-zero) vector orthogonal to $T_p(\mathbf{r}(U))$.*

The normal vectors are non-zero scalar multiples of $\mathbf{r}_u \wedge \mathbf{r}_v$ where $\wedge$ denotes the vector product in $\mathbb{R}^3$. The two unit vectors

$$\pm \frac{\mathbf{r}_u \wedge \mathbf{r}_v}{|\mathbf{r}_u \wedge \mathbf{r}_v|}$$

are the choices of **unit normal** to $\mathbf{r}(U)$ at p.

Definition 3.4 *The map* $\mathbf{n}$ *from* $\mathbf{r}(U)$ *to* S^2*, the unit sphere, which continuously sends* $\mathbf{r}(u,v)$ *to a unit normal* $\mathbf{n}(u,v)$ *is called the* ***Gauss map****.*

The definition of the differential of a map $f\colon \mathbb{R}^2 \to \mathbb{R}^2$ then extends to maps between parameterized surfaces in an obvious way.

Definition 3.5 *Let* X *and* Y *be smooth parameterized surfaces in* $\mathbb{R}^3$ *and let* $p \in X$*. For a smooth map* $f\colon X \to Y$ *(see Definition 1.10). Then the* ***differential*** *of* f *at* p *is the linear map*

$$\mathrm{d}f_p\colon T_pX \to T_{f(p)}Y$$

defined as follows. Let $\mathbf{v} \in T_pX$ *and let* $\gamma : (-\epsilon, \epsilon) \to X$ *be a smooth curve such that*

$$\gamma(0) = p \quad \textit{and} \quad \gamma'(0) = \mathbf{v}.$$

Then $f \circ \gamma$ *is a smooth curve in* Y *and as before we define*

$$\mathrm{d}f_p(v) = \mathrm{d}f_p(\gamma'(0)) = (f \circ \gamma)'(0).$$

A quick check shows that this definition is independent of the choice of curve γ*.*

Before we discuss any of the theory of surfaces, we should introduce some standard examples. We have already introduced differentiable atlases for the sphere and real projective plane, but we introduce two other parameterizations for (most of) the sphere here.

Example 3.6 ***(Parameterizing the sphere)*** *Consider the map* $\mathbf{r}_1\colon (-\pi, \pi) \times (0, \pi) \to \mathbb{R}^3$ *(see Figure 3.1) given by*

$$\mathbf{r}_1\colon (u,v) \mapsto (\cos u \sin v, \sin u \sin v, \cos v).$$

It is easy to check that the image of this map is contained in S^2*, the unit sphere centred at the origin. In fact the image is the whole sphere save for half a great circle. The parameter* u *is the angle between the projection of* $\mathbf{r}_1(u,v)$ *onto the* xy*-plane and the* x*-axis and* v *is the angle between* $\mathbf{r}_1(u,v)$ *and the* z*-axis.*

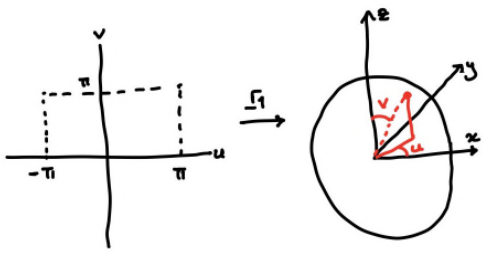


Figure 3.1 – spherical polars

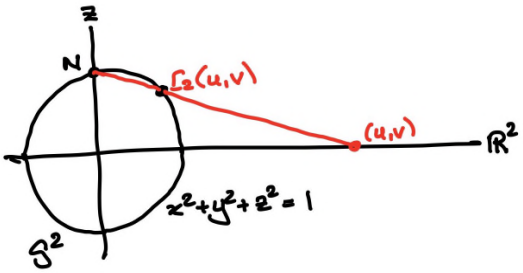


Figure 3.2 – stereographic projection

Consider also the map $\mathbf{r}_2\colon \mathbb{R}^2 \to \mathbb{R}^3$ *defined by*

$$\mathbf{r}_2\colon (u,v) \mapsto \left(\frac{2u}{u^2+v^2+1}, \frac{2v}{u^2+v^2+1}, \frac{u^2+v^2-1}{u^2+v^2+1} \right).$$

This again is a chart of the unit sphere. The map $\mathbf{r}_2$ *is in fact stereographic projection (see Figure 3.2) from the 'north pole'* $N = (0, 0, 1)$*; that is a point of* $(u, v) \in \mathbb{R}^2$ *is mapped to the intersection of the sphere with the line joining* $(u, v, 0)$ *and* N*. In this case the image of the sphere is the whole sphere minus* N*. This map is particularly relevant for setting up the extended complex plane with the* ***Riemann sphere****, a first example of a compact Riemann surface,*

Example 3.7 *(**Graphs**) Amongst the simplest examples of parameterized surfaces are graphs. Let* $f(x, y)$ *be a smooth function defined on an open set* $U \subseteq \mathbb{R}^2$*. Then the* ***graph*** *of* f *is the surface* $z = f(x, y)$ *and may be parameterized by*

$$\mathbf{r}(u, v) = (u, v, f(u, v)), \qquad (u, v) \in U.$$

These graphs seem almost too simple a family of surfaces to be of interest. One point of importance though is that any smooth surface in $\mathbb{R}^3$ *is, locally at least, a graph. That is:*

- *About any point of a smooth surface in* $\mathbb{R}^3$ *there is an open neighbourhood* U *such that* U *is a graph of the form* $z = f(x, y)$ *or* $y = f(x, z)$ *or* $x = f(y, z)$ *for some smooth function* f*. (Do Carmo, p.63).*

Indeed, for a general smooth surface $(x(u, v), y(u, v), z(u, v))$*, provided the normal is not horizontal or equivalently*

$$\frac{\partial(x, y)}{\partial(u, v)} = x_u y_v - x_v y_u \neq 0,$$

then the surface can be locally parameterized as $z = f(x, y)$ *for some* f*.*

Example 3.8 *(**The cone**) The punctured cone* $x^2 + y^2 = z^2$*,* $(z > 0)$ *in* $\mathbb{R}^3$ *may be smoothly parameterized by*

$$\mathbf{r}(u, v) = (u, v, \sqrt{u^2 + v^2}), \qquad u, v \in \mathbb{R},\ u^2 + v^2 \neq 0.$$

Note that the two sheeted cone $x^2 + y^2 = z^2$ *is not the image of any parameterization as no neighbourhood of the cone about* $(0, 0, 0)$ *is homeomorphic to an open subset of* $\mathbb{R}^2$*. (To see this consider the topological effect of removing the origin.)*

Consider now the one sheeted cone C *given by* $x^2 + y^2 = z^2$*,* $(z \geqslant 0)$*. This certainly is the image of a parameterization* $\mathbf{s}\colon \mathbb{R}^2 \to C$*, but for no such map is* C *smooth at the point* $(0, 0, 0)$*. To prove this we assume that the cone may be locally parameterized about* $(0, 0, 0)$ *as the graph of a smooth function. The only possibility (from* $z = f(x, y)$ *or* $y = f(x, z)$ *or* $x = f(y, z)$*) is a graph of the form* $z = f(x, y)$ *and by the definition of* C *we see that*

$$f(x, y) = \sqrt{x^2 + y^2}.$$

As f *is not differentiable at* $(0, 0)$ *then* $(0, 0, 0)$ *is not a smooth point of* C *for any parameterization. Such points on a surface are called* ***singular points****.*

Example 3.9 *(**Surfaces of revolution**) Surfaces may also be formed by taking a curve in $\mathbb{R}^3$ and using this curve to generate a surface. One such family are the surfaces of revolution. A **surface of revolution** is formed by rotating a smooth curve in, say, the xz-plane about the z-axis. For example, the cylinder in the above exercise is a surface of revolution.*

Assume the curve has equation $x = f(z) > 0$. Then the surface of revolution generated has equation $x^2 + y^2 = f(z)^2$. The surface cannot entirely be parameterized with one co-ordinate system but the map

$$\mathbf{r}(\theta, z) = (f(z)\cos\theta, f(z)\sin\theta, z), \qquad \theta \in (0, 2\pi), z \in \mathbb{R}$$

*parameterizes all of the surface except for the original generating curve. The curves of the form $\theta = const.$ are called **meridians**; this includes the original generating curve (where $\theta = 0$). Those curves with equations $z = const.$ are called **parallels**.*

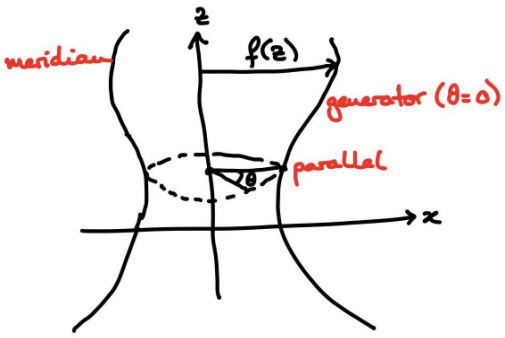


Figure 3.3 – surface of revolution

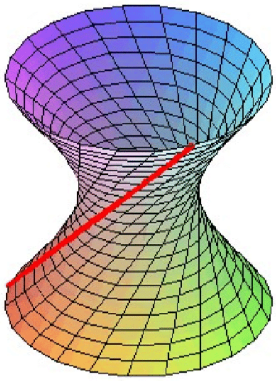


Figure 3.4 – hyperboloid of one sheet

Example 3.10 *(**Ruled surfaces**) Let $\gamma\colon I \to \mathbb{R}^3$ be a smooth curve in $\mathbb{R}^3$ and let $\mathbf{w}\colon I \to \mathbb{R}^3\backslash\{\mathbf{0}\}$ be a second non-vanishing vector function on I. Then the parameterized surface given by*

$$\mathbf{r}(u, v) = \gamma(u) + v\mathbf{w}(u) \quad u \in I, v \in \mathbb{R}$$

*is an example of a **ruled surface**. The curve γ is known as the directrix and the lines in the surface given by $u = constant$ are known as rulings.*

Note that the parameterization $\mathbf{r}$ above need not be a homeomorphism onto its image and so such a ruled surface may have self-intersections, although these may be avoided by limiting the domain of the co-ordinate v. For example, the image of the map

$$\mathbf{r}(u, v) = (v\cos u, v\sin u, v), \qquad u \in (0, 2\pi), v \in \mathbb{R},$$

is all of the two sheeted cone except for two rays (two halves of the line $x = z$). The map $\mathbf{r}$ is not a parameterization as $(0, 0, 0)$ is a self-intersection. However the restriction of $\mathbf{r}$ to $(0, 2\pi) \times (0, \infty)$ is a valid parameterization for the one sheeted cone with the omission of a single ruling.

Exercise 3.11 *Show that the hyperbolic paraboloid $z = xy$ and the hyperboloid of one sheet $x^2 + y^2 = z^2 + 1$ in $\mathbb{R}^3$ are ruled surfaces.*

3.1 The First Fundamental Form

Let $U \subseteq \mathbb{R}^2$ be an open subset of the plane and $\mathbf{r}\colon U \to \mathbb{R}^3$ be a parameterization of a smooth surface X. Let

$$\gamma\colon I \to X \quad \text{be given} \quad \gamma(t) = \mathbf{r}(u(t), v(t))$$

be a smooth curve lying in X.

Definition 3.12 *We define the* ***length*** *of γ to be*

$$\mathcal{L}(\gamma) = \int_I \left|\frac{\mathrm{d}\gamma}{\mathrm{d}t}\right| \mathrm{d}t. \tag{3.1}$$

Using the chain rule it is easy to see that the length of γ does not depend on the choice of parameter t. Now

$$\frac{\mathrm{d}\gamma}{\mathrm{d}t} = \frac{\mathrm{d}u}{\mathrm{d}t}\frac{\partial \mathbf{r}}{\partial u} + \frac{\mathrm{d}v}{\mathrm{d}t}\frac{\partial \mathbf{r}}{\partial v}$$

or written more concisely

$$\dot{\gamma} = \dot{u}\mathbf{r}_u + \dot{v}\mathbf{r}_v.$$

So the length of γ equals

$$\int_I \sqrt{E\dot{u}^2 + 2F\dot{u}\dot{v} + G\dot{v}^2}\,\mathrm{d}t \tag{3.2}$$

where

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

Definition 3.13 *The quadratic form $I_p\colon T_pX \to \mathbb{R}$ on the tangent space T_pX, defined by*

$$I\left({}_p\alpha\mathbf{r}_u + \beta\mathbf{r}_v\right) = E\alpha^2 + 2F\alpha\beta + G\beta^2$$

is called the ***first fundamental form*** *of X.*

Remark 3.14 *What does this actually mean? The first fundamental form is the restriction to T_pX of the quadratic form*

$$\mathbf{x} \mapsto |\mathbf{x}|^2.$$

Now $\{\mathbf{r}_u, \mathbf{r}_v\}$ is a basis for the tangent space and with respect to this basis the first fundamental form has coefficients $E, 2F$ and G. Geometrically it can be thought of as the square of the element of arc length, often conveyed as

$$\mathrm{d}s^2 = E\mathrm{d}u^2 + 2F\mathrm{d}u\mathrm{d}v + G\mathrm{d}v^2.$$

For $X = \mathbf{r}(U)$, a smooth parameterized surface, let

$$u\colon \mathbf{r}(u,v) \mapsto u \quad \textit{and} \quad v\colon \mathbf{r}(u,v) \mapsto v$$

denote the co-ordinate maps. For $p = \mathbf{r}(u_0, v_0)$, consider the differentials $\mathrm{d}u_p, \mathrm{d}v_p\colon T_pX \to \mathbb{R}$. We define two curves along the co-ordinate curves through p. Set

$$\begin{aligned} \gamma(t) &= \mathbf{r}(u_0 + t, v_0), \qquad t \in (-\epsilon, \epsilon), \\ \Gamma(t) &= \mathbf{r}(u_0, v_0 + t), \qquad t \in (-\epsilon, \epsilon). \end{aligned}$$

Note that $\gamma'(0) = \mathbf{r}_u(p)$ *and* $\Gamma'(0) = \mathbf{r}_v(p)$. *So*

$$\begin{aligned} \mathrm{d}u_p(\mathbf{r}_u) &= \mathrm{d}u_p(\gamma'(0)) = (u \circ \gamma)'(0) = (t \mapsto u_0 + t)'(0) = 1, \\ \mathrm{d}u_p(\mathbf{r}_v) &= \mathrm{d}u_p(\Gamma'(0)) = (u \circ \Gamma)'(0) = (t \mapsto u_0)'(0) = 0. \end{aligned}$$

Similarly $\mathrm{d}v_p(\mathbf{r}_u) = 0$ *and* $\mathrm{d}v_p(\mathbf{r}_v) = 1$. *So* $\mathrm{d}u_p$ *and* $\mathrm{d}v_p$ *are elements of the dual tangent space* T_pX; *in fact they are the dual basis of* $\{\mathbf{r}_u(p), \mathbf{r}_v(p)\}$. *So* $E\mathrm{d}u_p^2 + 2F\mathrm{d}u_p\mathrm{d}v_p + G\mathrm{d}v_p^2$ *is the quadratic form on* T_pX *given by*

$$I_p \colon \alpha\mathbf{r}_u + \beta\mathbf{r}_v \mapsto E\alpha^2 + 2F\alpha\beta + G\beta^2.$$

However one thinks about the first fundamental form, remember that the form is associated with the surface. When we change co-ordinates the quadratic form does not change, but its expression will generally look different in terms of the new co-ordinates.

Example 3.15 *Find the first fundamental form of the plane using (a) Cartesian co-ordinates and (b) polar co-ordinates.*

Solution. Using Cartesian co-ordinates we find

$$\mathbf{r}(u, v) = (u, v), \qquad u, v \in \mathbb{R}$$

and with polar co-ordinates

$$\mathbf{R}(r, \theta) = (r\cos\theta, r\sin\theta), \qquad r > 0, \theta \in (0, 2\pi).$$

So

$$\begin{aligned} \mathbf{r}_u &= (1, 0) & \mathbf{r}_v &= (0, 1), \\ \mathbf{R}_r &= (\cos\theta, \sin\theta), & \mathbf{R}_\theta &= (-r\sin\theta, r\cos\theta). \end{aligned}$$

With respect to the two co-ordinate systems the first fundamental form is:

$$\mathrm{d}u^2 + \mathrm{d}v^2 \qquad \text{and} \qquad \mathrm{d}r^2 + r^2\mathrm{d}\theta^2.$$

■

Remark 3.16 *It is always possible to introduce local co-ordinates such that the first fundamental form has certain preferential forms.*

- *[Do Carmo, p.183] There exists a local parameterization around any point of a surface such that* $F = 0$. *Such a parameterization is called* ***orthogonal****.*
- *[Do Carmo, p.227] There exists a local parameterization around any point of a surface such that* $F = 0$ *and* $E = G$. *Such a parameterization is called* ***isothermal****. This is equivalent to the parameterization being conformal from the plane; the existence of isothermal co-ordinates implies all smooth surfaces are locally conformal.*
- *[Do Carmo, p.287] Using geodesic polar co-ordinates, it is possible to parameterize a surface locally such that* $E = 1$ *and* $F = 0$.

The following argument was previously given in Prelims Geometry as a definition for area. Let $V \subseteq U$ be an open subset of U; we wish to calculate the area of $\mathbf{r}(V)$. Consider a small parallelogram with vertices

$$\mathbf{r}(u,v), \quad \mathbf{r}(u+\delta u, v), \quad \mathbf{r}(u, v+\delta v), \quad \mathbf{r}(u+\delta u, v+\delta v).$$

Now

$$\mathbf{r}(u+\delta u, v) - \mathbf{r}(u,v) = \mathbf{r}_u(u,v)\delta u + O(\delta u^2)$$

and there is a similar expression for varying v. So the area of the parallelogram is, ignoring higher order terms,

$$|\mathbf{r}_u \wedge \mathbf{r}_v|\, \delta u\, \delta v.$$

It thus seems reasonable to define:

Definition 3.17 *The **area** of $\mathbf{r}(V)$ equals*

$$\iint_V |\mathbf{r}_u \wedge \mathbf{r}_v|\, \mathrm{d}u\, \mathrm{d}v. \tag{3.3}$$

Now

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

Thus the expression (3.3) for the area of $\mathbf{r}(V)$ can be rewritten as

$$\iint_V \sqrt{EG - F^2}\, \mathrm{d}u\, \mathrm{d}v. \tag{3.4}$$

See Proposition 0.11 for a proof that this definition is independent of the choice of co-ordinates.

Example 3.18 *Show that the area of a sphere of radius a equals $4\pi a^2$.*

Solution. We may parameterize the sphere using spherical polar co-ordinates

$$\mathbf{r}(u,v) = (a\cos u \sin v, a \sin u \sin v, a \cos v), \qquad u \in (-\pi, \pi),\ v \in (0, \pi),$$

omitting only half a great circle. Then

$$\begin{aligned} \mathbf{r}_u &= (-a \sin u \sin v, a \cos u \sin v, 0), \\ \mathbf{r}_v &= (a \cos u \cos v, a \sin u \cos v, -a \sin v). \end{aligned}$$

Thus (with respect to the co-ordinates u and v) the first fundamental form is given by

$$E = a^2 \sin^2 v, \quad F = 0, \quad G = a^2$$

and the area is given by

$$\int_0^\pi \int_{-\pi}^{\pi} a^2 |\sin v|\, \mathrm{d}u\, \mathrm{d}v = 2\pi a^2 \int_0^\pi \sin v\, \mathrm{d}v = 4\pi a^2$$

as required. ■

Example 3.19 *The **tractoid** (see Figure 3.5) is the surface of revolution formed by rotating the curve*

$$x(t) = -(\cos t + \log \tan \frac{t}{2}), \qquad y(t) = \sin t, \qquad t \in (0, \pi/2)$$

*(known as the **tractrix**) about the x-axis.*
(a) Show that, when the tractrix is parameterized by arc-length s, the first fundamental form of the tractoid is

$$\mathrm{d}s^2 + e^{-2s}\mathrm{d}\theta^2. \tag{3.5}$$

(b) Show that the area of the tractoid equals 2π.

Solution. (a) We may parameterize the tractoid by writing

$$\mathbf{r}(t, \theta) = (x(t), y(t)\cos\theta, y(t)\sin\theta), \qquad t \in (0, \infty),\ \theta \in (0, 2\pi),$$

omitting only the original tractrix. Differentiating with respect to t and θ we find that

$$\begin{aligned} \mathbf{r}_t &= (-\cos t \cot t, \cos t \cos\theta, \cos t \sin\theta), \\ \mathbf{r}_\theta &= (0, -\sin t \sin\theta, \sin t \cos\theta). \end{aligned}$$

Thus the first fundamental form is given by

$$\cot^2 t\, \mathrm{d}t^2 + \sin^2 t\, \mathrm{d}\theta^2. \tag{3.6}$$

Now

$$\left(\frac{\mathrm{d}s}{\mathrm{d}t}\right)^2 = \left(\frac{\mathrm{d}x}{\mathrm{d}t}\right)^2 + \left(\frac{\mathrm{d}y}{\mathrm{d}t}\right)^2 = \left(\frac{\cos^2 t}{\sin t}\right)^2 + \cos^2 t = \cot^2 t \left(\cos^2 t + \sin^2 t\right) = \cot^2 t.$$

As s is decreasing with respect to t then $\mathrm{d}s/\mathrm{d}t = -\cot t$ and hence $s = -\log \sin t$. Substituting these expressions into (3.6) we obtain $E = 1, F = 0, G = e^{-2s}$ as in (3.5).

(b) The area of the tractoid is then given by the integral

$$\int_0^\infty \int_0^{2\pi} e^{-s}\, \mathrm{d}\theta\, \mathrm{d}s = 2\pi.$$

■

Exercise 3.20 *Show that the area of the torus in $\mathbb{R}^3$, given by*

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

for $u, v \in (0, 2\pi)$ and $a > b > 0$, equals $4\pi^2 ab$.

Properties of surfaces which depend solely on the first fundamental such as length and area (and geodesics and Gaussian curvature – see later) are called **intrinsic**. Maps between surfaces which preserve the intrinsic geometry are called *isometries*.

Definition 3.21 *An **isometry** between two surfaces X and Y is a diffeomorphism $f\colon X \to Y$ which maps curves in X to curves in Y of the same length. X and Y are then said to be isometric.*

As the first fundamental form represents an element of arc length then the following theorem should be intuitively clear.

Theorem 3.22 *Two smooth, parameterized surfaces X and Y are isometric if and only if there exists an open subset $U \subset \mathbb{R}^2$ and parameterizations*

$$\mathbf{r}\colon U \to X, \qquad \mathbf{s}\colon U \to Y,$$

such that the first fundamental forms of X and Y are the same.

Proof. Sufficiency is straightforward. Suppose two such parameterizations $\mathbf{r}$ and $\mathbf{s}$ exist with the same fundamental forms – I claim $f = \mathbf{s}\mathbf{r}^{-1}\colon X \to Y$ is the required isometry. Let C be a smooth curve in U. The lengths of $\mathbf{r}(C)$ and $\mathbf{s}(C) = f(\mathbf{r}(C))$ are identical as they are given by the same integral (3.2).

Conversely, suppose now that $f\colon X \to Y$ is an isometry of two smooth, parameterized surfaces and suppose that $\mathbf{r}\colon U \to X$ is a parameterization of X. Let $\mathbf{s} = f\mathbf{r}\colon U \to Y$. We shall write $E, 2F, G$ and $\tilde{E}, 2\tilde{F}, \tilde{G}$ for the coefficients of the first fundamental forms of X and Y with respect to $\mathbf{r}$ and $\mathbf{s}$. As f is an isometry we have that

$$\int_a^b \sqrt{E\dot{u}^2 + 2F\dot{u}\dot{v} + G\dot{v}^2}\,\mathrm{d}t = \int_a^b \sqrt{\tilde{E}\dot{u}^2 + 2\tilde{F}\dot{u}\dot{v} + \tilde{G}\dot{v}^2}\,\mathrm{d}t \tag{3.7}$$

for **all** smooth curves $(u(t), v(t))$, $a \leqslant t \leqslant b$, in U.

As the above is an identity for all b then

$$\sqrt{E\dot{u}^2 + 2F\dot{u}\dot{v} + G\dot{v}^2} = \sqrt{\tilde{E}\dot{u}^2 + 2\tilde{F}\dot{u}\dot{v} + \tilde{G}\dot{v}^2}$$

And choosing part of a co-ordinate curve, namely: $u(t) = u_0 + t$ and $v(t) = v_0$ in U, it follows that $E = \tilde{E}$. By a similar argument using $u = \text{const.}$ curves we may conclude that $G = \tilde{G}$. Finally then $F = \tilde{F}$. ■

Example 3.23 *The **catenoid** (with a meridian removed) and **helicoid** are respectively parameterized by*

$$\begin{aligned} \mathbf{r}(u, v) &= (u, \cosh u, \cos v, \cosh u \sin v), \qquad u \in \mathbb{R},\ v \in (0, 2\pi)\,, \\ \mathbf{s}(\tilde{u}, \tilde{v}) &= (\tilde{u}, \tilde{v}\cos\tilde{u}, \tilde{v}\sin\tilde{u}), \qquad \tilde{u} \in \mathbb{R},\ \tilde{v} \in \mathbb{R}. \end{aligned}$$

Show that the catenoid is isometric to part of the helicoid, in such a way that meridians of the catenoid map to rulings of the helicoid.

Solution. The first fundamental form of the catenoid equals

$$\cosh^2 u\,\mathrm{d}u^2 + \cosh^2 u\,\mathrm{d}v^2$$

and the first fundamental form of the helicoid equals

$$(1 + \tilde{v}^2)\,\mathrm{d}\tilde{u}^2 + \mathrm{d}\tilde{v}^2. \tag{3.8}$$

Now consider the map

$$\mathbf{r}(u,v) \mapsto \mathbf{s}(v, \sinh u), \quad \text{for} \quad u \in \mathbb{R}, v \in (0, 2\pi) \tag{3.9}$$

between the catenoid and the helicoid. Under the substitution $\tilde{u} = v$ and $\tilde{v} = \sinh u$ then the form (3.8) becomes

$$(1 + \sinh^2 u)\,\mathrm{d}v^2 + \mathrm{d}(\sinh u)^2 = \cosh^2 u\,\mathrm{d}u^2 + \cosh^2 u\,\mathrm{d}v^2$$

which is the first fundamental form of the catenoid. Thus the map (3.9) is indeed an isometry.

The meridians of the catenoid are given by the equations $v = \textit{constant}$. Under the above isometry the meridians map to the curves on the helicoid given by $\tilde{u} = \textit{constant}$ – i.e. the rulings. ■

Exercise 3.24 *(First part is Sheet 2, Part A, Exercise 1) Two curves on the same smooth parameterized surface are given parameterically by $t \mapsto (u(t), v(t))$ and $t \mapsto (\tilde{u}(t), \tilde{v}(t))$. Suppose that the curves intersect at $t = 0$. (i.e. $u(0) = \tilde{u}(0)$ and $v(0) = \tilde{v}(0)$.) Prove that the angle of intersection θ is given by*

$$\cos\theta = \frac{E\dot{u}\dot{\tilde{u}} + F(\dot{u}\dot{\tilde{v}} + \dot{\tilde{u}}\dot{v}) + G\dot{v}\dot{\tilde{v}}}{\sqrt{E\dot{u}^2 + 2F\dot{u}\dot{v} + G\dot{v}^2}\sqrt{E\dot{\tilde{u}}^2 + 2F\dot{\tilde{u}}\dot{\tilde{v}} + G\dot{\tilde{v}}^2}}$$

Deduce that a parameterization is conformal if and only if the first fundamental form satisfies $E = G$ and $F = 0$ everywhere.

Exercise 3.25 *A diffeomorphism between surfaces X and Y is said to be **conformal** if the angle between any two intersecting curves on X equals the angle between their images on Y and is said to be **area-preserving** if each subset of X is mapped to a subset of Y of equal area. Show that a diffeomorphism is an isometry if and only if it is area-preserving and conformal.*

Thus far we have not made any calculations of lengths and areas which couldn't have been done as easily with the old expressions (3.1), (3.3) as with the new expressions (3.2), (3.4) which are in terms of coefficients of the first fundamental form. The calculations in the following examples however can only be done using the new definitions of length and area.

Example 3.26 *The **flat torus** $\mathbb{T}$ is the surface in $\mathbb{R}^4$ given by*

$$\mathbb{T} = \{(x, y, z, t) \in \mathbb{R}^4 \mid x^2 + y^2 = z^2 + t^2 = 1\}.$$

Show that $\mathbb{T}$ is locally isometric to $\mathbb{R}^2$ and calculate the area of $\mathbb{T}$.

Solution. We may parameterize (a dense open subset of) $\mathbb{T}$ by

$$\mathbf{r}(u,v) = (\cos u, \sin u, \cos v, \sin v), \qquad u, v \in (0, 2\pi).$$

Then the first fundamental form of $\mathbb{T}$ is $\mathrm{d}u^2 + \mathrm{d}v^2$ and so $\mathbb{T}$ is locally isometric to the plane. $\mathbb{T}$ is certainly not globally isometric to $\mathbb{R}^2$ since $\mathbb{T}$ is compact and $\mathbb{R}^2$ is non-compact. (In fact the flat torus is isometric to no surface in $\mathbb{R}^3$ – see Sheet 2, Part B, Exercise 4.) The area of $\mathbb{T}$

is easily seen using (3.4) to equal $4\pi^2$ but as the vector product is not defined in $\mathbb{R}^4$ then our original definition (3.3) is not applicable. ■

So far we have only considered examples where the metric structure of the surface is precisely that induced on the surface by the Euclidean space (usually $\mathbb{R}^3$) in which the surface lies. There is no reason why we should limit ourselves to these cases – in fact there are good reasons not to.

From Example 3.19 the tractoid (with the original tractrix removed) has first fundamental form

$$\mathrm{d}s^2 + e^{-2s}\,\mathrm{d}\theta^2, \qquad s > 0,\ \theta \in (0, 2\pi),$$

when the tractrix is parameterized by arc-length s. The map f from the tractoid to $(0, 2\pi) \times (1, \infty)$ which sends the point on the tractoid with co-ordinates (s, θ) to (θ, e^s) is a diffeomorphism but is not an isometry. We could however ask:

Example 3.27 *In terms of the co-ordinates x and y, find the first fundamental form on $(0, 2\pi) \times (1, \infty)$ for which f is an isometry.*

Solution. The co-ordinates x and y are related to s and θ by

$$x = \theta, \quad \text{and} \quad y = e^s.$$

For f to be an isometry we need to endow $(0, 2\pi) \times (1, \infty)$ with the first fundamental form

$$\mathrm{d}s^2 + e^{-2s}\,\mathrm{d}\theta^2 = \mathrm{d}(\log y)^2 + \frac{1}{y^2}\mathrm{d}x^2 = \frac{\mathrm{d}x^2 + \mathrm{d}y^2}{y^2}.$$

■

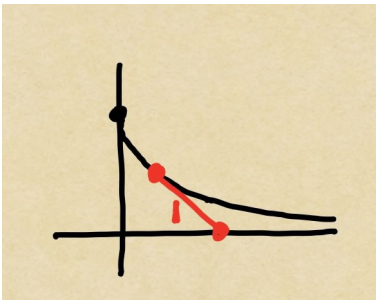


Figure 3.5 – tractrix

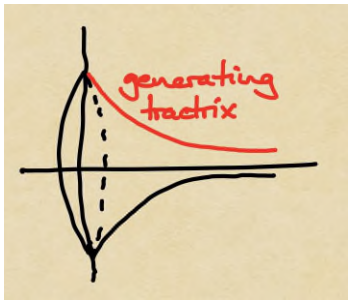


Figure 3.6 – tractoid

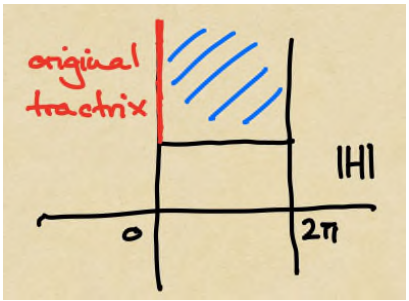


Figure 3.7 – tractoid as a subset of $\mathbb{H}$

What we have shown above is that the tractoid (without a meridian) is isometric to part of $\mathbb{H}$, the *hyperbolic plane* (Figure 3.7).

Example 3.28 ***(Poincaré's half plane model of the hyperbolic plane)*** *$\mathbb{H}$ is the surface created by endowing the upper half plane $\{(x, y) \mid y > 0\}$ with the first fundamental form*

$$\frac{\mathrm{d}x^2 + \mathrm{d}y^2}{y^2}. \tag{3.10}$$

$\mathbb{H}$ is of interest because it was the first model for a non-Euclidean geometry.

Whilst the infinite rectangle $(0, 2\pi) \times (1, \infty)$ with the first fundamental form (3.10) is isometric to a surface in $\mathbb{R}^3$, the hyperbolic plane is not. This is a consequence of *Hilbert's Theorem* (Do Carmo, p. 446). We could isometrically embed $\mathbb{H}$ in a higher dimensional Euclidean space, although the isometry may be a little complicated, but there is no need. From our formulas (3.2),(3.4) we may find the length and area of curves and regions in $\mathbb{H}$ without having to be working in a particular Euclidean space. Indeed we could create a *geometric* surface by endowing any open subset of $\mathbb{R}^2$ with any first fundamental form $E\mathrm{d}x^2 + 2F\mathrm{d}x\mathrm{d}y + G\mathrm{d}y^2$ provided that E, F, G are smooth functions and

$$E > 0, \quad G > 0, \quad EG - F^2 > 0.$$

Conversely any parameterized surface which is diffeomorphic to an open subset of $\mathbb{R}^2$ would be isometric to one of these surfaces.

Example 3.29 *Find the length of the curve $\gamma(t) = (0, t)$ for $1 \leqslant t \leqslant 2$ in $\mathbb{H}$.*

Solution. We have $E = G = y^{-2}$ and $F = 0$. Substituting these into (3.2) we find

$$\mathcal{L}(\gamma) = \int_1^2 \sqrt{\frac{1}{t^2}}\,\mathrm{d}t = [\log t]_1^2 = \log 2.$$

■

Exercise 3.30 *Show that the surfaces created by endowing $(0, \alpha) \times (0, \infty)$ with the first fundamental form (3.10) are isometric for any $\alpha > 0$.*

Definition 3.31 *A **smooth geometric surface** or **smooth Riemannian 2-manifold** is a Hausdorff topological space X together with*

(a) homeomorphisms $\phi_\alpha \colon U_\alpha \to V_\alpha$ between open sets $U_\alpha \subseteq X$ and open sets $V_\alpha \subseteq \mathbb{R}^2$,

(b) first fundamental forms $E_\alpha \mathrm{d}x^2 + 2F_\alpha \mathrm{d}x\mathrm{d}y + G_\alpha \mathrm{d}y^2$ on U_α where $E_\alpha, F_\alpha, G_\alpha$ are smooth functions satisfying

$$E_\alpha > 0, \quad G_\alpha > 0, \quad E_\alpha G_\alpha - (F_\alpha)^2 > 0,$$

such that

(a) $\bigcup_\alpha U_\alpha = X$,

(b) when $U_\alpha \cap U_\beta \neq \emptyset$ then

$$(\phi_\alpha) \circ \phi_\beta^{-1} \colon (\phi_\beta)(U_\alpha \cap U_\beta) \to (\phi_\alpha)(U_\alpha \cap U_\beta)$$

is an isometry.

Example 3.32 *(**The elliptic plane**) Topologically the elliptic plane is the real projective plane. Geometrically it is the surface endowed with the first fundamental form from the unit sphere.*

Let D denote the unit disc $\{(u, v) \mid u^2 + v^2 < 1\}$. Then $\mathbf{r}_1 \colon D \to S^2$, defined by

$$\mathbf{r}_1(u, v) = (u, v, \sqrt{1 - u^2 - v^2}),$$

is a parameterization of a unit hemisphere, so that $\mathbf{s}_1 = \pi \circ \mathbf{r}_1 \colon D \to S^2/\{\pm 1\} = \mathbb{P}$ *is a parameterization of (a dense open subset of) the real projective plane* $\mathbb{P}$*. The first fundamental form on* $\mathbf{r}_1(D)$ *is*

$$\frac{(1-v^2)\mathrm{d}u^2 + 2uv\mathrm{d}u\mathrm{d}v + (1-u^2)\mathrm{d}v^2}{1-u^2-v^2},$$

and we can endow $\mathbf{s}_1(D)$ *with this first fundamental form to form a geometric surface.*

A second parameterization $\mathbf{s}_2 = \pi \circ \mathbf{r}_2 \colon D \to \mathbb{P}$ *arises from the parameterization* $\mathbf{r}_2 \colon D \to S^2$ *given by*

$$\mathbf{r}_2(U,V) = (U, \sqrt{1-U^2-V^2}, V),$$

which is endowed with the same first fundamental form once u *is replaced with* U *and* v *with* V*. Now the transition map* $\mathbf{s}_2^{-1} \circ \mathbf{s}_1$ *is given by*

$$U(u,v) = u \qquad \text{and} \qquad V(u,v) = \sqrt{1-u^2-v^2}.$$

Substituting these values into the first fundamental form on $\mathbf{s}_2(D)$ *we note* $\mathrm{d}U = \mathrm{d}u$ *and*

$$\mathrm{d}V = \frac{-u\mathrm{d}u - v\mathrm{d}v}{\sqrt{1-u^2-v^2}}$$

and then

$$\begin{aligned}
&\frac{(1-V^2)\mathrm{d}U^2 + 2UV\mathrm{d}U\mathrm{d}V + (1-U^2)\mathrm{d}V^2}{1-U^2-V^2} \\
= \;& \frac{(u^2+v^2)\mathrm{d}u^2 - 2u\sqrt{1-u^2-v^2}\mathrm{d}u\left(\frac{u\mathrm{d}u+v\mathrm{d}v}{\sqrt{1-u^2-v^2}}\right) + (1-u^2)\left(\frac{u\mathrm{d}u+v\mathrm{d}v}{\sqrt{1-u^2-v^2}}\right)^2}{v^2} \\
= \;& \frac{(u^2+v^2)\mathrm{d}u^2 - 2u\mathrm{d}u\,(u\mathrm{d}u+v\mathrm{d}v) + \frac{(1-u^2)}{(1-u^2-v^2)}\,(u^2\mathrm{d}u^2 + 2uv\mathrm{d}u\mathrm{d}v + v^2\mathrm{d}v^2)}{v^2} \\
= \;& \frac{\left[(v^2-u^2)\,(1-u^2-v^2) + (1-u^2)\,u^2\right]\mathrm{d}u^2 + 2\left[-uv\,(1-u^2-v^2) + uv\,(1-u^2)\right]\mathrm{d}u\mathrm{d}v + (1-u^2)\,v^2\mathrm{d}v^2}{v^2\,(1-u^2-v^2)} \\
= \;& \frac{\left[v^2-v^4\right]\mathrm{d}u^2 + 2\left[uv^3\right]\mathrm{d}u\mathrm{d}v + \left[(1-u^2)\,v^2\right]\mathrm{d}v^2}{v^2\,(1-u^2-v^2)} \\
= \;& \frac{(1-v^2)\mathrm{d}u^2 + 2uv\mathrm{d}u\mathrm{d}v + (1-u^2)\mathrm{d}v^2}{1-u^2-v^2}.
\end{aligned}$$

Hence the transition map is an isometry as required, because one first fundamental form transforms in the other.

We can similarly extend the notion of orientability to abstract surfaces by requiring that transition maps are orientation-preserving as well as diffeomorphisms. It is even possible to define the tangent space for a smooth abstract surface, even though the surface is not embedded in any ambient Euclidean space. On an abstract surface we still have local co-ordinates, so it still possible to differentiate smooth functions with respect to those co-ordinates.

Definition 3.33 *Let* X *be a smooth abstract surface and* $p \in X$*. Let* V_p *denote the vector space (algebra, in fact) of all functions* $\varphi \colon X \to \mathbb{R}$ *which are smooth at* p*. Then the* ***tangent space*** *at* p*, written* T_pX*, is the set of all linear maps* $D \colon V_p \to \mathbb{R}$ *which satisfy the product rule*

$$D(\varphi\psi) = \varphi(p)\,D\psi + \psi(p)\,D\varphi \qquad \text{for all } \varphi, \psi \in V_p.$$

Such a D is called a ***derivation.*** *Note T_pX is a vector space with addition and scalar multiplication defined by*

$$(D_1 + D_2)\,\varphi = D_1\varphi + D_2\varphi, \qquad (\lambda D)\,\varphi = \lambda(D\varphi).$$

Given a smooth map $f\colon X \to Y$ between two smooth abstract surfaces X, Y with $p \in X$ the differential $\mathrm{d}f_p\colon T_pX \to T_{f(p)}Y$ is defined by

$$(\mathrm{d}f_p\,(D))\,(\alpha) = D(\alpha \circ f)$$

where $D \in T_pX$ and α is a real map $\alpha\colon Y \to \mathbb{R}$ which is smooth at $f(p)$.

Exercise 3.34 *T_pX is two dimensional and a basis is*

$$\left\{ \left.\frac{\partial}{\partial u}\right|_p, \left.\frac{\partial}{\partial v}\right|_p \right\}$$

where u and v are co-ordinates local to p.

3.2 Curvature and the Weingarten map

Let X be a smooth parameterized surface in $\mathbb{R}^3$ described by $\mathbf{r}\colon U \to X$ and let

$$\mathbf{n} = \frac{\mathbf{r}_u \wedge \mathbf{r}_v}{|\mathbf{r}_u \wedge \mathbf{r}_v|}$$

denote a choice of unit normal. When $\gamma(s)$ is a curve in X, parameterized by arc length then the curvature $\kappa(s)$ of γ at the point $\gamma(s)$ is simply the magnitude of $\ddot{\gamma}(s)$.

When looking at such a curve, the vector $\ddot{\gamma}(s)$ has two natural components, a tangential component and a normal component. As $\dot{\gamma}(s)$ is a unit vector for all s, its derivative $\ddot{\gamma}(s)$ is perpendicular to $\dot{\gamma}(s)$. So we may decompose $\ddot{\gamma}(s)$ in the form:

$$\ddot{\gamma} = k_n\mathbf{n} + k_g(\mathbf{n} \wedge \dot{\gamma}). \tag{3.11}$$

Definition 3.35 *We define:*
(a) $k_n(s)$ is the ***normal curvature*** *of γ at $\gamma(s)$.*
(b) $k_g(s)$ is the ***geodesic curvature*** *of γ at $\gamma(s)$.*
It follows that $\kappa^2 = \left|\dot{\mathbf{t}}\right|^2 = |\ddot{\gamma}|^2 = k_n^2 + k_g^2$.
(c) A curve in X whose geodesic curvature is everywhere zero is called a ***geodesic.***

We shall consider, for the moment, the normal curvature of curves and we shall use this to define a second quadratic form on the tangent space of a point of X. We shall see later (Theorem 4.2 and Sheet 2, Part B, Exercise 3) that the geodesics of a surface and the geodesic curvature of a curve are intrinsic; that is they depend only on the first fundamental form of the surface and the direction of the curve. This is very much not the case with normal curvature, which gives information on how a geometric surface has been embedded in $\mathbb{R}^3$.

The normal curvature k_n of γ equals $\ddot{\gamma} \cdot \mathbf{n}$. By the chain rule we have

$$\dot{\gamma} = \dot{u}\mathbf{r}_u + \dot{v}\mathbf{r}_v,$$

and applying the chain rule again we find

$$\ddot{\gamma} = \ddot{u}\mathbf{r}_u + \ddot{v}\mathbf{r}_v + \dot{u}^2\mathbf{r}_{uu} + 2\dot{u}\dot{v}\mathbf{r}_{uv} + \dot{v}^2\mathbf{r}_{vv}.$$

Hence the normal curvature $k_n = \ddot{\gamma} \cdot \mathbf{n}$ equals

$$k_n = L\dot{u}^2 + 2M\dot{u}\dot{v} + N\dot{v}^2,$$

where

$$\begin{aligned} L &= \mathbf{r}_{uu} \cdot \mathbf{n} = -\mathbf{r}_u \cdot \mathbf{n}_u, \\ M &= \mathbf{r}_{uv} \cdot \mathbf{n} = -\mathbf{r}_u \cdot \mathbf{n}_v = -\mathbf{r}_v \cdot \mathbf{n}_u, \\ N &= \mathbf{r}_{vv} \cdot \mathbf{n} = -\mathbf{r}_v \cdot \mathbf{n}_v. \end{aligned} \tag{3.12}$$

Note that the alternative expressions for L, M, N come from the differentiating the equations

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

Definition 3.36 *The quadratic form $II_p \colon T_pX \to \mathbb{R}$ given by*

$$\alpha\mathbf{r}_u + \beta\mathbf{r}_v \mapsto L\alpha^2 + 2M\alpha\beta + N\beta^2$$

*is called the **second fundamental form** of X. (Note that some authors, including Do Carmo, use e, f, g instead of L, M, N for the coefficients of the second fundamental form.)*

The first fundamental form describes the intrinsic properties of the surface, whereas the second fundamental form relates to the surface's embedding in $\mathbb{R}^3$. Although the proof of the following theorem is far beyond the scope of this course, I include an abridged statement of:

Theorem 3.37 *(Off-syllabus)* ***The Fundamental Theorem of the Local Theory of Surfaces.***

Let E, F, G, L, M, N be differentiable functions on an open set $U \subset \mathbb{R}^2$ which satisfy
(a) $E > 0, G > 0, EG - F^2 > 0$,
(b) certain compatibility equations (Remark 3.56, Do Carmo p.235).

Then for each $p \in U$ there is an open set $V \subset U$ containing p and a smooth parameterization $\mathbf{r}(V)$ of a surface in $\mathbb{R}^3$ with $E, 2F, G$ and $L, 2M, N$ as the coefficients of the first and second fundamental forms. Further a second surface $\tilde{\mathbf{r}}(V)$ in $\mathbb{R}^3$ with the same first and second fundamental forms differs from $\mathbf{r}(V)$ only by a rigid motion of $\mathbb{R}^3$.

One equation of compatability is the Gauss formula, which we will meet shortly, and which expresses the Gaussian curvature – ostensibly defined in terms of both fundamental forms – solely in terms of the first fundamental form.

Example 3.38 *Show that the xy-plane and cylinder $x^2 + y^2 = a^2$ are locally isometric but have different second fundamental forms.*

Solution. A parameterization of the xy-plane is $\mathbf{r}(u, v) = (u, v, 0)$ which leads to

$$E = 1, \quad F = 0, \quad G = 1, \quad L = 0, \quad M = 0, \quad N = 0.$$

The cylinder, except for one meridian, can be parameterized by $\mathbf{s}(u,v) = (a\cos(u/a), a\sin(u/a), v)$ where $0 < u < 2\pi a$ and $v \in \mathbb{R}$. This leads to

$$E = 1, \quad F = 0, \quad G = 1, \quad L = -a^{-1}, \quad M = 0, \quad N = 0.$$

Thus the cylinder and plane are locally isometric. They are not globally isometric as they are not homeomorphic – the cylinder is not simply connected whereas the plane is. ■

In order to define the curvature of the surface at a point we need to introduce the *Weingarten map* or *shape operator*. The Weingarten map is the differential of the Gauss (normal) map $\mathbf{n}$ and consequently is written as $\mathrm{d}\mathbf{n}_p$ in some texts. Curvature, for a curve, is a measure of how quickly the tangent is varying. Similarly for a surface we need to investigate how quickly the tangent plane, or equivalently the normal to the surface is varying. Note that as $\mathbf{n}\cdot\mathbf{n} = 1$ then

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

Thus $\mathbf{n}_u$ and $\mathbf{n}_v$ are tangents vectors to the surface.

Definition 3.39 *The* ***Weingarten map*** *(or* ***shape operator****) at the point p is the linear map $W_p\colon T_pX \to T_pX$ defined by*

$$W_p\mathbf{r}_u = \mathbf{n}_u, \qquad W_p\mathbf{r}_v = \mathbf{n}_v. \tag{3.13}$$

More generally note that $W_p(\gamma'(s)) = (\mathbf{n}\circ\gamma)'(s)$ and so $W_p = \mathrm{d}\mathbf{n}_p$ is the differential of the Gauss map.

Proposition 3.40 *The Weingarten map $W_p\colon T_pX \to T_pX$ is a self-adjoint linear map independent of the choice of parameters u and v. In particular, as W_p is self-adjoint, it is diagonalisable.*

Proof. Let $\mathbf{s}(\tilde{u},\tilde{v})$ be a second parameterization for X with $\mathbf{s}(\tilde{u},\tilde{v}) = \mathbf{r}(u,v)$. Then by the chain rule we have

$$\mathbf{s}_{\tilde{u}} = \frac{\partial u}{\partial\tilde{u}}\mathbf{r}_u + \frac{\partial v}{\partial\tilde{u}}\mathbf{r}_v, \qquad \mathbf{s}_{\tilde{v}} = \frac{\partial u}{\partial\tilde{v}}\mathbf{r}_u + \frac{\partial v}{\partial\tilde{v}}\mathbf{r}_v.$$

Hence by the above definition of the Weingarten map and the chain rule we have

$$W_p\mathbf{s}_{\tilde{u}} = \frac{\partial u}{\partial\tilde{u}}\mathbf{n}_u + \frac{\partial v}{\partial\tilde{u}}\mathbf{n}_v = \mathbf{n}_{\tilde{u}}, \qquad W_p\mathbf{s}_{\tilde{v}} = \frac{\partial u}{\partial\tilde{v}}\mathbf{n}_u + \frac{\partial v}{\partial\tilde{v}}\mathbf{n}_v = \mathbf{n}_{\tilde{v}}.$$

It is also easy to check that W_p is a self-adjoint linear map – that is

$$(W_p\mathbf{x})\cdot\mathbf{y} = \mathbf{x}\cdot(W_p\mathbf{y}) \tag{3.14}$$

for any two tangent vectors $\mathbf{x},\mathbf{y} \in T_pX$. We note from equation (3.12) that

$$W_p\mathbf{r}_u\cdot\mathbf{r}_v = \mathbf{n}_u\cdot\mathbf{r}_v = \mathbf{n}_v\cdot\mathbf{r}_u = W_p\mathbf{r}_v\cdot\mathbf{r}_u.$$

Equation (3.14) then follows for all tangent vectors $\mathbf{x},\mathbf{y}$ by linearity. ■

As W_p is self-adjoint it is diagonalizable and has real eigenvalues. Let γ be a curve in X with $\gamma(0) = p$. Then

$$\begin{aligned} W_p(\gamma'(0)) \cdot \gamma'(0) &= \mathbf{n}'(\gamma(0)) \cdot \gamma'(0) \\ &= -\mathbf{n} \cdot \gamma''(0) = -k_n. \end{aligned}$$

Thus the eigenvalues of W_p are $-k_1$ and $-k_2$ where k_1 and k_2 are the extreme values of the normal curvature, called the **principal curvatures** of X at p and the eigenvectors of W_p are the **principal directions**. The **lines of curvature** are curves whose tangents are the principal directions.

We make the following definitions:

Definition 3.41 *The **Gaussian curvature** $K(p)$ at the point p is the product of the principal curvatures or equivalently the determinant $\det W_p$ of the Weingarten map.*

Definition 3.42 *(Off-syllabus) The average of the principal curvatures is known as the **mean curvature** at p. It is given by the formula*

$$H = \frac{LG - 2MF + NE}{2(EG - F^2)}.$$

The mean curvature is important in the study of minimal surfaces. A minimal surface is a surface with an area that is a local minimum, such as with soap films. A soap film – in order to reduce the surface tension – has minimal area compared with all perturbations of the surface. This is equivalent to the mean curvature of the surface being zero (Segal, Theorem 9.1).

The tangent vectors $\mathbf{r}_u$ and $\mathbf{r}_v$ form a basis for the tangent plane T_pX and $W_p \colon T_pX \to T_pX$ is a linear map. We now work out the matrix for W_p with respect to this basis.

Let us suppose that the matrix for W_p with respect to the basis $\{\mathbf{r}_u, \mathbf{r}_v\}$ is

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix}.$$

Then

$$W_p\mathbf{r}_u = \mathbf{n}_u = A\mathbf{r}_u + C\mathbf{r}_v, \tag{3.15}$$

$$W_p\mathbf{r}_v = \mathbf{n}_v = B\mathbf{r}_u + D\mathbf{r}_v. \tag{3.16}$$

Dotting equation (3.15) with $\mathbf{r}_u$ and with $\mathbf{r}_v$ we find

$$-L = AE + CF, \qquad -M = AF + CG.$$

Doing the same for equation (3.16) we obtain

$$-M = BE + DF, \qquad -N = BF + DG.$$

Putting these equations into matrix form gives

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

and hence with respect to the basis $\{\mathbf{r}_u, \mathbf{r}_v\}$

$$W_p = \frac{1}{EG - F^2} \begin{pmatrix} -G & F \\ F & -E \end{pmatrix} \begin{pmatrix} L & M \\ M & N \end{pmatrix}. \tag{3.17}$$

Corollary 3.43 *The Gaussian curvature $K(p)$ at p, which equals $\det W_p$, is given by the formula*

$$K(p) = \frac{LN - M^2}{EG - F^2}.$$

Despite the above expression for K, which is in terms of the coefficients of the first *and* second fundamental forms, the Gaussian curvature may be written solely in terms of the coefficients of the first fundamental form and is invariant under isometries. This is a theorem due to Gauss and known as the *Theorema Egregium* (Theorem 3.48).

Gauss originally did not define K by the above formula but rather as the following more intuitive limit. Let U be a small open subset of X about the point p. Then if we let the area of U tend to zero (see Sheet 3, Part B, Exercise 3)

$$|K| = \lim_{\text{Area}(U)\to 0} \frac{\text{Area}(\mathbf{n}(U))}{\text{Area}(U)}.$$

The more 'curved' the surface at a point, the greater the variety in the normal vectors about the point.

We end this section with two worked examples – we continue with the earlier examples – the sphere and the tractoid – where we calculated the first fundamental form.

Example 3.44 *Find the Gaussian curvature of a sphere of radius a.*

Solution. In Example 3.18 we parameterized the sphere with

$$\mathbf{r}(u, v) = (a\cos u \sin v, a \sin u \sin v, a \cos v), \qquad u \in (-\pi, \pi),\ v \in (0, \pi),$$

omitting only half a great circle and found

$$E = a^2 \sin^2 v, \quad F = 0, \quad G = a^2.$$

The outward-pointing unit normal equals

$$\mathbf{n}(u, v) = (\cos u \sin v, \sin u \sin v, \cos v) = \frac{1}{a}\mathbf{r}(u, v).$$

So we can avoid further calculation by noting

$$\begin{aligned} L &= \mathbf{r}_{uu} \cdot \mathbf{n} = -\mathbf{r}_u \cdot \mathbf{n}_u = -\frac{1}{a}\mathbf{r}_u \cdot \mathbf{r}_u = -\frac{E}{a} = -a\sin^2 v; \\ M &= \mathbf{r}_{uv} \cdot \mathbf{n} = -\mathbf{r}_u \cdot \mathbf{n}_v = -\frac{1}{a}\mathbf{r}_u \cdot \mathbf{r}_v = -\frac{F}{a} = 0; \\ N &= \mathbf{r}_{vv} \cdot \mathbf{n} = -\mathbf{r}_v \cdot \mathbf{n}_v = -\frac{1}{a}\mathbf{r}_v \cdot \mathbf{r}_v = -\frac{G}{a} = -a. \end{aligned}$$

Hence

$$K = \frac{LN - M^2}{EG - F^2} = \frac{a^2 \sin^2 v}{a^4 \sin^2 v} = \frac{1}{a^2}.$$

■

Example 3.45 *A torus of revolution is formed by rotating the circle with equation*

$$(x-b)^2+y^2=a^2, \qquad (b>a),$$

about the y-axis. Parameterize the torus and find its Gaussian curvature.

Solution. We can parametrize (an open dense subset of) the torus as

$$\mathbf{r}(u,v)=((b+a\sin u)\cos v,\ (b+a\sin u)\sin v, a\cos u) \qquad 0<u,v<2\pi.$$

We have

$$\mathbf{r}_u=(a\cos u\cos v, a\cos u\sin v, -a\sin u), \qquad \mathbf{r}_v=(-(b+a\sin u)\sin v,(b+a\sin u)\cos v,0)$$

giving

$$\begin{aligned} E &= a^2\cos^2 u\left(\cos^2 v+\sin^2 v\right)+a^2\sin^2 u=a^2,\\ F &= a(b+a\sin u)(-\cos u\sin v\cos v+\cos v\cos u\sin v)=0,\\ G &= (b+a\sin u)\left(\sin^2 v+\cos^2 v\right)=b+a\sin u. \end{aligned}$$

Further

$$\begin{aligned} \mathbf{r}_u\wedge\mathbf{r}_v &= \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k}\\ a\cos u\cos v & a\cos u\sin v & -a\sin u\\ -(b+a\sin u)\sin v & (b+a\sin u)\cos v & 0 \end{vmatrix}\\ &= a(b+a\sin u)\begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k}\\ \cos u\cos v & \cos u\sin v & -\sin u\\ -\sin v & \cos v & 0 \end{vmatrix}\\ &= a(b+a\sin u)(\sin u\cos v,\sin y\sin v,\cos u), \end{aligned}$$

giving

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

We then have

$$\begin{aligned} \mathbf{r}_{uu} &= (-a\sin u\cos v,-a\sin u\sin v,-a\cos u),\\ \mathbf{r}_{uv} &= (-a\cos u\sin v,a\cos u\cos v,0),\\ \mathbf{r}_{vv} &= (-(b+a\sin u)\cos v,-(b+a\sin u)\sin v,0), \end{aligned}$$

and so

$$\begin{aligned} L &= -a\sin^2 u\left(\cos^2 v+\sin^2 v\right)-a\cos^2 u=-a,\\ M &= -a\cos u\sin v\sin u\cos v+a\cos u\sin v\sin u\cos v=0,\\ N &= -(b+a\sin u)\sin u\left(\cos^2 v+\sin^2 v\right)=-(b+a\sin u)\sin u. \end{aligned}$$

Hence

$$K=\frac{LN-M^2}{EG-F^2}=\frac{a(b+a\sin u)\sin u}{a^2(b+a\sin u)}=\frac{1}{a}\sin u.$$

Note that $K>0$ on the outside of the torus when $0<u<\pi$ and $K<0$ when $\pi<u<2\pi$. ■

Remark 3.46 *(**Parity of Gaussian curvature**) The sign of Gaussian curvature can be readily appreciated. If we choose an outward pointing normal in the example of the torus, on the outside of the outside of the torus the lines of curvature are both bending away from the normal, the principal curvatures are negative and their product K is positive. If we had instead had an inward pointing normal then the principal curvatures would still have had the same sign and $K > 0$ would still be true. On the inside of the torus, one line of curvature is around the hole of the torus and one through the hole of the torus. The principal curvatures have different signs and then $K < 0$.*

Exercise 3.47 *Find the lines of curvature and the principal curvatures on a surface of revolution in terms of the distance ρ of the generating curve from the axis. Show that the Gaussian curvature K equals $\kappa \cos\phi/\rho$ where κ is the curvature of the generating curve and ϕ is the angle between the axis and the tangent line to the curve.*

3.3 Theorema Egregium

Theorem 3.48 *(**Theorema Egregium**, Gauss, 1827) Gaussian curvature is intrinsic, and so preserved by isometries.*

Remark 3.49 *Recall that we define Gaussian curvature as*

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

The first fundamental form is intrinsic but the second fundamental form is not (as we saw earlier with Example 3.38). Hence there is no reason to expect that K is intrinsic. The Latin title 'Theorema Egregium' translates as 'remarkable theorem'.

Proof. Let $(u, v) \mapsto \mathbf{r}(u, v)$ be a parameterization for a patch of surface X in $\mathbb{R}^3$. Let $\mathbf{n}$ be a unit normal vector field on X and let the first and second fundamental forms respectively be

$$E\mathrm{d}u^2 + 2F\mathrm{d}u\mathrm{d}v + G\mathrm{d}v^2 \quad \text{and} \quad L\mathrm{d}u^2 + 2M\mathrm{d}u\mathrm{d}v + N\mathrm{d}v^2.$$

And recall the Weingarten map $W = \mathrm{d}\mathbf{n}$ (equation (3.17)) is represented by the matrix

$$\begin{pmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \end{pmatrix} = -\begin{pmatrix} E & F \\ F & G \end{pmatrix}^{-1} \begin{pmatrix} L & M \\ M & N \end{pmatrix}$$

with respect to the basis $\{\mathbf{r}_u, \mathbf{r}_v\}$ for the tangent space.

We introduce now the **Christoffel symbols** Γ^k_{ij}, defined by writing

$$\begin{aligned} \mathbf{r}_{uu} &= \Gamma^1_{11}\mathbf{r}_u + \Gamma^2_{11}\mathbf{r}_v + L\mathbf{n}, \\ \mathbf{r}_{uv} &= \Gamma^1_{12}\mathbf{r}_u + \Gamma^2_{12}\mathbf{r}_v + M\mathbf{n}, \\ \mathbf{r}_{vv} &= \Gamma^1_{22}\mathbf{r}_u + \Gamma^2_{22}\mathbf{r}_v + N\mathbf{n}. \end{aligned}$$

Our aim will be first to show that the Christoffel symbols are intrinsic – that is they depend only on E, F and G and their derivatives – and then show that the Gaussian curvature can be written in terms of the Christoffel symbols. ■

Lemma 3.50 *The Christoffel symbols depend only on E, F and G and their derivatives.*

Proof. Dotting the equations above with $\mathbf{r}_u$ and $\mathbf{r}_v$ we find

$$\begin{cases} \Gamma^1_{11}E + \Gamma^2_{11}F &=& \mathbf{r}_{uu}\cdot\mathbf{r}_u &=& \frac{1}{2}(\mathbf{r}_u\cdot\mathbf{r}_u)_u &=& \frac{1}{2}E_u, \\ \Gamma^1_{11}F + \Gamma^2_{11}G &=& \mathbf{r}_{uu}\cdot\mathbf{r}_v &=& (\mathbf{r}_u\cdot\mathbf{r}_v)_u - \frac{1}{2}(\mathbf{r}_u\cdot\mathbf{r}_u)_v &=& F_u - \frac{1}{2}E_v, \\ \Gamma^1_{12}E + \Gamma^2_{12}F &=& \mathbf{r}_{uv}\cdot\mathbf{r}_u &=& \frac{1}{2}(\mathbf{r}_u\cdot\mathbf{r}_u)_v &=& \frac{1}{2}E_v, \\ \Gamma^1_{12}F + \Gamma^2_{12}G &=& \mathbf{r}_{uv}\cdot\mathbf{r}_v &=& \frac{1}{2}(\mathbf{r}_v\cdot\mathbf{r}_v)_u &=& \frac{1}{2}G_u, \\ \Gamma^1_{22}E + \Gamma^2_{22}F &=& \mathbf{r}_{vv}\cdot\mathbf{r}_u &=& (\mathbf{r}_u\cdot\mathbf{r}_v)_v - \frac{1}{2}(\mathbf{r}_v\cdot\mathbf{r}_v)_u &=& F_v - \frac{1}{2}G_u, \\ \Gamma^1_{22}F + \Gamma^2_{22}G &=& \mathbf{r}_{vv}\cdot\mathbf{r}_v &=& \frac{1}{2}(\mathbf{r}_v\cdot\mathbf{r}_v)_v &=& \frac{1}{2}G_v. \end{cases}$$

Each of the braced equations are invertible as the determinant $EG-F^2$ is non-zero. Thus each Christoffel symbol may be written in terms of E, F, G and their derivatives. ■

Corollary 3.51 *Suppose that the parameterization* $\mathbf{r}$ *is orthogonal, that is* $F = 0$*. Then:*

$$\Gamma^1_{11} = E_u/2E,\ \Gamma^1_{12} = E_v/2E,\ \Gamma^1_{22} = -G_u/2E,$$

$$\Gamma^2_{11} = -E_v/2G,\ \Gamma^2_{12} = G_u/2G,\ \Gamma^2_{22} = G_v/2G.$$

Lemma 3.52 ***(The Gauss formula)***

$$(\Gamma^2_{12})_u - (\Gamma^2_{11})_v + \Gamma^1_{12}\Gamma^2_{11} + \Gamma^2_{12}\Gamma^2_{12} - \Gamma^2_{11}\Gamma^2_{22} - \Gamma^1_{11}\Gamma^2_{12} = -EK,$$

where K denotes the Gaussian curvature.

Proof. Note by the product rule that

$$(\mathbf{r}_{uu})_v = \Gamma^1_{11}\mathbf{r}_{uv} + (\Gamma^1_{11})_v\mathbf{r}_u + \Gamma^2_{11}\mathbf{r}_{vv} + (\Gamma^2_{11})_v\mathbf{r}_v + L\mathbf{n}_v + L_v\mathbf{n},$$

and that

$$(\mathbf{r}_{uv})_u = \Gamma^1_{12}\mathbf{r}_{uu} + (\Gamma^1_{12})_u\mathbf{r}_u + \Gamma^2_{12}\mathbf{r}_{uv} + (\Gamma^2_{12})_u\mathbf{r}_v + M\mathbf{n}_u + M_u\mathbf{n}.$$

We may write $(\mathbf{r}_{uu})_v$ and $(\mathbf{r}_{uv})_u$ in terms of the basis $\{\mathbf{r}_u, \mathbf{r}_v, \mathbf{n}\}$. By comparing the coefficients of $\mathbf{r}_v$ in these expressions we obtain

$$\Gamma^1_{11}\Gamma^2_{12} + \Gamma^2_{11}\Gamma^2_{22} + (\Gamma^2_{11})_v + Lw_{22} = \Gamma^1_{12}\Gamma^2_{11} + \Gamma^2_{12}\Gamma^2_{12} + (\Gamma^2_{12})_u + Mw_{21}.$$

Hence

$$\begin{aligned} & (\Gamma^2_{12})_u - (\Gamma^2_{11})_v + \Gamma^1_{12}\Gamma^2_{11} + \Gamma^2_{12}\Gamma^2_{12} - \Gamma^2_{11}\Gamma^2_{22} - \Gamma^1_{11}\Gamma^2_{12} \\ = \quad & Lw_{22} - Mw_{21} \\ = \quad & \frac{L(FM-EN) - M(LF-EM)}{EG-F^2} \\ = \quad & -EK. \end{aligned}$$

These two lemmas prove our claims. The Christoffel symbols are intrinsic, so by the Gauss formula K is also intrinsic. ■

Corollary 3.53 *When $F = 0$ the Gaussian curvature K equals*

$$K = \frac{-1}{2\sqrt{EG}} \left\{ \left(\frac{E_v}{\sqrt{EG}} \right)_v + \left(\frac{G_u}{\sqrt{EG}} \right)_u \right\}.$$

Solution. From the Gauss formula, and the above formulae for the Christoffel symbols when $F = 0$, we have that $-EK$ equals

$$\begin{aligned} & \left(\frac{G_u}{2G} \right)_u + \left(\frac{E_v}{2G} \right)_v - \frac{E_v^2}{4EG} + \frac{G_u^2}{4G^2} + \frac{E_v G_v}{4G^2} - \frac{E_u G_u}{4EG} \\ = \; & \frac{GG_{uu} - G_u^2}{2G^2} + \frac{GE_{vv} - E_v G_v}{2G^2} - \frac{E_v^2}{4EG} + \frac{G_u^2}{4G^2} + \frac{E_v G_v}{4G^2} - \frac{E_u G_u}{4EG} \\ = \; & \frac{G_{uu}}{2G} + \frac{E_{vv}}{2G} - \frac{E_v^2}{4EG} - \frac{G_u^2}{4G^2} - \frac{E_v G_v}{4G^2} - \frac{E_u G_u}{4EG}. \end{aligned}$$

Hence $-2\sqrt{EG}K$ equals

$$\left(\frac{E_{vv}}{\sqrt{EG}} - \frac{E_v}{2EG\sqrt{EG}} (E_v G + G_v E) \right) + \left(\frac{G_{uu}}{\sqrt{EG}} - \frac{G_u}{2EG\sqrt{EG}} (G_u E + E_u G) \right).$$

to give the required result. ■

Example 3.54 *The (Poincaré half-plane model for the)* ***hyperbolic plane*** *is the half-plane $\mathbb{H} = \{(u, v) \in \mathbb{R}^2 : v > 0\}$ with the first fundamental form $E = G = v^{-2}$ and $F = 0$. Find the Gaussian curvature of $\mathbb{H}$.*

Solution.

$$K = \frac{-v^2}{2} \frac{\mathrm{d}}{\mathrm{d}v} \left(\frac{-2}{v} \right) = -1.$$

■

Example 3.55 *Show that there exists no surface $\mathbf{r}(u, v)$ with first and second fundamental forms respectively*

$$\mathrm{d}u^2 + \mathrm{d}v^2 \quad \textit{and} \quad \mathrm{d}u^2 - \mathrm{d}v^2.$$

Solution. On the one hand the surface has Gaussian curvature everywhere -1 as $E = G = L = 1, F = M = 0, N = -1$. On the other the surface is isometric to a subset of the plane and hence has Gaussian curvature 0. ■

Remark 3.56 *(**Equations of compatability**) The Gauss formula is a necessary condition connecting the coefficients of the first and second fundamental forms – it is one of the equations of compatability that form part of the fundamental theorem (Theorem 3.37). The other equations are called the Mainardi-Codazzi equations and they require*

$$\begin{aligned} L_v - M_u &= L\Gamma_{12}^1 + M(\Gamma_{12}^2 - \Gamma_{11}^1) - N\Gamma_{11}^2, \\ M_v - N_u &= L\Gamma_{22}^1 + M(\Gamma_{22}^2 - \Gamma_{12}^1) - N\Gamma_{11}^2. \end{aligned}$$

4. GEODESICS

We gave in the previous section the definition of a geodesic curve. Namely a geodesic is a curve with zero geodesic curvature or equivalently:

Definition 4.1 *A curve $\gamma\colon I \to X$, parameterized by arc length on a surface X, is a **geodesic** if for all $s \in I$ the vector $\ddot{\gamma}(s)$ is normal to the surface at the point $\gamma(s)$.*

Geodesics are also the curves of shortest length on a surface – at least 'locally'. This means that given a geodesic between two points on a surface, varying the geodesic slightly will produce curves of greater length. For example, given two points on a sphere the great circle containing these two points is a geodesic. If the points are not antipodal, then there will be a shorter and longer arc connecting them. However both arcs are geodesics and locally are the shortest paths between the points.

We will see that geodesics are determined by the first fundamental form. Consequently an isometry between two surfaces will map geodesics in the first surface to geodesics in the second.

Theorem 4.2 *Let X be a smooth parameterized surface and γ be a smooth curve on X parameterized by arc length s. Then γ is a geodesic if and only the parameters $(u(s), v(s))$ of $\gamma(s)$ satisfy*

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}s}(E\dot{u} + F\dot{v}) &= \frac{1}{2}(E_u\dot{u}^2 + 2F_u\dot{u}\dot{v} + G_u\dot{v}^2) \\ \frac{\mathrm{d}}{\mathrm{d}s}(F\dot{u} + G\dot{v}) &= \frac{1}{2}(E_v\dot{u}^2 + 2F_v\dot{u}\dot{v} + G_v\dot{v}^2) \end{aligned} \tag{4.18}$$

for all s, where $E\mathrm{d}u^2 + 2F\mathrm{d}u\mathrm{d}v + G\mathrm{d}v^2$ is the first fundamental form of X.

Proof. As $\mathbf{r}_u$ and $\mathbf{r}_v$ are independent tangent vectors then $\ddot{\gamma}(s)$ is normal to the surface if and only if $\ddot{\gamma}(s) \cdot \mathbf{r}_u = 0$ and $\ddot{\gamma}(s) \cdot \mathbf{r}_v = 0$. Now

$$\dot{\gamma}(s) = \dot{u}\mathbf{r}_u + \dot{v}\mathbf{r}_v.$$

Thus

$$\begin{aligned} 0 &= \ddot{\gamma} \cdot \mathbf{r}_u = \frac{\mathrm{d}}{\mathrm{d}s}(\dot{\gamma} \cdot \mathbf{r}_u) - \dot{\gamma} \cdot \dot{\mathbf{r}}_u \\ &= \frac{\mathrm{d}}{\mathrm{d}s}(E\dot{u} + F\dot{v}) - (\dot{u}\mathbf{r}_u + \dot{v}\mathbf{r}_v) \cdot (\mathbf{r}_{uu}\dot{u} + \mathbf{r}_{uv}\dot{v}) \\ &= \frac{\mathrm{d}}{\mathrm{d}s}(E\dot{u} + F\dot{v}) - ((\mathbf{r}_{uu} \cdot \mathbf{r}_u)\dot{u}^2 + (\mathbf{r}_{uu} \cdot \mathbf{r}_v + \mathbf{r}_{uv} \cdot \mathbf{r}_u)\dot{u}\dot{v} + (\mathbf{r}_{uv} \cdot \mathbf{r}_v)\dot{v}^2) \\ &= \frac{\mathrm{d}}{\mathrm{d}s}(E\dot{u} + F\dot{v}) - \frac{1}{2}(E_u\dot{u}^2 + 2F_u\dot{u}\dot{v} + G_u\dot{v}^2), \end{aligned}$$

as required. The second geodesic equation follows similarly. ■

Given two points on a surface there need not be a geodesic connecting the two points. For example in $\mathbb{R}^2$ the geodesics are line segments. So in the punctured plane $\mathbb{R}^2\backslash\{\mathbf{0}\}$ there is no geodesic connecting $(1, 0)$ and $(-1, 0)$. Also if a geodesic exists between two points it need not be unique (see the examples of the sphere and cylinder below.) However geodesics always exist locally (Do Carmo p.255):

Theorem 4.3 *Given a point $p \in X$ and a non-zero vector $\mathbf{v} \in T_pX$ then there exists $\varepsilon > 0$ and a unique geodesic $\gamma\colon (-\varepsilon, \varepsilon) \to X$ parameterized by arc-length such that $\gamma(0) = p$ and $\gamma'(0) = \mathbf{v}$.*

The proof of this theorem is beyond this syllabus and in any case largely relates to the analysis of differential equations. The existence and uniqueness of geodesics, at least locally, leads to the notion of *geodesic polar co-ordinates* (mentioned in Remark 3.16). When polar co-ordinates are used to parameterize the plane, from the origin, we obtain a first fundamental form with $E = 1$ and $F = 0$. More generally, we can locally parameterize a surface around a point p, by assigning co-ordinates r and θ to the point of the surface that is distance r from p when measured along the geodesic making an angle θ at p with some fixed tangential direction. When we do this we find that $E = 1$ and $F = 0$ (Do Carmo p.287).

For many surfaces geodesics are not just locally defined but many be extended indefinitely – such surfaces are called *complete surfaces*. Note that this may mean that the geodesic wraps back on to itself as with a great circle on a sphere. This notion of completeness coincides with the notion of complete metric spaces. Given a connected surface X it can be shown that, given two points a, b of the surface, there is a piecewise-smooth curve between the points. The *intrinsic distance* $d(a, b)$ between the points can then be defined as the infimum

$$d(a, b) = \inf_{\gamma} \mathcal{L}(\gamma)$$

where $\mathcal{L}(\gamma)$ is the length of a curve γ and the infimum is taken over all piecewise smooth curves γ in X which connect a and b. The function d is a metric on X and the *Hopf-Rinow theorem* states that:

Theorem 4.4 ***(Hopf-Rinow)*** *(Off-syllabus) For a connected, smooth geometric surface X the following are equivalent:*

(a) (X, d) is a complete metric space.

(b) A geodesic can be indefinitely extended.

The following examples are left to Sheet 3, Part B, Exercise 1.

Example 4.5 *(a) The geodesics on a sphere are arcs of great circles.*

(b) The geodesics on a cylinder are the meridians, the latitudes and helices. So between two points of the cylinder, that do not lie on the same meridian or parallel, there are infinitely many geodesics between the points.

Example 4.6 *(a) Prove that a meridian on a surface of revolution is a geodesic.*

(b) When is a parallel of latitude a geodesic on such a surface?

Solution. Suppose that the surface of revolution is generated by rotating the curve $y = f(x)$ about the x-axis and parameterize it as

$$\mathbf{r}(x, \theta) = (x, f(x)\cos\theta, f(x)\sin\theta), \qquad x \in \mathbb{R},\ \theta \in (-\pi, \pi).$$

By Sheet 2, Part A, Exercise 2, the first fundamental form equals

$$(1 + f'(x)^2)\mathrm{d}x^2 + f(x)^2\mathrm{d}\theta^2$$

and the geodesic equations are

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}s}((1+f'(x)^2)\dot{x}) &= f'(x)(f''(x)\dot{x}^2+f(x)\dot{\theta}^2),\\ \frac{\mathrm{d}}{\mathrm{d}s}(f(x)^2\dot{\theta}) &= 0.\end{aligned}$$

(a) Along a meridian $\dot{\theta}=0$ and $\dot{x}=(1+f'(x)^2)^{-1/2}$. The second equation is then trivially true and substituting into the first equation we find

$$\frac{\mathrm{d}}{\mathrm{d}s}((1+f'(x)^2)\dot{x}) = \dot{x}\frac{\mathrm{d}}{\mathrm{d}x}(1+f'(x)^2)^{1/2} = \frac{f'(x)f''(x)}{1+f'(x)^2} = f'(x)f''(x)\dot{x}^2$$

as required.

(b) A parallel is given by the equation $\dot{x}=0$. Thus the two geodesic equations now read as $f(x)f'(x)\dot{\theta}^2=0$ and $f(x)^2\ddot{\theta}=0$. As the geodesic is a circle parameterized by arc-length then $\dot{\theta}$ is a non-zero constant and $\ddot{\theta}=0$. As $f(x)>0$, then the equations hold if and only if $f'(x)=0$. ■

We now prove an earlier comment on geodesics namely that they are locally curves of least length. That is, however a geodesic between two points is perturbed, we produce curves of greater length

Theorem 4.7 *Let $\gamma\colon [a,b]\to X$ be a smooth geodesic in X. Let γ_δ, where $\delta\in(-\varepsilon,\varepsilon)$, be a family of smooth curves*

$$\gamma_\delta\colon [a,b]\to X$$

with $\gamma_0=\gamma$ and $\gamma_\delta(a)=\gamma(a),\gamma_\delta(b)=\gamma(b)$ for all $\delta\in(-\varepsilon,\varepsilon)$ and let $\mathcal{L}(\delta)=\mathcal{L}(\gamma_\delta)$. Then $\mathcal{L}'(0)=0$.

Proof. (Proof non-examinable) Let $R(\delta,t)=E\dot{u}^2+2F\dot{u}\dot{v}+G\dot{v}^2$ where $\gamma_\delta(t)=\mathbf{r}(u(\delta,t),v(\delta,t))$ and the dot denotes differentiation with respect to t. Then

$$\mathcal{L}(\delta)=\int_a^b \sqrt{R}\ \mathrm{d}t$$

giving

$$\mathcal{L}'(0)=\frac{\mathrm{d}}{\mathrm{d}\delta}\bigg|_{\delta=0}\int_a^b\sqrt{R}\,\mathrm{d}t=\int_a^b\frac{\partial\sqrt{R}}{\partial\delta}\bigg|_{\delta=0}\mathrm{d}t=\frac{1}{2}\int_a^b\frac{1}{\sqrt{R}}\frac{\partial R}{\partial\delta}\bigg|_{\delta=0}\mathrm{d}t, \tag{4.19}$$

by differentiation under the integral sign. Now

$$\begin{aligned}\frac{\partial R}{\partial\delta} &= \{E_u\dot{u}^2+2F_u\dot{u}\dot{v}+G_u\dot{v}^2\}\frac{\partial u}{\partial\delta}\\ &+ \{E_v\dot{u}^2+2F_v\dot{u}\dot{v}+G_v\dot{v}^2\}\frac{\partial v}{\partial\delta}\\ &+ 2(E\dot{u}+F\dot{v})\frac{\partial\dot{u}}{\partial\delta}+2(F\dot{u}+G\dot{v})\frac{\partial\dot{v}}{\partial\delta}.\end{aligned}$$

As $\gamma = \gamma_0$ is a geodesic then substituting in the geodesic equations (4.18)

$$\begin{aligned}\left.\frac{\partial R}{\partial \delta}\right|_{\delta=0} &= 2\left[\frac{\mathrm{d}}{\mathrm{d}t}(E\dot{u}+F\dot{v})\frac{\partial u}{\partial \delta}+(E\dot{u}+F\dot{v})\frac{\partial \dot{u}}{\partial \delta}\right.\\ &\quad\left.+\ \frac{\mathrm{d}}{\mathrm{d}t}(F\dot{u}+G\dot{v})\frac{\partial v}{\partial \delta}+(F\dot{u}+G\dot{v})\frac{\partial \dot{v}}{\partial \delta}\right]_{\delta=0}\\ &= 2\frac{\mathrm{d}}{\mathrm{d}t}\left\{(E\dot{u}+F\dot{v})\left.\frac{\partial u}{\partial \delta}\right|_{\delta=0}+(F\dot{u}+g\dot{v})\left.\frac{\partial v}{\partial \delta}\right|_{\delta=0}\right\}. \qquad (4.20)\end{aligned}$$

We may assume without loss of generality that $\gamma = \gamma_0$ is parameterized by arc length so that $R(0,t) = 1$. Hence, substituting (4.20) into (4.19),

$$\begin{aligned}\mathcal{L}'(0) &= \int_a^b \frac{\mathrm{d}}{\mathrm{d}t}\left\{(E\dot{u}+F\dot{v})\left.\frac{\partial u}{\partial \delta}\right|_{\delta=0}+(F\dot{u}+g\dot{v})\left.\frac{\partial v}{\partial \delta}\right|_{\delta=0}\right\}\mathrm{d}t\\ &= \left[(E\dot{u}+F\dot{v})\left.\frac{\partial u}{\partial \delta}\right|_{\delta=0}+(F\dot{u}+g\dot{v})\left.\frac{\partial v}{\partial \delta}\right|_{\delta=0}\right]_{t=a}^{t=b}.\end{aligned}$$

However $u(\delta,a), u(\delta,b), v(\delta,a)$ and $v(\delta,b)$ are all constant giving $\partial u/\partial\delta = \partial v/\partial\delta = 0$ when $t = a$ and $t = b$ and hence $\mathcal{L}'(0) = 0$. ■

Example 4.8 *What are the geodesics in the hyperbolic plane $\mathbb{H}$? (See Examples 3.28 and 3.54.)*

Solution. Method 1: If we substitute $E = G = y^{-2}$ and $F = 0$ into the geodesic equations (4.18) then we find

$$\frac{\mathrm{d}}{\mathrm{d}s}\left(\frac{\dot{x}}{y^2}\right) = 0, \qquad \frac{\mathrm{d}}{\mathrm{d}s}\left(\frac{\dot{y}}{y^2}\right) = \frac{-(\dot{x}^2+\dot{y}^2)}{y^3}.$$

The first equation yields $\dot{x} = cy^2$ for some constant c. So the half-lines $x =$ constant are then geodesics corresponding to $c = 0$. Assume that $c \neq 0$. The second equation may be rewritten as

$$\frac{\ddot{y}y-\dot{y}^2}{y^2} = \frac{-\dot{x}^2}{y^2},$$

or equivalently

$$\frac{\mathrm{d}}{\mathrm{d}s}\left(\frac{\dot{y}}{y}\right) = -c\dot{x}.$$

Integrating we find that $\dot{y} = (b-cx)y$ for some constant b. Now

$$\frac{\mathrm{d}y}{\mathrm{d}x} = \frac{\dot{y}}{\dot{x}} = \frac{b-cx}{cy},$$

and solving this differential equation gives

$$\frac{1}{2}c(x^2+y^2) - bx = a,$$

for some constant a, which is the equation of a semicircle in $\mathbb{H}$ which cuts the x-axis orthogonally.

Method 2: Alternatively we could consider what the isometries of $\mathbb{H}$ might be and use the fact that geodesics are mapped to other geodesics by isometries. For ease of notation we now introduce a complex variable $z = x + iy$ so that the first fundamental form on $\mathbb{H}$ is now given by

$$\frac{-4|\mathrm{d}z|^2}{(z-\bar{z})^2}.$$

Then I claim the map

$$w\colon z \mapsto \frac{az+b}{cz+d},$$

where a, b, c, d are real numbers satisfying $ad - bc = 1$, is an isometry of $\mathbb{H}$. From standard theorems concerning Möbius transformations we can see that w maps the upper half plane onto the upper half plane; as a, b, c, d are real, the real axis is mapped to the real axis and the imaginary part of the image of i equals

$$\mathrm{Im}\left(\frac{ai+b}{ci+d}\right) = \frac{\mathrm{Im}\left((ai+b)\left(d-ci\right)\right)}{c^2+d^2} = \frac{ad-bc}{c^2+d^2} = \frac{1}{c^2+d^2} > 0.$$

To check w is an isometry we need to prove that $\mathbb{H}$ when parameterized by w and z has the same first fundamental form. Firstly note

$$\mathrm{d}w = \frac{\mathrm{d}z}{(cz+d)^2}.$$

So

$$\frac{-4|\mathrm{d}w|^2}{(w-\bar{w})^2} = \frac{\frac{-4|\mathrm{d}z|^2}{|cz+d|^4}}{\left(\frac{az+b}{cz+d} - \frac{a\bar{z}+b}{c\bar{z}+d}\right)^2} = \frac{-4|\mathrm{d}z|^2}{((az+b)(c\bar{z}+d) - (cz+d)(a\bar{z}+b))^2}.$$

The denominator in the final expression above factorises as $(ad-bc)^2(z-\bar{z})^2$ showing that

$$\frac{-4|\mathrm{d}w|^2}{(w-\bar{w})^2} = \frac{-4|\mathrm{d}z|^2}{(z-\bar{z})^2}$$

and consequently w is an isometry.

Note now that $x = 0, y = e^{-s}$ is a solution to the geodesic equations for $\mathbb{H}$ (Sheet 3, Part A, Exercise 1) and so the positive imaginary axis is a geodesic. As we show below, there is a Möbius map of the same form as w which maps any other half line or semicircle orthogonal to the positive imaginary axis, showing that these too are examples of geodesics. From Theorem 4.3 we know that these are all the geodesics of $\mathbb{H}$.

Given another half-line $\mathrm{Re}\, z = k$ then the Möbius map $z \mapsto z - k$ (where $a = 1, b = -k, c = 0, d = 1$ so that $ad - bc = 1$) takes the half-line to the positive imaginary axis. For the semicircle perpendicular to the real axis, meeting at p and q (where $p < q$), the Möbius map

$$z \mapsto \frac{1}{\sqrt{q-p}}\left(\frac{p-z}{z-q}\right)$$

takes the semicircle to the positive imaginary axis. Again we check

$$ad - bc = \left(\frac{-1}{\sqrt{q-p}}\right)\left(\frac{-q}{\sqrt{q-p}}\right) - \left(\frac{p}{\sqrt{q-p}}\right)\left(\frac{1}{\sqrt{q-p}}\right) = \frac{q-p}{q-p} = 1.$$

■

Remark 4.9 *(**Historical context**) The hyperbolic plane $\mathbb{H}$ is of interest because it is an example of a non-Euclidean geometry. A Euclidean geometry is one that satisfies certain axioms including the **parallel postulate** which states that:*

- *given a line l and a point p not on l then there is a unique line through p (known as a parallel) which does not meet l.*

If we read 'geodesic' for 'line' in the above, then we see that given a line l in $\mathbb{H}$ and a point p not on the line then there are infinitely many lines through p not meeting l. (In Figure 4.1 M_1, M_2, M_3 are parallels of L through P.)

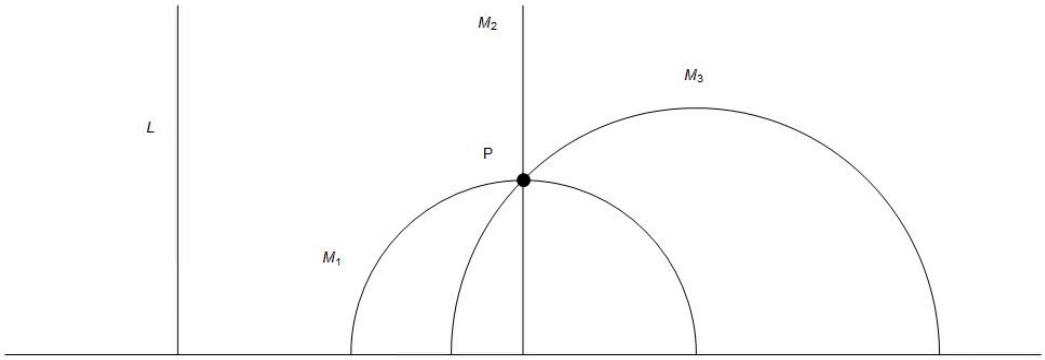


Figure 4.1 – hyperbolic parallels

For literal centuries, mathematicians had been trying to deduce the axiom of parallels from Euclid's other axioms. Instead all they managed to find were alternative, equivalent formulations for the parallel postulate. The above formulation is in fact due to Ludlam (1785) though it is usually attritbuted to Playfair; other formulations include:

- *parallel lines are everywhere equidistant.*
- *the sum of the angles of a triangle equals two right angles.*
- *given a triangle, we can construct a similar triangle of any area.*
- *Pythagoras' theorem.*
- *three non-collinear points always lie on a circle.*

In the nineteenth century certain mathematicians – notably Bolyai, Lobachevsky and Gauss – began to suspect that the parallel postulate was independent of the other axioms and proved alternative theory where more than one parallel existed. Such theory might still have contained inconsistencies, but this was shown not to be the case when Beltrami, Klein and Poincaré found models for the hyperbolic plane which showed the new geometry to be every bit as consistent as Euclidean geometry.

The elliptic plane (Example 3.32) is another example of non-Euclidean geometry, in which case there are no parallels to a line. The elliptic plane had previously been discounted as a non-Euclidean geometry as it did not seem to meet another axiom of Euclid that lines can be extended indefinitely. But if we permit lines to be extended repeatedly on to themselves then the elliptic plane is a valid non-Euclidean geometry.

5. GEOMETRY & ANALYSIS MEET TOPOLOGY

In this chapter we will meet some startling results which connect the topology of a surface with global aspects of geometry and analysis. For example, the *global Gauss-Bonnet theorem* says that for a closed geometric surface X,

$$\iint_X K \, \mathrm{d}A = 2\pi\chi(X).$$

What is striking about this result is that the term on the RHS is manifestly topological in nature whilst the total curvature on the LHS is ostensibly geometric. It is possible to distort a surface locally to change its Gaussian curvature without changing its topology, but the above theorem shows there will be knock-on effects elsewhere on the surface as the total curvature must remain constant. There are other corollaries to this result such as (a) the sphere is the only orientable closed surface which can have positive curvature everywhere and (b) the torus is the only orientable closed surface which can be everywhere flat (Example 3.26).

5.1 The Gauss-Bonnet theorems

We begin first with a proof of the *local Gauss-Bonnet theorem*. The statement of this theorem is on the syllabus *but its proof is not*; I include the proof here for completeness' sake.

Theorem 5.1 *(**Local Gauss-Bonnet Theorem** – first version) (Proof off syllabus) Let γ be a smooth, simple, closed curve on a patch of surface X, enclosing a region R. Then*

$$\int_\gamma k_g \, \mathrm{d}s + \iint_R K \, \mathrm{d}A = 2\pi.$$

Proof. We will assume that $X = \mathbf{r}(U)$ where $\mathbf{r}$ is an orthogonal parameterization, so that $F = 0$. (The existence of such fields was mentioned in Remark 3.16.) We then set

$$\mathbf{e}_1 = \frac{\mathbf{r}_u}{\sqrt{E}}, \qquad \mathbf{e}_2 = \frac{\mathbf{r}_v}{\sqrt{G}},$$

to be smooth, orthonormal, tangent, vector fields $\mathbf{e}_1, \mathbf{e}_2 \colon V \to \mathbb{R}^3$. Let $\theta(s)$ denote the angle between the unit vector $\dot\gamma(s)$ and $\mathbf{e}_1$ at the point $\gamma(s)$, so that

$$\dot\gamma = \mathbf{e}_1 \cos\theta + \mathbf{e}_2 \sin\theta$$

giving

$$\ddot\gamma = \dot\theta \left(-\mathbf{e}_1 \sin\theta + \mathbf{e}_2 \cos\theta\right) + \left(\dot{\mathbf{e}}_1 \cos\theta + \dot{\mathbf{e}}_2 \sin\theta\right).$$

With $\mathbf{n} = \mathbf{e}_1 \wedge \mathbf{e}_2$ then

$$\mathbf{n} \wedge \dot\gamma = -\mathbf{e}_1 \sin\theta + \mathbf{e}_2 \cos\theta,$$

as $\mathbf{n} \wedge \mathbf{e}_1 = \mathbf{e}_2$ and $\mathbf{n} \wedge \mathbf{e}_2 = -\mathbf{e}_1$, so that

$$\ddot{\gamma} = \dot{\theta}\left(\mathbf{n} \wedge \dot{\gamma}\right) + \left(\dot{\mathbf{e}}_1 \cos\theta + \dot{\mathbf{e}}_2 \sin\theta\right).$$

Thus

$$\begin{aligned} k_g &= \ddot{\gamma} \cdot (\mathbf{n} \wedge \dot{\gamma}) \\ &= \dot{\theta} + (\dot{\mathbf{e}}_1 \cos\theta + \dot{\mathbf{e}}_2 \sin\theta) \cdot (-\mathbf{e}_1 \sin\theta + \mathbf{e}_2 \cos\theta) \\ &= \dot{\theta} - \mathbf{e}_1 \cdot \dot{\mathbf{e}}_2 \end{aligned}$$

because $\dot{\mathbf{e}}_1 \cdot \mathbf{e}_1 = 0 = \dot{\mathbf{e}}_2 \cdot \mathbf{e}_2$ and $\dot{\mathbf{e}}_1 \cdot \mathbf{e}_2 = -\mathbf{e}_1 \cdot \dot{\mathbf{e}}_2$ from differentiating $\mathbf{e}_1 \cdot \mathbf{e}_2 = 0$ and $\mathbf{e}_1 \cdot \mathbf{e}_1 = 1 = \mathbf{e}_2 \cdot \mathbf{e}_2 = 0$; thus we have

$$2\pi - \int_\gamma k_g \,\mathrm{d}s = \Delta\theta - \int_\gamma k_g \,\mathrm{d}s = \int_\gamma \left(\dot{\theta} - k_g\right) \,\mathrm{d}s = \int_\gamma \mathbf{e}_1 \cdot \dot{\mathbf{e}}_2 \,\mathrm{d}s.$$

We will then apply Green's theorem to this last integral. Recall Green's theorem states: for a smooth, simple, closed curve β in an open set $V \subseteq \mathbb{R}^2$, bounding a region S, with P, Q being two smooth functions defined on V

$$\int_\beta (P \,\mathrm{d}u + Q \,\mathrm{d}v) = \iint_S \left(\frac{\partial Q}{\partial u} - \frac{\partial P}{\partial v}\right) \mathrm{d}u \,\mathrm{d}v.$$

Let $\beta = \mathbf{r}^{-1}(\gamma)$. We have

$$\dot{\mathbf{e}}_2 = \frac{\partial \mathbf{e}_2}{\partial u}\frac{\mathrm{d}u}{\mathrm{d}s} + \frac{\partial \mathbf{e}_2}{\partial v}\frac{\mathrm{d}v}{\mathrm{d}s},$$

so that $P = \mathbf{e}_1 \cdot \partial \mathbf{e}_2/\partial u$ and $Q = \mathbf{e}_1 \cdot \partial \mathbf{e}_2/\partial v$. Then

$$\frac{\partial Q}{\partial u} - \frac{\partial P}{\partial v} = \frac{\partial \mathbf{e}_1}{\partial u} \cdot \frac{\partial \mathbf{e}_2}{\partial v} - \frac{\partial \mathbf{e}_1}{\partial v} \cdot \frac{\partial \mathbf{e}_2}{\partial u}.$$

Differentiating $\mathbf{e}_1 = \mathbf{r}_u/\sqrt{E}$ and $\mathbf{e}_2 = \mathbf{r}_v/\sqrt{G}$ we find

$$\frac{\partial \mathbf{e}_1}{\partial u} = \frac{1}{\sqrt{E}}\mathbf{r}_{uu} - \frac{E_u}{2E^{3/2}}\mathbf{r}_u, \qquad \frac{\partial \mathbf{e}_2}{\partial v} = \frac{1}{\sqrt{G}}\mathbf{r}_{vv} - \frac{G_v}{2G^{3/2}}\mathbf{r}_v.$$

Noting that $\mathbf{r}_u, \mathbf{r}_v$ and $\mathbf{n}$ are mutually orthogonal, we find

$$\begin{aligned} \frac{\partial \mathbf{e}_1}{\partial u} \cdot \frac{\partial \mathbf{e}_2}{\partial v} &= \frac{\mathbf{r}_{uu} \cdot \mathbf{r}_{vv}}{\sqrt{EG}} - \frac{E_u \mathbf{r}_u \cdot \mathbf{r}_{vv}}{2E^{3/2}\sqrt{G}} - \frac{G_v \mathbf{r}_{uu} \cdot \mathbf{r}_v}{2G^{3/2}\sqrt{E}} \\ &= \frac{\Gamma^1_{11}\Gamma^1_{22}E + \Gamma^2_{11}\Gamma^2_{22}G + LN}{\sqrt{EG}} - \frac{E_u\Gamma^1_{22}}{2\sqrt{EG}} - \frac{G_v\Gamma^2_{11}}{2\sqrt{EG}}. \end{aligned}$$

From Corollary 3.51, when $F = 0$, we have

$$\begin{aligned} \Gamma^1_{11} &= E_u/2E, \quad \Gamma^1_{12} = E_v/2E, \quad \Gamma^1_{22} = -G_u/2E, \\ \Gamma^2_{11} &= -E_v/2G, \quad \Gamma^2_{12} = G_u/2G, \quad \Gamma^2_{22} = G_v/2G. \end{aligned}$$

and so the above simplifies to

$$\frac{\partial \mathbf{e}_1}{\partial u} \cdot \frac{\partial \mathbf{e}_2}{\partial v} = \frac{LN}{\sqrt{EG}}$$

Similarly

$$\frac{\partial \mathbf{e}_1}{\partial v} \cdot \frac{\partial \mathbf{e}_2}{\partial u} = \frac{\Gamma^1_{12}\Gamma^1_{21}E + \Gamma^2_{12}\Gamma^2_{21}G - M^2}{\sqrt{EG}} - \frac{E_v\Gamma^1_{21}}{2\sqrt{EG}} - \frac{G_u\Gamma^2_{12}}{2\sqrt{EG}} = \frac{-M^2}{\sqrt{EG}}.$$

Hence

$$I = \iint_S \frac{LN - M^2}{\sqrt{EG}}\,\mathrm{d}u\,\mathrm{d}v = \iint_S K\sqrt{EG}\,\mathrm{d}u\,\mathrm{d}v = \iint_S K\,\mathrm{d}A.$$

■

Theorem 5.2 *(**Local Gauss-Bonnet Theorem** – second version.) (Proof off syllabus)*

Let γ be a piecewise-smooth simple, closed curve on a patch of surface X, enclosing a region R. Then

$$\int_\gamma k_g\,\mathrm{d}s + \iint_R K\,\mathrm{d}A + \sum_{i=1}^{n} \alpha_i = 2\pi$$

where $\alpha_1, \ldots, \alpha_n$ are the external angles at the points where γ is not smooth.

Proof. The proof is almost identical to the proof of the first version save that at those points where γ is not smooth there is a jump discontinuity in $\theta(s)$ of α_i where α_i is the external angle. The only amendment needed to the proof is that

$$\int_\gamma \dot{\theta}\,\mathrm{d}s = \Delta\theta = 2\pi - \sum_{i=1}^{n} \alpha_i.$$

■

Example 5.3 *Note that when we use internal angles $\beta_i = \pi - \alpha_i$, and when the curvilinear polygon R is bounded by geodesics, then we obtain*

$$\iint_R K\,\mathrm{d}A = \sum_{i=1}^{n} \beta_i - (n-2)\pi.$$

Thus the internal angle sum exceeds $(n-2)\pi$ by the total curvature. Focusing on triangles: in the plane, where $K = 0$, we have

$$\beta_1 + \beta_2 + \beta_3 = \pi,$$

whilst in the hyperbolic plane (where $K = -1$) we have Lambert's Theorem

$$A = \pi - \beta_1 - \beta_2 - \beta_3,$$

and on the sphere or elliptic plane (where $K = 1$) then we have Girard's Theorem

$$A = \beta_1 + \beta_2 + \beta_3 - \pi.$$

Theorem 5.4 ***(Global Gauss-Bonnet Theorem)*** *Let X be a smooth, closed, orientable surface. Then*

$$\iint_R K \, \mathrm{d}A = 2\pi\chi(X).$$

Proof. Say that X is subdivided by smooth curves into curvilinear polygons. We apply the local GBT to each of these polygons and sum each of the resulting equations.

The contributions $\int_\gamma k_g \, \mathrm{d}s$ cancel out as each edge bounds two polygons but with different orientations. For one orientation k_g is negative what it equals on the reverse orientation. The sum of the total curvature from each polygon equals the total curvature on the surface. And, using internal angles, we also need to sum the expressions

$$\sum_{i=1}^{n} \beta_i - n\pi + 2\pi.$$

The sum of the internal angles equals $2\pi V$ where V is the number of vertices. This is because at each vertex the internal angles add up to 2π. Now we have F faces so that

$$\sum_{\text{faces}} 2\pi = 2\pi F,$$

and each edge bounds two faces so that

$$\sum_{\text{faces}} n_{\text{face}}\pi = \sum_{\text{edges}} 2\pi = 2\pi E,$$

finally yielding

$$\sum_{\text{faces}} \left(\sum_{i=1}^{n} \beta_i - n_{\text{face}}\pi + 2\pi \right) = 2\pi V - 2\pi E + 2\pi F = 2\pi\chi(X).$$

■

Remark 5.5 *(a) We have assumed, without proof, that every compact, smooth surface has a subdivision. This is true – in fact this is more generally true for any separable smooth surface.*

(b) The above proof is for closed orientable surfaces. It relies on orientability when we refer to the opposite orientations of two curves. However the theorem also holds for non-orientable closed surfaces.

(c) In the next chapter we will discuss closed surfaces of constant curvature. The global Gauss-Bonnet theorem makes plain that only certain surfaces might be endowed with first fundamental forms with constant Gaussian curvature. A closed geometric surface with constant positive/zero/negative Gaussian curvature is necessarily a sphere/torus/torus with more than one hole. That is because their Euler characteristics are $2/0/ \leqslant -2$. The theorem only gives necessity. As a sphere has constant positive curvature and as the flat torus has constant zero curvature, such surfaces are clearly possible. When we study quotients of the hyperbolic plane we will construct surfaces of constant curvature -1 of each positive genus.

Example 5.6 *Show that the catenoid $x^2 + y^2 = \cosh^2 z$ has a single, simple, closed geodesic.*

Solution. The 'waist' $z = 0$ is a simple, closed geodesic by Example 4.6(b) – it is a latitude where the radius function is at a minimum.

The catenoid can be parameterized by

$$\mathbf{r}(u, v) = (\cosh v \cos u, \cosh v \sin u, v), \qquad 0 < u < 2\pi,\ v \in \mathbb{R}.$$

The Gaussian curvature at the point $\mathbf{r}(u, v)$ equals $K(u, v) = -\cosh^{-4} v < 0$. Note that there cannot be a simple, closed geodesic that does not wrap once around the catenoid. By the local Gauss-Bonnet theorem we would then have

$$0 > \iint_R K \, \mathrm{d}A = 2\pi,$$

which is a contradiction. Suppose now that there were two simple, closed geodesics wrapping once around the catenoid. If these geodesics do not intersect and enclose a region R between them then we would have

$$0 > \iint_R K \, \mathrm{d}A = 2\pi\chi(R) = 0,$$

as the Euler characteristic of R (which is a cylinder) equals 0. Again we have a contradiction.

Finally suppose that the two geodesics do intersect and let R be the region bounded by them. Should they intersect once we would have

$$\iint_R K \, \mathrm{d}A + (\pi - \beta_1) + (\pi - \beta_2) = 2\pi,$$

where β_1 and β_2 are the two internal angles at the point of intersection. The LHS is less than 2π and so again we have a contradiction. Should the geodesics intersect more than once then we can focus on the geodesics between two points of intersection to get the same contradiction. ■

5.2 The Poincaré-Hopf theorem

Suppose that we are given a tangent vector $\mathbf{v}(x)$ at each point x of a smooth, closed surface X in $\mathbb{R}^3$. We can think of $\mathbf{v}(x)$ as the velocity at x of some fluid flow $\mathbf{v}$ on the surface. A point where $\mathbf{v}(x) = \mathbf{0}$ is called a **stationary (or singular) point** of the flow. It is a well known fact – the *hairy ball theorem* – that a flow on a sphere must have at least one stationary point. This is a consequence of the sphere's topology and we will more generally prove the *Poincaré-Hopf theorem* for surfaces which states that

$$\chi(X) = \sum_{\substack{\text{stationary} \\ \text{points } x}} \operatorname{index}(\mathbf{v}(x))$$

where the *index* (or *multiplicity*) is an integer associated with each stationary point, and assuming there to be finitely many stationary points.

If $x \in X$ is an isolated stationary point of $\mathbf{v}$ then we can find a small neighbourhood U of x such that $\mathbf{v}$ is non-zero on $U \backslash \{x\}$. Now let $\mathbf{e}$ be another smooth, nowhere zero, vector field defined on U; we will use $\mathbf{e}$ as a reference direction with which to compare the behaviour of

$\mathbf{v}(x)$. Let $\gamma(t)$ be a simple, closed, anticlockwise smooth curve in U which encircles x. Then $\mathbf{v}$ and $\mathbf{e}$ are both non-zero on γ and we defined the index as the winding number of $\mathbf{v}$ with respect to $\mathbf{e}$ as γ is traversed once. That is:

$$\text{index } = \frac{\Delta\psi}{2\pi} = \frac{1}{2\pi}\int_\gamma \frac{\mathrm{d}\psi}{\mathrm{d}t}\,\mathrm{d}t$$

where ψ is the angle between $\mathbf{v}$ and $\mathbf{e}$. Whilst ψ is only defined up to multiples of 2π this does not affect the total change $\Delta\psi$ in ψ.

Remark 5.7 *It is not immediately clear that the index of a stationary point is well-defined. It may depend on the choice of vector field* $\mathbf{e}$ *or on the curve* γ*. In Do Carmo, p.280, it is shown that*

$$\Delta\psi = \iint_R K\,\mathrm{d}A,$$

where R *is the region bounded by* γ*. This then is independent of the choice of* $\mathbf{e}$*.*

Say now that γ_0 *and* γ_1 *are two simple, closed, positively oriented curves around* x*. These curves are then homotopic and it is possible to create a family of simple, closed, positively oriented curves* γ_t*, where* $0 \leqslant t \leqslant 1$*, which continuously deform from* γ_0 *to* γ_1*. Let* $I(t)$ *be the index as calculated using* γ_t*. Then* $I(t)$ *is a continuous, integer-valued function on* $[0,1]$*, so by connectedness* $I(t)$ *is constant and in particular* $I(0) = I(1)$*. This shows that the index is independent of the choice of* γ*.*

Example 5.8 *Find the index of each of the following stationary points at the origin:*

(a) source: $\mathbf{v}(x,y) = (x,y)$.
(b) sink: $\mathbf{v}(x,y) = (-x,-y)$.
(c) vortex: $\mathbf{v}(x,y) = (-y,x)$.
(d) bifurcation: $\mathbf{v}(x,y) = (x,-y)$.
(e) dipole: $\mathbf{v}(x,y) = (x^2-y^2, 2xy)$.

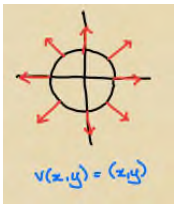


Figure. 5.1
source

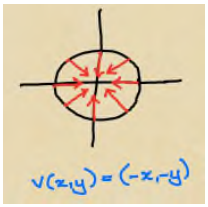


Figure 5.2
sink

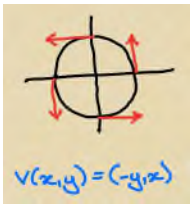


Figure 5.3
vortex

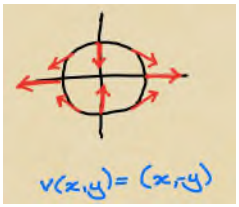


Figure 5.4
bifurcation

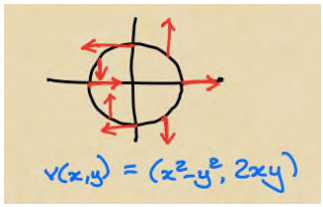


Figure 5.5
dipole

Solution. In each case, we will take γ to be the curve $\gamma(t) = (\cos t, \sin t)$ and $\mathbf{e} = (1,0)$.

(a) $\mathbf{v}(\cos t, \sin t) = (\cos t, \sin t)$ and so we may take $\psi = t$. Thus the index is 1.

(b) $\mathbf{v}(\cos t, \sin t) = (-\cos t, -\sin t) = (\cos(t+\pi), \sin(t+\pi))$ and so we may take $\psi = t+\pi$. Again the index is 1.

(c) $\mathbf{v}(\cos t, \sin t) = (-\sin t, \cos t) = (\cos(t+\pi/2), \sin(t+\pi/2))$ and so we may take $\psi = t+\pi/2$. Once more the index is 1.

(d) $\mathbf{v}(\cos t, \sin t) = (\cos t, -\sin t) = (\cos(-t), \sin(-t))$ and so we may take $\psi = -t$. Thus the index is -1.

(e) $\mathbf{v}(\cos t, \sin t) = (\cos^2 t - \sin^2 t, 2\sin t \cos t) = (\cos(2t), \sin(2t))$ and so we may take $\psi = 2t$. Thus the index is 2. ■

Theorem 5.9 *(Poincaré 1881, Hopf 1926) Let* $\mathbf{v}$ *be a smooth vector field on a smooth closed orientable surface* X *with finitely many stationary points. Then*

$$\chi(X) = \sum_{\substack{\text{stationary} \\ \text{points } x}} \text{index}(\mathbf{v}(x)).$$

Remark 5.10 *Henri Poincaré proved the above theorem for surfaces in 1881. Heinz Hopf generalized the result to higher-dimenstional manifolds in 1926. Any continuous map of the unit circle can be assigned its degree – an integer describing how many times the circle wraps onto itself in an anticlockwise fashion and the index of a stationary point can be seen in this light. The degree of a map from a higher-dimensional sphere to itself can similarly be defined (Brouwer 1911) and the index of stationary points in higher dimensions can be simiarly understood.*

Proof. Let $x_1, \ldots, x_n$ be the stationary points of the vector field $\mathbf{v}$. Choose a smooth, simple, closed curve γ_i around each x_i enclosing a region R_i. Let

$$Y = X \setminus \bigcup_{i=1}^{n} R_i.$$

At each point $y \in Y$ we may choose an orthonormal basis $\{\mathbf{e}_1(y), \mathbf{e}_2(y)\}$ for the tangent space at y and such that $\mathbf{e}_1(y)$ is in the direction of the non-zero $\mathbf{v}(y)$. Applying the argument of the local Gauss-Bonnet theorem to the region Y we obtain

$$\iint_Y K \, \mathrm{d}A = -\sum_{i=1}^{n} \int_{\gamma_i} \mathbf{e}_1 \cdot \dot{\mathbf{e}}_2 \, \mathrm{d}s.$$

The negative sign is because the γ_i are oriented clockwise as part of the boundary of Y.

Choosing a similar orthonormal basis $\{\mathbf{f}_1, \mathbf{f}_2\}$ for the points in R_i we find that

$$\iint_{R_i} K \, \mathrm{d}A = \int_{\gamma_i} \mathbf{f}_1 \cdot \dot{\mathbf{f}}_2 \, \mathrm{d}s.$$

Adding each of these equations ($i = 1, \ldots, n$) to the previous equation and applying the global Gauss-Bonnet theorem we obtain

$$2\pi\chi(X) = \iint_X K \, \mathrm{d}A = \sum_{i=1}^{n} \int_{\gamma_i} \left(\mathbf{f}_1 \cdot \dot{\mathbf{f}}_2 - \mathbf{e}_1 \cdot \dot{\mathbf{e}}_2 \right) \mathrm{d}s.$$

From the proof of the local Gauss-Bonnet theorem we know that

$$\mathbf{e}_1 \cdot \dot{\mathbf{e}}_2 = \dot{\theta} - k_g, \qquad \mathbf{f}_1 \cdot \dot{\mathbf{f}}_2 = \dot{\varphi} - k_g$$

where θ and φ are the angles between $\dot{\gamma}$ and $\mathbf{e}_1$ and $\mathbf{f}_1$ respectively. Setting $\psi = \varphi - \theta$ to be the angle between $\mathbf{f}_1$ and $\mathbf{e}_1$ we obtain

$$\chi(X) = \sum_{i=1}^{n} \frac{1}{2\pi} \int_{\gamma_i} \dot{\psi} \, \mathrm{d}s = \sum_{i=1}^{n} \text{index}(\mathbf{v}(x_i))$$

as required. ■

Corollary 5.11 *(**Hairy ball theorem**) A smooth vector field on a sphere must have at least one stationary point.*

Proof. The follows from the fact that the Euler characteristic of a sphere is two. ■

5.3 Analysis on a closed surface

We now apply the Poincaré-Hopf theorem to demonstrate a first result in *Morse theory*. Morse theory, named after Marston Morse, includes a wide selection of results relating a surface's topology to the behaviour of smooth real functions on the surface.

Proposition 5.12 *(**Gradient vector field**) Let X be a smooth surface in $\mathbb{R}^3$, $p \in X$ and $f\colon X \to \mathbb{R}$ be smooth. Then there is a unique tangent vector, denoted $(\operatorname{grad}_X f)(p)$ or $(\nabla_X f)(p)$, such that*

$$(\operatorname{grad}_X f)(p) \cdot \mathbf{v} = \mathrm{d}f_p(\mathbf{v}) \tag{5.21}$$

for any tangent vector $\mathbf{v} \in T_pX$.

Proof. Parameterize X locally as $\mathbf{r}(u, v)$. It follows that

$$(\operatorname{grad}_X f)(p) \cdot \mathbf{r}_u = \mathrm{d}f_p(\mathbf{r}_u) = f_u(p), \qquad (\operatorname{grad}_X f)(p) \cdot \mathbf{r}_v = \mathrm{d}f_p(\mathbf{r}_v) = f_v(p).$$

As $\mathbf{r}_u$ and $\mathbf{r}_v$ form a basis for T_pX then this specifies $(\operatorname{grad}_X f)(p)$ uniquely. As the scalar product and $\mathrm{d}f_p$ are both linear, then (5.21) holds on the entire tangent space. ■

Exercise 5.13 *In terms of the local co-ordinates u, v, show that*

$$\operatorname{grad}_X f = \left(\frac{f_u G - f_v F}{EG - F^2}\right)\mathbf{r}_u + \left(\frac{f_v E - f_u F}{EG - F^2}\right)\mathbf{r}_v.$$

It then follows that $\operatorname{grad}_X f = 0$ if and only if $f_u = f_v = 0$. This is left to Sheet 3, Part A, Exercise 1.

Note that when $X = \mathbb{R}^2$, parameterized with Cartesian co-ordinates x, y, then

$$\operatorname{grad}_X f = f_x\mathbf{i} + f_y\mathbf{j}$$

concurs with the usual definition of ∇f.

Definition 5.14 *Given a smooth surface X in $\mathbb{R}^3$ and a smooth function $f\colon X \to \mathbb{R}$, we say that $p \in X$ is a **critical point** of f if $(\operatorname{grad}_X f)(p) = 0$. Equivalently, if $\mathbf{r}(u, v)$ is a local parameterization around p, then p is a critical point if and only if*

$$\frac{\partial f}{\partial u}(p) = 0 = \frac{\partial f}{\partial v}(p).$$

Example 5.15 *Let $f(x, y) = \cos \pi x + \cos \pi y$ on $\mathbb{R}^2$. Then*

$$\frac{\partial f}{\partial x} = -\pi \sin \pi x, \qquad \frac{\partial f}{\partial y} = -\pi \sin \pi y$$

are zero when x and y are integers.

Definition 5.16 *A critical point p is said to be* ***non-degenerate*** *if the Hessian matrix*

$$\begin{pmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial x \partial y} \\ \frac{\partial^2 f}{\partial y \partial x} & \frac{\partial^2 f}{\partial y^2} \end{pmatrix}$$

is nonsingular. If, further, the Hessian matrix is:

- *positive-definite then p is a local minimum;*
- *negative-definite then p is a local maximum;*
- *indefinite then p is a saddle point.*

A smooth real-valued function with only non-degerate critical points is called a ***Morse function****.*

It will become apparent in the proof of the next proposition that these definitions do indeed correspond to standard notions of a minima, maxima and saddle points.

Example 5.17 *With the above $f(x, y)$,the Hessian equals*

$$-\pi^2 \begin{pmatrix} \cos \pi x & 0 \\ 0 & \cos \pi y \end{pmatrix}$$

At $(0,0), (1,0)$, and $(1,1)$ this respectively equals

$$-\pi^2 \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \qquad -\pi^2 \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}, \qquad \pi^2 \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

These are respectively negative definite, indefinite and positive definite and so the points are respectively a maximum, a saddle point and a minimum.

Proposition 5.18 *Let f be a Morse function on a smooth patch X which has a critical point at p. Then $\operatorname{grad}_x f$ has:*
(a) index 1 at p if f has a minimum or a maximum at p.
(b) index -1 at p if f has a saddle point at p.

Proof. Take a conformal, local parameterization near $p = \mathbf{r}(0,0)$ and without loss of generality assume that $f(p) = 0$. In terms of these local co-ordinates Taylor's theorem states that

$$f(\mathbf{r}(u,v)) = \frac{1}{2}\left(f_{uu}(p)u^2 + 2f_{uv}(p)uv + f_{vv}(p)v^2\right) + \text{ higher order terms.}$$

Further, by the spectral theorem, we can may rotate the uv-plane so that

$$f(\mathbf{r}(u,v)) = \lambda u^2 + \mu v^2 + \text{ higher order terms,}$$

By assuming u and v to be suitably small we note that f has the same type of critical point as

$$g(\mathbf{r}(u,v)) = \lambda u^2 + \mu v^2.$$

This is (a) a minimum if $\lambda, \mu > 0$, (b) a maximum if $\lambda, \mu < 0$, (c) a saddle point if $\lambda\mu < 0$.

Now

$$\begin{aligned}(\operatorname{grad}_X g)(p) \cdot \mathbf{r}_u &= \frac{\mathrm{d}}{\mathrm{d}t} g(\mathbf{r}(t,0)) = 2\lambda t = 2\lambda u; \\ (\operatorname{grad}_X g)(p) \cdot \mathbf{r}_v &= \frac{\mathrm{d}}{\mathrm{d}t} g(\mathbf{r}(0,t)) = 2\mu t = 2\mu v.\end{aligned}$$

Taking $\mathbf{r}_u$ as the reference field and recalling that the parameterization is conformal, this means that the angle ψ between $\mathbf{r}_u$ and $\operatorname{grad}_X g$ satisfies

$$(\cos\psi, \sin\psi) = \frac{(\lambda u, \mu v)}{\sqrt{\lambda^2 u^2 + \mu^2 v^2}}.$$

A suitably small closed curve $\lambda^2 u^2 + \mu^2 v^2 = r^2$ around the point $p = \mathbf{r}(0,0)$ can be parameterized by

$$u = \frac{r}{|\lambda|}\cos t, \qquad v = \frac{r}{|\mu|}\sin t,$$

giving

$$\cos\psi = \frac{\lambda}{|\lambda|}\cos t, \qquad \sin\psi = \frac{\mu}{|\mu|}\sin t.$$

- Minimum: $\lambda, \mu > 0$ so that $\psi = t$ and the index is 1
- Maximum: $\lambda, \mu < 0$ so that $\psi = \pi + t$ and the index is 1.
- Saddle: $\lambda < 0 < \mu$ so that $\psi = \pi - t$ and the index is -1.
- Saddle: $\lambda > 0 > \mu$ so that $\psi = 2\pi - t$ and the index is -1.

■

Theorem 5.19 *Given a Morse function f on a smooth, orientable surface X then*

$$\chi(X) = \#\left(\textit{maxima}\right) - \#\left(\textit{saddles}\right) + \#\left(\textit{minima}\right).$$

Proof. Apply the Poincaré-Hopf theorem to the vector field $\operatorname{grad}_X(f)$, taking note of the previous proposition. ■

Example 5.20 *The function $f(x,y) = \cos\pi x + \cos\pi y$ has period 2 in both the x and y variables. $f(x,y)$ descends to a well-defined smooth function $\tilde{f}(x,y)$ on $\mathbb{R}^2/(2\mathbb{Z})^2$ which is diffeomorphic to the torus $\mathbb{T}$. $\tilde{f}(x,y)$ has a maximum at $(0,0)$, saddle points at $(1,0)$ and $(0,1)$ and a minimum at $(1,1)$. Hence*

$$\chi(\mathbb{T}) = 1 - 2 + 1 = 0.$$

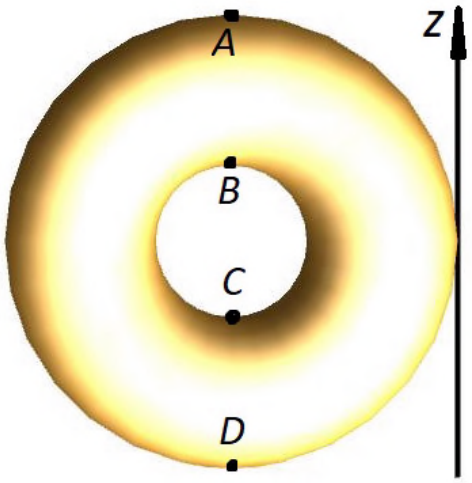


Figure 5.6 – height function on a torus

Example 5.21 *Consider the height function z on the torus $\mathbb{T}$ as depicted in Figure 5.6. There is a maximum at the top of the torus (point A), a minimum at the bottom of the torus (point D) and two saddle points at points B and C. Hence*

$$\chi(\mathbb{T}) = 1 - 2 + 1 = 0.$$

Remark 5.22 *Theorem 5.19 is part of a broader subject called Morse theory, a subject within differential topology which relates differentiable functions on a surface to the surface's topology. It is named after Marston Morse (1892-1977) who first wrote on the subject in 1925.*

Revisiting the example of the height function on a torus, consider the sets

$$X_h = \{(x, y, z) \in \mathbb{T} \mid z \leqslant h\}.$$

Note that the topology of these sets only changes as h achieves the value of one of the critical points' heights. In fact Morse showed that two such sets X_h and X_k would have the same 'homotopy type' if no critical height lay between h and k. This notion of homotopy equivalence is a type of topological equivalence, though weaker than that of being homeomorphic. Further Morse showed how the topology of X_h changes as h passes through a critical height. When h passes through a maximum (at A) or a minimum (at D) a 2-cell (a disc) is attached to the set, but when h passes through a saddle point (at B and C) a 0-cell (a point) is adjoined to the set.

6. HYPERBOLIC SURFACES

6.1 Models for the Hyperbolic Plane

Definition 6.1 *(a) The* ***hyperbolic plane*** $\mathbb{H}$ *is the geometric surface formed from the upper half-plane*

$$\mathbb{H} = \{z \in \mathbb{C} \mid \operatorname{Im} z > 0\},$$

endowed with the first fundamental form

$$\frac{\mathrm{d}x^2 + \mathrm{d}y^2}{y^2}.$$

This is the ***Poincaré's half-plane model.***

(b) We showed earlier (Example 3.54) that the Gaussian curvature of $\mathbb{H}$ *equals* -1*. We also showed in Example 4.8 that the geodesics of* $\mathbb{H}$ *are the half-lines perpendicular to the real axis and the semicircles that meet the real axis at right angles. Note that there is a unique geodesic between any two points of* $\mathbb{H}$*.*

(c) As $E = G$ *in the above first fundamental form then angles are measured in* $\mathbb{H}$ *in the same way that they are in* $\mathbb{C}$*.*

(d) In Method 2 of Example 4.8, we showed that, given real numbers a, b, c, d *such that* $ad - bc = 1$*,*

$$w(z) = \frac{az + b}{cz + d}$$

is a bijective isometry of $\mathbb{H}$*.*

The Möbius map

$$w = \frac{z - i}{z + i}, \qquad z = \frac{i(1 + w)}{1 - w}.$$

takes $\mathbb{H}$ conformally to the disc

$$\mathbb{D} = \{z \in \mathbb{C} \mid |z| < 1\}.$$

If we assign a first fundamental form to $\mathbb{D}$ in such a way that the Möbius map is an isometry then $\mathbb{D}$ is a second model for the hyperbolic plane known as **Poincaré's disc model.** Again we note the above first fundamental form can be rewritten in terms of z as

$$\frac{-4\,|\mathrm{d}z|^2}{(z - \bar{z})^2}.$$

Applying the change of variable we note that

$$|\mathrm{d}z|^2 = \left|\frac{(1 - w) + (1 + w)}{(1 - w)^2}\right|^2 |\mathrm{d}w|^2 = \frac{4\,|\mathrm{d}w|^2}{|1 - w|^4}$$

and

$$\begin{aligned}(z-\bar{z})^2 &= \left(\frac{i(1+w)}{1-w}+\frac{i(1+\bar{w})}{1-\bar{w}}\right)^2 \\ &= -\left(\frac{(1+w)(1-\bar{w})+(1-w)(1+\bar{w})}{(1-w)(1-\bar{w})}\right)^2 \\ &= .-4\left(\frac{1-w\bar{w}}{(1-w)(1-\bar{w})}\right)^2 \\ &= -4\frac{\left|1-|w|^2\right|^2}{|1-w|^4}\end{aligned}$$

Hence

$$\frac{-4|\mathrm{d}z|^2}{(z-\bar{z})^2}=\frac{-16|\mathrm{d}w|^2}{|1-w|^4}\times\frac{|1-w|^4}{-4\left|1-|w|^2\right|^2}=\frac{4|\mathrm{d}w|^2}{\left|1-|w|^2\right|^2}.$$

Proposition 6.2 *(a)* ***Poincaré's disc model*** *for the hyperbolic plane is the disc*

$$\mathbb{D}=\{z\in\mathbb{C}\mid |z|<1\}$$

endowed with the first fundamental form

$$\frac{4|\mathrm{d}z|^2}{\left|1-|z|^2\right|^2}.$$

(b) The Gaussian curvature equals -1 *and the geodesics are the diameters and the circular arcs that meet the unit circles in right angles. Angles are measured in* $\mathbb{D}$ *the same way they are in* $\mathbb{H}$*.*

(c) The orientation-preserving isometries of $\mathbb{D}$ *take the form*

$$f(z)=\frac{e^{i\theta}(z-a)}{1-\bar{a}z}$$

where $|a|<1$ *and* $\theta\in\mathbb{R}$*.*

Proof. (a) was proved in the discussion previous to this proposition and (b) follows as the map w is an isometry and conformal map from $\mathbb{H}$ to $\mathbb{D}$ which preserves curvature, geodesics and angles.

To prove (c) we first need to show that the circle $|z|=1$ maps to itself; if $|z|=1$ then

$$|f(z)|=\left|\frac{e^{i\theta}(z-a)}{1-\bar{a}z}\right|=\left|\frac{\bar{z}(z-a)}{1-\bar{a}z}\right|=\left|\frac{1-a\bar{z}}{1-\bar{a}z}\right|=1.$$

And as $f(a)=0$ then f maps $\mathbb{D}$ bijectively onto $\mathbb{D}$. Also f is orientation-preserving as it is holomorphic. Further

$$|\mathrm{d}f|=\left|\frac{(1-\bar{a}z)+\bar{a}(z-a)}{(1-\bar{a}z)^2}\right||\mathrm{d}z|=\left|\frac{1-|a|^2}{(1-\bar{a}z)^2}\right||\mathrm{d}z|$$

and then

$$\begin{aligned}
\frac{4\left|\mathrm{d}f\right|^2}{\left|1-\left|f(z)\right|^2\right|^2} &= 4\left|\mathrm{d}z\right|^2\left|\frac{1-|a|^2}{(1-\bar{a}z)^2}\right|^2\left|1-\left|\frac{z-a}{1-\bar{a}z}\right|^2\right|^{-2} \\
&= 4\left|\mathrm{d}z\right|^2\left|\frac{1-|a|^2}{(1-\bar{a}z)^2}\right|^2\left|\frac{(1-\bar{a}z)^2}{|1-\bar{a}z|^2-|z-a|^2}\right|^2 \\
&= \frac{4\left|\mathrm{d}z\right|^2\left(1-|a|^2\right)^2}{\left(1+a\bar{a}z\bar{z}-z\bar{z}-a\bar{a}\right)^2} \\
&= \frac{4\left|\mathrm{d}z\right|^2\left(1-|a|^2\right)^2}{\left(1-|a|^2\right)^2\left(1-|z|^2\right)^2} \\
&= \frac{4\left|\mathrm{d}z\right|^2}{\left|1-|z|^2\right|^2},
\end{aligned}$$

showing f is an isometry of $\mathbb{D}$. Further these are *all* the isometries of $\mathbb{D}$. Given an isometry g of $\mathbb{D}$ then, by setting $a = g(0)$ and choosing θ appropriately, we note $f^{-1} \circ g$ is an isometry which sends 0 to 0 and the interval $(0, 1)$ to itself. For orientation, distance and angles to be preserved, it follows that $f^{-1} \circ g$ is the identity and hence $g = f$. ■

Example 6.3 *(a) Let $0 < r < 1$. Find the distance in $\mathbb{D}$ between 0 and r as measured along the real axis.*

(b) Find the distance in $\mathbb{D}$ between $a, b \in \mathbb{D}$.

(c) Deduce a formula for the distance in $\mathbb{H}$ between $p, q \in \mathbb{H}$.

Solution. (a) The distance between 0 and r equals

$$\begin{aligned}
\int_0^r \frac{2\mathrm{d}x}{1-x^2} &= \int_0^r \left(\frac{1}{1-x}+\frac{1}{1+x}\right)\mathrm{d}x \\
&= \left[\log\left(\frac{1+x}{1-x}\right)\right]_0^r \\
&= \log\left(\frac{1+r}{1-r}\right) \\
&= 2\tanh^{-1} r.
\end{aligned}$$

(b) Given points $a, b \in \mathbb{D}$ then the Möbius map

$$\frac{e^{i\theta}(z-a)}{1-\bar{a}z}$$

is an isometry of $\mathbb{D}$ which takes a to 0 and for an appropriate choice of θ takes b to the positive real axis. Hence the distance between a and b equals

$$d_{\mathbb{D}}(a,b) = 2\tanh^{-1}\left|\frac{b-a}{1-\bar{a}b}\right|,$$

as measured along the geodesic between them.
(c) Recall that the map $(z-i)/(z+i)$ is an isometry between $\mathbb{H}$ and $\mathbb{D}$. So given $p, q \in \mathbb{H}$ the distance between them equals

$$\begin{aligned} d_{\mathbb{H}}(p.q) &= d_{\mathbb{D}}\left(\frac{p-i}{p+i}, \frac{q-i}{q+i}\right) \\ &= 2\tanh^{-1}\left|\frac{\left(\frac{q-i}{q+i}\right)-\left(\frac{p-i}{p+i}\right)}{1-\left(\frac{\bar{p}+i}{\bar{p}-i}\right)\left(\frac{q-i}{q+i}\right)}\right| \\ &= 2\tanh^{-1}\left|\frac{(\bar{p}-i)\left((q-i)(p+i)-(p-i)(q+i)\right)}{(p+i)\left((q+i)(\bar{p}-i)-(\bar{p}+i)(q-i)\right)}\right| \\ &= 2\tanh^{-1}\left|\frac{-2ip+2iq}{2i\bar{p}-2iq}\right| \\ &= 2\tanh^{-1}\left|\frac{q-p}{q-\bar{p}}\right|. \end{aligned}$$

■

Remark 6.4 *Poincare's models for the hyperbolic plane date to 1882. There were other models for the hyperbolic plane, most notably one due to Eugenio Beltrami (1868) and Felix Klein (1871). This model again uses the open unit disc, the geodesics are the line segments in the disc, but the model is not conformal with both distance and angle being measured in a non-Euclidean fashion.*

6.2 Hyperbolic geometry and trigonometry

We have yet to show that the hyperbolic distance function $d_{\mathbb{D}}$ is a metric. Certainly $d_{\mathbb{D}}(z,w) \geqslant 0$ and $d_{\mathbb{D}}(z,w) = 0$ if and only if $z = w$. Also symmetry follows as for $z, w \in \mathbb{D}$ then $|w-z| = |z-w|$ and $|1-\bar{z}w| = |1-\bar{w}z|$ as they are conjugates of one another. As there is an isometry of $\mathbb{D}$ taking any of a triangle's vertices to 0, the triangle inequality follows from:

Proposition 6.5 *For $z, w \in \mathbb{D}$,*

$$d_{\mathbb{D}}(z,w) \leqslant d_{\mathbb{D}}(0,z) + d_{\mathbb{D}}(0,w),$$

with equality if and only if z/w is real and negative.

This is turn will follow from:

Proposition 6.6 ***(Hyperbolic cosine rule)*** *Consider a hyperbolic triangle with vertices $0, z, w$. Write*

$$a = d_{\mathbb{D}}(0,z), \qquad b = d_{\mathbb{D}}(0,w), \qquad c = d_{\mathbb{D}}(z,w)$$

and angle C at 0. Then

$$\cosh c = \cosh a \cosh b - \sinh a \sinh b \cos C.$$

Note that if the approximations $\cosh x \approx 1 + x^2/2$ *and* $\sinh x \approx x$ *apply then the hyperbolic cosine rule approximates to*

$$c^2 = a^2 + b^2 - 2ab\cos C$$

which is the usual Euclidean cosine rule.

Proof. (Of the triangle inequality) This follows from the cosine rule as $\cos C \geqslant -1$ so that

$$\begin{aligned} \cosh c &\leqslant \cosh a \cosh b + \sinh a \sinh b \\ &= \cosh(a+b). \end{aligned}$$

As cosh is strictly increasing for non-negative arguments then

$$c \leqslant a + b.$$

Further we only have equality when $\cos C = -1$ and $C = \pi$ in which case w/z is real and negative. ■

Proof. (Of the cosine rule) Without loss of generality we may assume that z is positive. Then

$$z = \tanh\frac{d_{\mathbb{D}}(0,z)}{2}, \qquad w = e^{iC}\tanh\frac{d_{\mathbb{D}}(0,w)}{2}.$$

By the hyperbolic tangent half-angle formulae (the hyperbolic 't-formulae') then

$$\cosh d_{\mathbb{D}}(0,z) = \frac{1+z^2}{1-z^2}, \qquad \cosh d_{\mathbb{D}}(0,w) = \frac{1+|w|^2}{1-|w|^2},$$

and by definition

$$\tanh\frac{1}{2}d_{\mathbb{D}}(z,w) = \left|\frac{w-z}{1-zw}\right|.$$

So

$$\begin{aligned} \cosh d_{\mathbb{D}}(a,b) &= \frac{|1-zw|^2 + |w-z|^2}{|1-zw|^2 - |w-z|^2} \\ &= \frac{(1+z^2)\left(1+|w|^2\right) - 2\left(zw + z\bar{w}\right)}{(1-z^2)\left(1-|w|^2\right)} \\ &= \left(\frac{1+z^2}{1-z^2}\right)\left(\frac{1+|w|^2}{1-|w|^2}\right) - \left(\frac{2z}{1-z^2}\right)\left(\frac{2|w|}{1-|w|^2}\right)\left(\frac{w+\bar{w}}{2|w|}\right) \\ &= \cosh a \cosh b - \sinh a \sinh b \cos C, \end{aligned}$$

recalling

$$\sinh a = \frac{2z}{1-z^2}, \qquad \sinh b = \frac{2|w|}{1-|w|^2}, \qquad \cos\arg w = \frac{\operatorname{Re} w}{|w|}.$$

■

Remark 6.7 *(**Dual hyperbolic cosine rule**) As with spherical geometry, there is a second 'dual' cosine rule which has no equivalent in Euclidean geometry. In the hyperbolic case this reads as*

$$\cos C = -\cos A \cos B + \sin A \sin B \cosh c.$$

Proposition 6.8 *(**Hyperbolic sine rule**) For a hyperbolic triangle in $\mathbb{D}$ with angles A, B, C and sides a, b, c then*

$$\frac{\sin A}{\sinh a} = \frac{\sin B}{\sinh b} = \frac{\sin C}{\sinh c}.$$

Solution. This is Sheet 3, Part C, Exercise 1. ■

Example 6.9 *In $\mathbb{D}$ a circle of radius R has area $4\pi \sinh^2(R/2)$ and circumfernece $2\pi \sinh R$. Note that, for small values of R, these formulae approximate to πR^2 and $2\pi R$.*

Solution. The circle of radius R, centred on the origin in $\mathbb{D}$, corresponds to the circle $|z| = \tanh(R/2)$. So its interior has area equalling

$$\begin{aligned}
&\iint_{|z| \leqslant \tanh(R/2)} \sqrt{EG - F^2}\, \mathrm{d}x\, \mathrm{d}y \\
= &\int_0^{\tanh(R/2)} \int_0^{2\pi} \frac{4}{(1-r^2)^2}\, r\, \mathrm{d}r\, \mathrm{d}\theta \\
= &\ 2\pi \left[\frac{2}{1-r^2}\right]_0^{\tanh(R/2)} \\
= &\ 4\pi \left(\cosh^2\left(\frac{R}{2}\right) - 1\right) \\
= &\ 4\pi \sinh^2\left(\frac{R}{2}\right).
\end{aligned}$$

The circle can be parameterized as $x = \tanh(R/2)\cos t$ and $y = \tanh(R/2)\sin t$, so it has circumference

$$\begin{aligned}
&\int_{t=0}^{2\pi} \sqrt{E\dot{x}^2 + G\dot{y}^2}\, \mathrm{d}t \\
= &\ \tanh(R/2) \int_{t=0}^{2\pi} \frac{2}{1 - \tanh^2(R/2)}\, \mathrm{d}t \\
= &\ \frac{4\pi \tanh(R/2)}{\operatorname{sech}^2(R/2)} \\
= &\ 4\pi \sinh\left(\frac{R}{2}\right) \cosh\left(\frac{R}{2}\right) \\
= &\ 2\pi \sinh R.
\end{aligned}$$

■

Remark 6.10 *It follows that a circle in $\mathbb{D}$ is a Euclidean circle – this is because isometries of $\mathbb{D}$ are Möbius maps which send circles to circles. However the centre of a hyperbolic circle will not in general coincide with the Euclidean centre. Further it is also now clear, that through three non-collinear points there need not be a circle in $\mathbb{D}$. For example the points $0, 1-\varepsilon$ and $i(1-\varepsilon)$ are not collinear – as they lie on two different diameters – but any hyperbolic (and so Euclidean) circle passing through these three points will not entirely lie in $\mathbb{D}$.*

Theorem 6.11 ***(Lambert's Formula)*** *Given a triangle T in $\mathbb{D}$, bounded by geodesics, its area equals*

$$\pi - \alpha - \beta - \gamma$$

where α, β, γ are the three angles.

Proof. The local Gauss-Bonnet theorem states

$$\int_\gamma k_g \, \mathrm{d}s + \iint_R K \, \mathrm{d}A + \sum_{i=1}^{n} \alpha_i = 2\pi.$$

As $k_g = 0$ on a geodesic, and recalling that $K = -1$, we find

$$0 - A + (\pi - \alpha) + (\pi - \beta) + (\pi - \gamma) = 2\pi$$

which rearranges to the required result. Note that the maximal area π can be achieved by having all three vertices on the boundary of $\mathbb{D}$. Such a triangle is called *triasymptotic*. ■

Proposition 6.12 ***(The angle of parallelism)*** *Let l be a line in $\mathbb{D}$ and P be a point at distance $d > 0$ from l. This distance d is measured along a perpendicular from P to a point O on l. Then a line through P will meet l if the angle the line makes with OP is less than*

$$\Pi(d) = \sin^{-1} \operatorname{sech} d.$$

$\Pi(d)$ is known as the angle of parallelism.

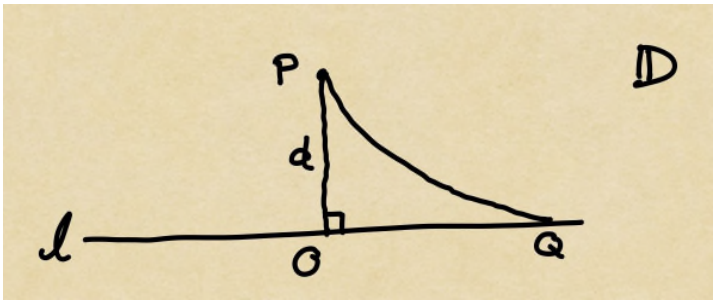


Figure 6.1 – the angle of parallelism

Proof. Without loss of generality we can take l to be the real axis and P to be on the positive imaginary axis, so represented by the complex number $\tanh(d/2)$. The point O is then the origin. Say that a second line passes through P, making an angle θ, and intersects l at the point Q. By the sine rule

$$\sin\theta = \frac{\sinh d_{\mathbb{D}}(O, Q)}{\sinh d_{\mathbb{D}}(P, Q)}.$$

And by the hyperbolic cosine rule we have

$$\cosh d_{\mathbb{D}}(P,Q) = \cosh d_{\mathbb{D}}(O,Q)\cosh d.$$

Eliminating $d_{\mathbb{D}}(O,Q)$ we find

$$\begin{aligned}\sin\theta &= \frac{1}{\sinh d_{\mathbb{D}}(P,Q)}\sqrt{\frac{\cosh^2 d_{\mathbb{D}}(P,Q)}{\cosh^2 d}-1}\\ &= \sqrt{\coth^2 d_{\mathbb{D}}(P,Q)\,\mathrm{sech}^2 d - \mathrm{cosech}^2 d_{\mathbb{D}}(P,Q)}.\end{aligned}$$

Now as $d_{\mathbb{D}}(P,Q)\to\infty$ then $\coth^2 d_{\mathbb{D}}(P,Q)\to 1$ and $\mathrm{cosech} d_{\mathbb{D}}(P,Q)\to 0$. So the limiting case for when we can solve for θ is when

$$\sin\theta = \mathrm{sech} d.$$

This θ is the required formula for the angle of parallelism. ■

6.3 Compact Hyperbolic Surfaces

As commented earlier, a closed geometric surface with constant curvature $K=-1$ is necessarily a torus with genus $g>1$. This is a consequence of the global Gauss-Bonnet theorem. It is not hard to appreciate how such a surface might be made from a polygon. The canonical identification space for such a torus is a $4g$-gon with edges identified as

$$a_1 a_2 a_1^{-1} a_2^{-1}\cdots a_{2g-1}a_{2g}a_{2g-1}^{-1}a_{2g}^{-1}.$$

When forming a topological surface the edges are identified by homeomorphisms and nothing further needs to be required. However, to create a hyperbolic surface we need to begin with a polygon that is already a geometric surface with boundary – so we take a regular polygon from the hyperbolic plane – and then the identifications need to be made using isometries. Further, the internal angles of the polygon, that are identified as the same vertex, need to add up to a whole angle.

Example 6.13 *Consider a regular octagon in $\mathbb{D}$, such as the one sketched as in Figure 6.2.*

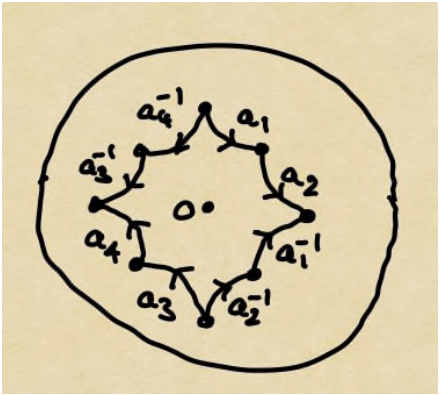


Figure 6.2 – an identified octagon from $\mathbb{D}$

For any d in the range $0 < d \leqslant \infty$ such an octagon can be constructed which has the vertices at hyperbolic distance d from the origin. Recall for a regular hyperbolic octagon

$$-\textit{area of octagon} = \iint_R K \,\mathrm{d}A = 8\beta - 6\pi,$$

where β is the common internal angle. As $d \to 0$ then $\beta \to 3\pi/4$ and when $d \to \infty$ then $\beta \to 0$. For any such octagon, a topological surface can be formed by identifying the eight edges as depicted and the eight vertices are then all identified to the same vertex. However to form a geometric surface we need to identify the edges with isometries and need $\beta = \pi/4$ so that the internal angles sum to a whole angle. But this is possible for a unique choice of d as β is a decreasing function of d.

For this particular choice of d, the global Gauss-Bonnet theorem tells us that

$$-4\pi = 2\pi\chi(X) = \iint_X K \,\mathrm{d}A = -\textit{area of } X.$$

So the surface's area equals 4π. More generally, for $g > 1$, a regular $2g$-gon can be identified to form a hyperbolic surface of genus g.

Example 6.14 *Find the distance of the vertices from the origin of the octagon in Figure 6.2 and the complex numbers representing those vertices.*

Solution. We have noted that the internal angles of the octagon are $\pi/4$, so the octagon can be naturally divided into 8 isosceles triangles with angles $A = \pi/4, B = C = \pi/8$. The equal length sides are then b and c. The dual hyperbolic cosine rule states

$$\cos C = -\cos A \cos B + \sin A \sin B \cosh c$$

and so

$$\cosh c = \frac{\cos\frac{\pi}{8} + \cos\frac{\pi}{4}\cos\frac{\pi}{8}}{\sin\frac{\pi}{4}\sin\frac{\pi}{8}} = \cot\left(\frac{\pi}{8}\right)\left(\frac{1+\cos(\pi/4)}{\sin(\pi/4)}\right).$$

Noting $\cot(\pi/8) = 1 + \sqrt{2}$ and $\sin(\pi/4) = \cos(\pi/4) = 1/\sqrt{2}$ we then have

$$\cosh c = \left(1+\sqrt{2}\right)^2.$$

The right-most vertex is then at $z = \tanh(c/2)$. By the hyperbolic tangent half-angle formulae we have

$$\frac{1+z^2}{1-z^2} = \left(1+\sqrt{2}\right)^2.$$

Solving for z we find $z = 2^{-1/4}$. Thus the vertices of the octagon are $2^{-1/4}\omega^k$ where $\omega = e^{i\pi/4}$ and $0 \leqslant k \leqslant 7$. ■

Example 6.15 *(**A non-orientable hyperbolic surface**) We can create a hyperbolic surface X, that is homeomorphic to a torus of genus 3, by identifying the edges of a regular dodecagon*

in $\mathbb{D}$, centred on the origin, in the canonical way. X can then be embedded in $\mathbb{R}^3$ – as a smooth surface – and in such a way that is symmetric about the origin. The antipodal map $\sigma(x, y, z) = (-x, -y, -z)$ is a self-inverse diffeomorphism of X with $X/\langle\sigma\rangle$ being non-orientable – for example, a symmetric band within the torus would become a Möbius band. But $X/\langle\sigma\rangle$ can be endowed with the hyperbolic structure that X has. Specifically $X/\langle\sigma\rangle$ is the sphere with four cross-caps.

Example 6.16 *(**Pseudosphere**) The tractoid is a hyperbolic surface, but not a complete one as its geodesics cannot be extended indefinitely. Omitting one meridian, it is isometric to the semi-infinite strip $(0, 2\pi) \times (1, \infty)$ and we see that the geodesic $x = \pi$ cannot be extended. The completion of the tractoid is the pseudosphere $\mathbb{H}/\Gamma$ where Γ is the group of isometries generated by $z \mapsto z + 2\pi$. (See Figure 6.3.)*

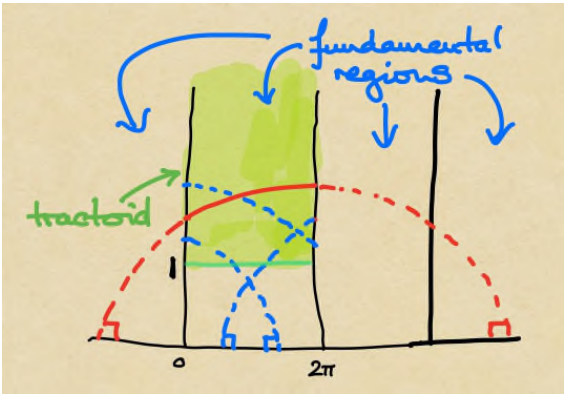


Figure 6.3 – the pseudosphere

The pseudosphere is complete but not compact.

The subject of hyperbolic surfaces is treated in detail in Stillwell. I include here just some of the key theorems.

Theorem 6.17 *(**Killing-Hopf theorem**) (Stillwell, p.111) Each complete, connected hyperbolic surface is of the form $\mathbb{H}/\Gamma$ where Γ is a discontinuous group of isometries of $\mathbb{H}$ which acts freely on $\mathbb{H}$.*

*To say that Γ is **discontinuous** means that no orbit (of Γ's action) has a limit point.*

*To say that Γ acts **freely** means that if $g.x = x$ for $g \in \Gamma$ and $x \in \mathbb{H}$ then g is the identity.*

Definition 6.18 *Given a free, discontinuous action of Γ on $\mathbb{H}$ then a **fundamental region** $R \subseteq \mathbb{H}$ for the action is a region of $\mathbb{H}$ which contains a representative of each orbit such that the interior of R contains at most one element of an orbit. Thus $\mathbb{H}/\Gamma$ is represented by R with some identifications on its boundary.*

Theorem 6.19 *(Stillwell, p.123) A hyperbolic surface is formed from a hyperbolic polygon provided*

(i) the edges are pairwise identified with isometries and

(ii) the sum of the internal angles, around vertices that are identified together, equals a whole angle.

Theorem 6.20 *(Stillwell, p.130) For any compact hyperbolic surface $\mathbb{H}/\Gamma$ there is a polygonal fundamental region for Γ.*

Theorem 6.21 *(Poincaré, 1882 – Stillwell p.180) A compact polygon P, satisfying the edge and angle conditions (i) and (ii) above, is a fundamental region for the group Γ generated by the edge-pairing transformations of P.*

Example 6.22 *Find the edge-pairing isometry which identifies the edge a_1 with a_1^{-1} as in Figure 6.2,*

Solution. Recall that the orientation-preserving isometries of $\mathbb{D}$ take the form

$$f(z) = e^{i\theta}\frac{z-a}{1-\bar{a}z}$$

where $a \in \mathbb{D}$ and $0 \leqslant \theta < 2\pi$. The vertices of the octagon are $\alpha\omega^k$ where $\alpha = 2^{-1/4}$ and $\omega = e^{i\pi/4}$. The map

$$f_1(z) = \frac{z-\alpha i}{1+\alpha i z}$$

takes the rear of edge a_1 to the origin and the front of the edge to

$$\begin{aligned}
\frac{\alpha\omega - \alpha i}{1+\alpha i\alpha\omega} &= \alpha\left(\frac{\frac{1+i}{\sqrt{2}} - i}{(1+i)/2}\right) \\
&= \frac{\alpha}{\sqrt{2}}\left(1+(1-\sqrt{2})i\right)(1-i) \\
&= \frac{\alpha}{\sqrt{2}}\left(\left(2-\sqrt{2}\right) - \sqrt{2}i\right) \\
&= \alpha\left(\left(\sqrt{2}-1\right) - i\right)
\end{aligned}$$

which has argument

$$\tan^{-1}\left(\frac{-1}{\sqrt{2}-1}\right) = -\tan^{-1}\left(\sqrt{2}+1\right) = -\frac{3\pi}{8}.$$

Thus the function $g_1(z) = e^{3\pi i/8}f_1(z)$ takes a_1 to the positive real axis from 0. We can argue the same to find a function $g_2(z)$ which takes the rear of a_1^{-1} to the origin with image along the positive real axis (details omitted). The edge-pairing isometry we are seeking is then $g_2^{-1} \circ g_1$.
■

7. RIEMANN SURFACES

7.1 Examples

Definition 7.1 *A Riemann surface is a connected topological surface S with a holomorphic atlas. A holomorphic atlas is a collection of charts $\{\varphi_i \colon U_i \to V_i\}$ where $V_i \subseteq \mathbb{C}$ and the transition maps are biholomorphic – that is holomorphic bijections with holomorphic inverses. Note that not all authors assume Riemann surfaces to be connected.*

The definition of holomorphic maps between Riemann surfaces can then be made in a like manner to how we defined smooth maps between smooth surfaces. Recall that holomorphic maps are much more rigid than smooth functions – as, for example, becomes apparent with the identity theorem. The issue of classifying Riemann surfaces up to biholomorphism (the correct notion of isomorphism here) is much more subtle than in the smooth case, with a great variety in the possible complex structures that a certain topological type can be endowed with. On these points we note:

Proposition 7.2 *Riemann surfaces are orientable.*

Proof. The transition maps are holomorphic, with non-zero derivatives, and so are orientation-preserving. ■

And the following is also true – left to Sheet 4, Exercise 3.

- A holomorphic function on a compact Riemann surface is constant.

Example 7.3 *(**The complex plane**) $\mathbb{C}$ is a Riemann surface. The identity map $\iota \colon \mathbb{C} \to \mathbb{C}$ forms a holomorphic atlas by itself.*

Example 7.4 *(**Riemann mapping theorem**) Every simply connected, non-empty proper open subset $U \subsetneq \mathbb{C}$ is biholomorphic to an open half-plane.*

Example 7.5 *(**Annuli**) An annulus*

$$A = \{z \in \mathbb{C} \mid r_1 < |z| < r_2\}$$

is not homeomorphic to an open half-plane – as it is not simply connected – and so is not biholomorphic to it. All such annuli are diffeomorphic to one another, but there is a famous theorem of complex analysis which shows two such annuli are biholomorphic if they have the same ratio of radii r_2/r_1.

Example 7.6 *(**The Riemann sphere**) The Riemann sphere can also be thought of as the complex projective line or the extended complex plane $\mathbb{C}_\infty = \mathbb{C} \cup \{\infty\}$. We can provide a holomorphic atlas with two charts*

$$\begin{aligned} U_1 &= \mathbb{C}_\infty \setminus \{\infty\}, \qquad V_1 = \mathbb{C}, \qquad \varphi_1(z) = z; \\ U_2 &= \mathbb{C}_\infty \setminus \{0\}, \qquad V_2 = \mathbb{C}, \qquad \varphi_2(z) = z^{-1}. \end{aligned}$$

Then $\varphi_1(U_1 \cap U_2) = \mathbb{C}\setminus\{0\}$ *and*

$$\varphi_2 \circ \varphi_1^{-1} = \frac{1}{z}$$

which is biholomorphism between $\varphi_1(U_1 \cap U_2)$ *and* $\varphi_2(U_1 \cap U_2)$.

Example 7.7 ***(Meromorphic maps)*** *In light of the previous example, a meromorphic map on a Riemann surface can be considered as a holomorphic map to the Riemann sphere. Say that* $f\colon \mathbb{C} \to \mathbb{C}_\infty$ *is a meromorphic function – so holomorphic except for finitely many poles. When* $f(z)$ *is finite then* $\varphi_1 \circ f(z) = f(z)$ *is holomorphic and when* $f(z) = \infty$ *then* $\varphi_2 \circ f(z) = 1/f(z)$ *which is holomorphic with a zero of the same order as the order of the pole of* $f(z)$.

Example 7.8 ***(Uniformization theorem)*** *Every simply-connected Riemann surface is biholomorphic to one of (a) the Riemann sphere, (b) the complex plane, (c) the open, upper half-plane.*

Example 7.9 ***(Complex structures on the torus)*** *Consider the lattice*

$$\Lambda = \mathbb{Z} \oplus \omega\mathbb{Z}$$

where $\omega \in \mathbb{C}\setminus\mathbb{R}$. *Then* $\mathbb{C}/\Lambda$ *is homeomorphic to a torus (Figure 7.1) and naturally inherits the structure of a Riemann surface from* $\mathbb{C}$.

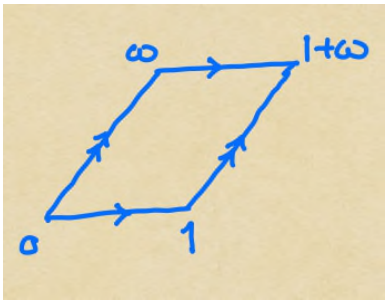


Figure 7.1 – parallelogram in lattice

But in general these complex tori are not biholomorphic to one another. It turns out (Kirwan p.141) that two complex tori $\mathbb{C}/\Lambda$ *and* $\mathbb{C}/\tilde{\Lambda}$ *are biholomorphic if and only if* $\Lambda = a\tilde{\Lambda}$ *or equally if* $J(\Lambda) = J(\tilde{\Lambda})$ *where*

$$J(\Lambda) = \frac{g_2^3}{g_2^3 - 27g_3^2}, \qquad g_2 = 60 \sum_{w\in\Lambda\setminus\{0\}} \frac{1}{w^4}, \qquad g_3 = 140 \sum_{w\in\Lambda\setminus\{0\}} \frac{1}{w^6}.$$

In particular, there are uncountably many biholomorphism classes of complex structures on a torus.

Example 7.10 ***(The Riemann surface of*** $\sqrt{z}$***)*** *The affine surface is*

$$\Sigma = \left\{(z, w) \in \mathbb{C}^2 \mid w^2 = z\right\}.$$

Topologically this is not complicated. The map $w \mapsto (w^2, w)$ *is a homeomorphism from* $\mathbb{C}$ *to* Σ *and so* Σ *is topologically a plane. When we include a point at infinity it is topologically a sphere.*

However if we wish to understand Σ as the Riemann surface of the multifunction of $\sqrt{z}$ then we need to define $\sqrt{z}$ on a cut plane. This was already discussed in section 0.4 with regard to the topology of Σ. Here we focus on Σ as a Riemann surface. For $(z,w) \in \Sigma$ where $z \neq 0$ then we can use either z or w as a local holomorphic co-ordinate, but around $(0,0)$ we need to use w as the local co-ordinate. This is because 0 is a branch point of $\sqrt{z}$, something we will discuss in more detail later. There is a similar issue if we want to include the point at infinity which is again a branch point of $\sqrt{z}$.

Example 7.11 *(**A non-singular cubic**) The cubic*

$$y^2 = x(x-1)(x-\lambda), \qquad \lambda \neq 0, 1,$$

is a non-singular (affine) cubic Σ in $\mathbb{C}^2$. It has a projectivized version $\bar{\Sigma}$ with equation

$$y^2 z = x(x-z)(x-\lambda z)$$

which has a single point at infinity $[z\colon x\colon y] = [0\colon 0\colon 1]$. The above equations are known as Legendre form.

For $x \neq 0, 1, \lambda, \infty$ there are two values of y. We make cuts between 0 and 1 and between λ and ∞. We can then define two holomorphic branches on this cut plane (Figure 7.2)

$$\pm\sqrt{x(x-1)(x-\lambda)}$$

and most points of Σ take the form $\left(x, \sqrt{x(x-1)(x-\lambda)}\right)$ or $(x, -\sqrt{x(x-1)(x-\lambda)})$. Let

$$\begin{aligned}
\Sigma_+ &= \left\{\left(x, \sqrt{x(x-1)(x-\lambda)}\right) \mid x \in \text{ cut plane}\right\} \subseteq \mathbb{C}^2;\\
\Sigma_- &= \left\{\left(x, -\sqrt{x(x-1)(x-\lambda)}\right) \mid x \in \text{ cut plane}\right\} \subseteq \mathbb{C}^2.
\end{aligned}$$

The points that are missing from $\bar{\Sigma}$ are the point at infinity and those points associated with a value of x on the cuts.

*Near the point $x = 0$ then $y^2 \approx \lambda x$. So as we move around the point $x = 0$ there is a sign change in the branches, just as there is with the standard holomorphic branches of the square root. This explains why the tabs A and B are so aligned as in Figure 7.3. The same argument can be made for the second cut. Including the point at infinity, we see that $\bar{\Sigma}$ is homeomorphic to a torus. Such an algebraic curve is called an **elliptic curve**.*

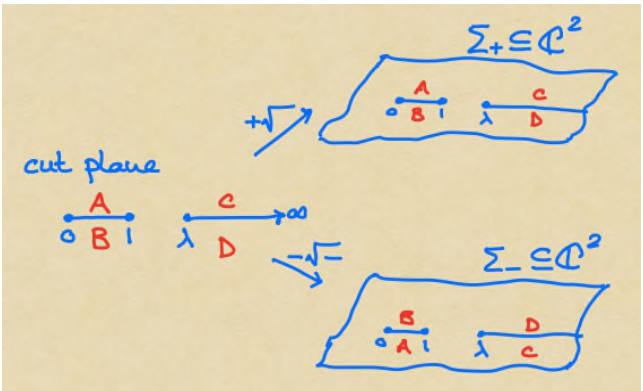


Figure 7.2 – branches on cut plane

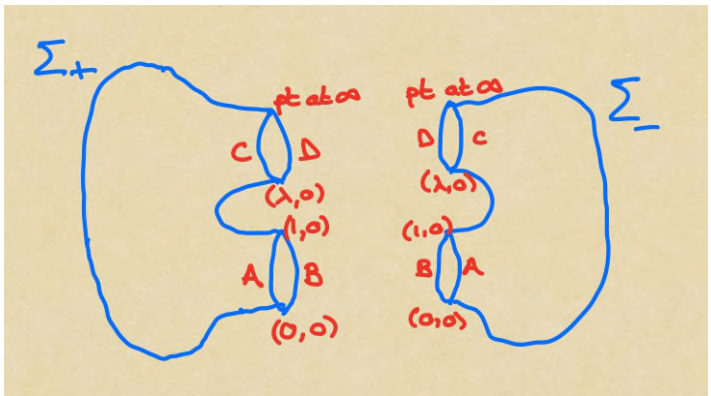


Figure 7.3 – gluing Σ_+ and Σ_-

It can be shown that the values

$$\lambda, \quad \lambda^{-1}, \quad 1-\lambda, \quad \frac{1}{1-\lambda}, \quad \frac{\lambda}{\lambda-1}, \quad \frac{\lambda-1}{\lambda},$$

lead to biholomorphic complex tori.

Example 7.12 *(**Hyperelliptic curves**) We can extend the analysis of the previous Riemann surface to curves with equations*

$$y^2 = (x-\alpha_1)(x-\alpha_2)\cdots(x-\alpha_n)$$

where $n \geqslant 2$ and the α_i are distinct. In section 0.4 we saw that when $n = 2$ the Riemann surface is a sphere and in the previous example obtain a torus when $n = 3$. When we increase n by 2 then we need to introduce a further cut in the plane and add a further handle to the surface. So the surface has genus $(n-1)/2$ when n is odd and genus $(n-2)/2$ when n is even.

However the projective curve is singular at its point at infinity – see the following remark. So we cannot just assign a complex structure on the surface, inherited from complex projective plane. Away from the branch points α_i we can use either x or y as the local holomorphic co-ordinate. At the branch points we need to use y as in the previous example with the square root. For now, assume $n = 2k$ is even and we introduce at ∞ the following co-ordinates

$$X = \frac{1}{x}, \qquad Y = \frac{y}{x^k}.$$

The defining equation now reads as

$$\left(\frac{1}{X}-\alpha_1\right)\left(\frac{1}{X}-\alpha_2\right)\cdots\left(\frac{1}{X}-\alpha_{2k}\right) = \left(\frac{Y}{X^k}\right)^2$$

which rearranges to

$$(1-\alpha_1 X)(1-\alpha_2 X)\cdots(1-\alpha_{2k}X) = Y^2.$$

Near infinity, when $X \approx 0$, we have $Y^2 \approx 1$ and so we can compactify Σ with two points at infinity associated with $(X,Y) = (0,1)$ and $(X,Y) = (0,-1)$. Near these two points X is a local holomorphic co-ordinate. A similar approach can be taken when n is odd, with just one point being needed at infinity.

This shows that complex structures can be assigned to a torus of any genus $g \geqslant 2$; the above Riemann surfaces are called hyperelliptic curves. All complex structures (up to biholomorphism) on the torus of a given genus can be studied via a classifying space known as a 'moduli space'. All complex structures for genus 2 arise as hyperelliptic curves but for $g > 2$ the hyperelliptic curves are not generic within the moduli space.

Remark 7.13 *(Off-syllabus) The above non-singular cubic is the zero set in the complex projective plane of the function*

$$F(x,y,z) = y^2 z - x(x-z)(x-\lambda z).$$

A singular point of the cubic is any point satisfying $F = \nabla F = 0$ and a quick check shows none exist on the cubic. The cubic's complex structure is inherited from the ambient projective space.

On Sheet 0, Exercise 5, we met the cubic

$$y^2 - x(x-1)^2 = 0$$

and a check shows this curve to be singular at $(x, y) = (1, 0)$. *This singularity is called a node. The complex projective curve is topologically a pinched torus.*

When $n > 3$, *the hyperelliptic curve*

$$G(x, y, z) = y^2 z^{n-2} - (x - \alpha_1 z)(x - \alpha_2 z) \cdots (x - \alpha_n z) = 0$$

can be checked to have a singularity at its point at infinity, where $x = 0, y = 1, z, = 0$. *This is why we complete the complex structure of the hyperelliptic curves in a different manner.*

7.2 The Riemann-Hurwitz formula

Proposition 7.14 ***(Local form of a holomorphic map)*** *For any holomorphic map* $f\colon S \to R$ *between Riemann surfaces, with* $f(s) = r$, *we can choose local complex co-ordinates around* $s \in S, r \in R$ *in terms of which* f *is the map*

$$f\colon D \to D \quad \textit{given by} \quad f(z) = z^n.$$

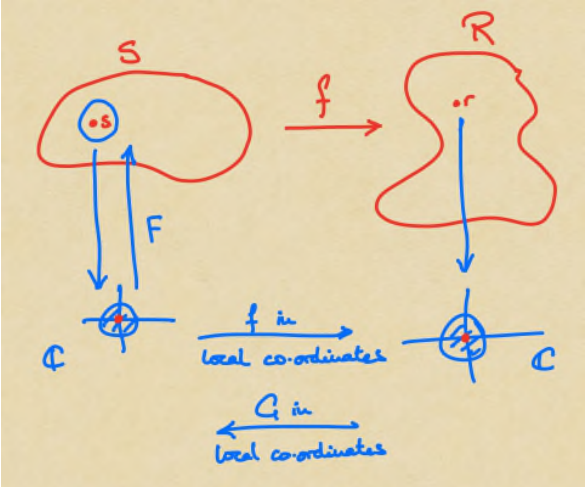


Figure 7.4 – local form of a holomorphic map

Proof. We can assume that, by translating if necessary, the local co-ordinates are chosen with s, r corresponding to $0 \in \mathbb{C}$, so that $f(0) = 0$ in local co-ordinates. The Taylor series for $f(z)$ therefore begins

$$f(z) = a_n z^n + a_{n+1} z^{n+1} + a_{n+2} z^{n+2} + \cdots$$

where $n \geqslant 1$ and $a_n \neq 0$ is the first non-zero coefficient. A holomorphic nth root of f is then defined near 0 with

$$f(z)^{1/n} = a_n^{1/n} z + \cdots .$$

The derivative of this nth root at 0 is $a_n^{1/n} \neq 0$ and so, by the inverse function theorem, there is a local holomorphic inverse G defined near 0. Then $G(0) = 0$ and

$$f(G(z))^{1/n} = z.$$

We can now change co-ordinates in the domain using the local biholomorphism G. To be explicit, if F is the original parameterization, defined locally near $s \in S$, then the new parameterization is $F \circ G$, defined near 0. The new local expression for f becomes

$$z \mapsto f(G(z)) = z^n.$$

■

Corollary 7.15 *(**Open Mapping Theorem**) A non-constant holomorphic map between Riemann surfaces is an open map. That is the image of an open set is open.*

Proof. This is left to Sheet 4, Exercise 3. ■

Definition 7.16 *Let $f\colon S \to R$ be a holomorphic map between Riemann surfaces and let $s \in S$. Then there are local co-ordinates around s and $f(s)$ such that f has the form $z \mapsto z^n$. The number n is called the **valency** of f at s and is written $v_f(s)$. Geometrically this is the number of solutions to the equation $f(z) = w$ for small $w \neq 0$. Thus the valency does not depend on the choice of co-ordinates.*

*If $n > 1$ then we say that $f(s)$ is a **branch point** and s is a **ramification point**. Note that s is a ramification point if and only if $f'(s) = 0$ in local co-ordinates.*

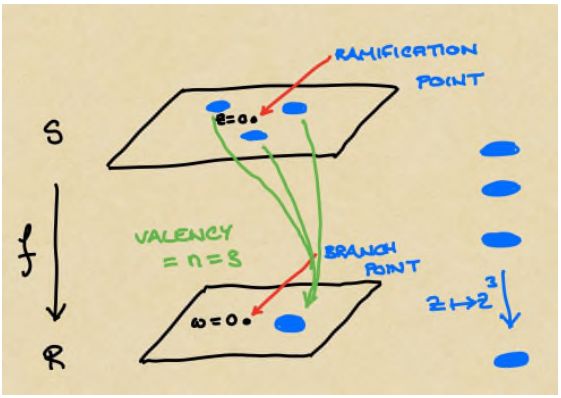


Figure 7.5 – local picture at a ramification point

Lemma 7.17 *A holomorphic function on a compact Riemann surface has finitely many ramification points.*

Proof. Locally $f(z) = z^n$ and so $f'(z) = nz^{n-1} \neq 0$ for $z \neq 0$. Hence the ramification points of f form a discrete set and hence a finite set, as S is (sequentially) compact. ■

Example 7.18 *Consider the map $f(z) = z^2$ from the Riemann sphere to itself. The ramification points of f are 0 and ∞ with the valency equalling 2 at each point.*

Proposition 7.19 *(**Degree of a map**) Given a non-constant holomorphic map $f\colon S \to R$ between compact Riemann surfaces, the **degree** of f is defined to be*

$$\deg(f) = \sum_{s \in f^{-1}(r)} v_f(s),$$

for any $r \in R$. This definition is independent of the choice of r.

Proof. Since S is compact, and f is not constant, then $f^{-1}(r)$ is finite. Choose local co-ordinates around each point p of $f^{-1}(r)$ such that around each point f is given by $z \mapsto z^{v_f(p)}$. Without loss of generality we may assume that the domains D_p of these local co-ordinates are disjoint, that their images are the same open neighbourhood V of r and that

$$f^{-1}(V) = \bigcup_{p \in f^{-1}(r)} D_p.$$

It follows that

$$\sum_{p \in f^{-1}(r)} v_f(p)$$

is the number of distinct solutions to the equation $f(z) = w$ for $w \in V \backslash \{r\}$. Thus this sum is locally constant and since R is connected then this sum is constant on R. ■

Corollary 7.20 *For all points $r \in R$, except branch points, there are precisely $\deg(f)$ points in S which map to r.*

Example 7.21 *For the earlier example of $f(z) = z^2$ on the Riemann sphere, we have $\deg(f) = 2$. For $r \neq 0, \infty$ then $f^{-1}(r)$ consists of the two squares roots of r, each of which has valency 1. For $r = 0, \infty$ then $f^{-1}(r)$ is a singleton with valency 2.*

Theorem 7.22 ***(Riemann-Hurwitz Formula)*** *For any non-constant holomorphic function $f \colon S \to R$ between compact Riemann surfaces*

$$\chi(S) = \deg(f)\chi(R) - \sum_{\substack{\text{ramification} \\ \text{points } p}} (v_f(p) - 1)\,.$$

*The sum on the RHS is referred to as the **branching index** of f.*

Example 7.23 *For our earlier map $f(z) = z^2$ on the Riemann sphere, the above equation holds as $\chi(R) = \chi(S) = 2 = \deg(f)$ and $v_f(0) = v_f(\infty) = 2$. So we arrive at*

$$2 = 2 \times 2 - (2 - 1) - (2 - 1)\,,$$

which is true.

Proof. Pick a triangulation for R so that the branch points belong to the vertices of the triangulation. We want the pre-image to yield a triangulation of S. So we subdivide the triangles into smaller triangles if necessary, so that each triangle $T \subseteq R$ lies inside an open set $V \subseteq R$ small enough so that $f^{-1}(V) \to V$ can be written in the usual local form on each connected component $U \subseteq S$ of $f^{-1}(V)$. If the local form of $f \colon U \to V$ is $z \mapsto z$, then the pre-image of T is a triangle. But if the local form is $z \mapsto z^n$ where $n > 1$, then $f^{-1}(T)$ consists of n triangles.

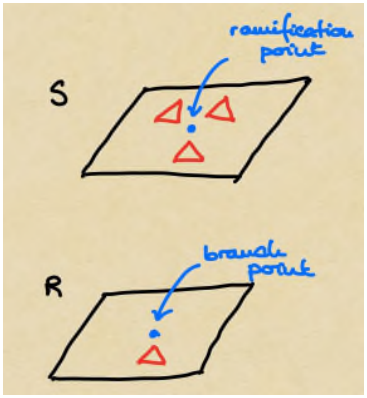


Figure 7.6a

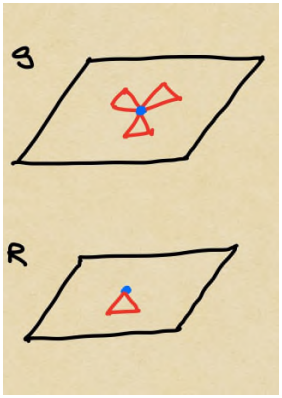


Figure 7.6b

If the vertex is not a branch point then these n triangles have n times as many vertices, edges and faces as T does (Figure 7.6a). Thus they contribute n times to $\chi(S)$. However if the vertex is a branch point, then the n triangles meet at the corresponding ramification point. We have lost $v_f(p) - 1$ vertices (Figure 7.6b). So the subdivision of S satisfies

$$V(S) = \deg(f)V(R) - \sum_{\substack{\text{ramification} \\ \text{points } p}} (v_f(p) - 1)$$

and $E(S) = \deg(f)E(R)$ and $F(S) = \deg(f)F(R)$. The result follows. ■

Example 7.24 *Suppose that $g(S) < g(R)$. Then any holomorphic map $f\colon S \to R$ is constant.*

Solution. Assume for a contradiction that f is not constant. We have

$$\chi(S) = \deg(f)\chi(R) - B$$

where $B \geqslant 0$ is the branching index. We then have

$$2 - 2g(S) = \deg(f)\,(2 - 2g(R)) - B$$

Rearranging gives

$$2g(R) - 2g(S) = (\deg(f) - 1)\,(2 - 2g(R)) - B.$$

The LHS is positive but, as $g(R) \geqslant 1$, the RHS is at most zero. ■

7.3 Meromorphic and Elliptic Functions

Recall that we can identify a meromorphic function f on S with a holomorphic map $f\colon S \to \mathbb{C}_\infty$ provided that f is not identically ∞. We can note here that f has equal number of poles and zeros on S, counting multiplicities, as that number is just $\deg(f)$. The following result may come as a surprise, but is a first intimation of connections with algebraic geometry.

Theorem 7.25 *(a) The meromorphic functions $\mathbb{C}_\infty \to \mathbb{C}_\infty$ are rational functions.*
(b) The biholomorphisms of $\mathbb{C}_\infty$ are the Möbius maps.

Remark 7.26 *Note that as a consequence of (a) the meromorphic functions on $\mathbb{C}_\infty$ form a field, the function field of $\mathbb{C}_\infty$.*

Proof. (a) As a consequence of the identity theorem, zeros of holomorphic functions are isolated. As $\mathbb{C}_\infty$ is compact, this means that $f\colon \mathbb{C}_\infty \to \mathbb{C}_\infty$ has finitely many zeros $z_1, \ldots, z_n$ and finitely many poles $p_1, \ldots, p_m$ in $\mathbb{C}$ – we will attend to a possible zero or pole at ∞ in a moment. Let $a_1, \ldots, a_n$ and $b_1, \ldots, b_m$ be the orders of the zeros and the poles and set

$$g(z) = \prod_{j=1}^{n} (z - z_j)^{a_j} \prod_{k=1}^{m} (z - p_k)^{-b_k}.$$

Then f/g is meromorphic and it no longer has any zeros or poles in $\mathbb{C}$.

By the earlier comment, f/g has an equal number of zeros and poles and cannot have both at ∞. This means that $f/g\colon \mathbb{C}_\infty \to \mathbb{C}$ is holomorphic and so, by Sheet 4, Exercise 3(ii), is constant. Hence $f = \text{constant} \times g(z)$ is rational.

(b) By part (a) a biholomorphism is a rational function. As a biholomorphism is bijective, then there can be at most one zero and one pole. If that zero and pole have any multiplicity then then rational function will not be injective locally and so the numerator and denominator must have degree one and be independent of one another. That is, the biholomorphism must be a Möbius map. ■

Recall now that we have met complex tori both as the Riemann surface of the multifunction

$$\sqrt{(z - e_1)(z - e_2)(z - e_3)} \qquad e_1, e_2, e_3 \text{ are distinct,}$$

and also as the quotient

$$\frac{\mathbb{C}}{\mathbb{Z}\omega_1 \oplus \mathbb{Z}\omega_2} \qquad \frac{\omega_2}{\omega_1} \notin \mathbb{R}.$$

We have not, thus far, made any connection between these two definitions.

Definition 7.27 *An **elliptic function** is a meromorphic function f on $\mathbb{C}$ which is doubly periodic – that is f is periodic in two independent directions ω_1 and ω_2.*

Definition 7.28 *The **Weierstrass $\wp$-function** (or **Weierstrass elliptic function**) associated with the lattice $\Lambda = \mathbb{Z}\omega_1 \oplus \mathbb{Z}\omega_2$ equals*

$$\wp(z) = \frac{1}{z^2} + \sum_{0 \neq \omega \in \Lambda} \left(\frac{1}{(z - \omega)^2} - \frac{1}{\omega^2} \right).$$

Remark 7.29 *It is impossible to find a meromorphic function on $\mathbb{C}/\Lambda$ with only one simple pole as this would mean $\mathbb{C}/\Lambda$, a torus, is homeomorphic to $\mathbb{C}_\infty$, a sphere. If instead we take a meromorphic function with a double pole, then we can WLOG assume it to be at 0 with Laurent coefficient $c_{-2} = 1$. In order to make the function doubly periodic then we might expect to include the sum*

$$\sum_{0 \neq \omega \in \Lambda} \frac{1}{(z - \omega)^2}$$

but this is unfortunately divergent. However the inclusion of second term (which is itself divergent) makes the infinite sum convergent.

I list below some important facts about $\wp(z)$ but they are, in the main, turgid to prove. The important role of $\wp(z)$ is in providing a link between our two descriptions of complex tori.

- On a domain bounded away from the poles, $\wp(z)$ converges to an elliptic function.
- In the fundamental parallelogram $\{\alpha\omega_1 + \beta\omega_2 \mid 0 \leqslant \alpha, \beta \leqslant 1\}$, $\wp$ has a pole of order 2 at $z = 0$.
- $\wp\colon \mathbb{C}/\Lambda \to \mathbb{C}_\infty$ has degree 2.
- $\wp' = 0$ at $\omega_1/2, \omega_2/2$ and $(\omega_1 + \omega_2)/2$.
- In the fundamental parallelogram $\wp$ has ramification points at $0, \omega_1/2, \omega_2/2$ and $(\omega_1 + \omega_2)/2$.
- The valencies at the ramification points are each 2.
- The branch points of $\wp$ are denoted

$$e_1 = \wp\left(\frac{\omega_1}{2}\right), \qquad e_2 = \wp\left(\frac{\omega_2}{2}\right), \qquad e_3 = \wp\left(\frac{\omega_1 + \omega_2}{2}\right), \qquad \infty = \wp(0).$$

- $\wp$ satisfies the differential equation

$$\wp'(z)^2 = 4\left(\wp(z) - e_1\right)\left(\wp(z) - e_2\right)\left(\wp(z) - e_3\right).$$

Finally we have the following theorem:

Theorem 7.30 *(a) The following is a biholomorphism*

$$\begin{array}{rcl} \mathbb{C}/\Lambda & \to & \left\{(z,w) \in \mathbb{C}^2 \mid w^2 = 4\left(z - e_1\right)\left(z - e_2\right)\left(z - e_3\right)\right\} \cup \{\infty\}; \\ z & \mapsto & \left(\wp(z), \wp'(z)\right). \end{array}$$

(a) It can be shown that the function field of meromorphic functions on the complex torus $\mathbb{C}/\Lambda$ *is* $\mathbb{C}\left(\wp, \wp'\right)$.