Geometry

Cambridge University Mathematical Tripos: Part IB

17th May 2024

Contents

1 Surfaces

1.1 Basic definitions

Definition. A *topological surface* is a topological space Σ such that
  (i) for all points $p \in \Sigma$, there exists an open neighbourhood $p \in U \subset \Sigma$ such that U is homeomorphic to $\mathbb{R}^2$, or a disc $D^2 \subset \mathbb{R}^2$, with its usual Euclidean topology;
  (ii) Σ is Hausdorff and second countable.

Remark. $\mathbb{R}^2$ is homeomorphic to the open disc $D(0,1) = \{x \in \mathbb{R}^2 : \|x\| < 1\}$. Recall that a space X is Hausdorff if two points $p \neq q \in X$ have open neighbourhoods U, V such that $U \cap V = \varnothing$. A space X is *second countable* if it has a countable base; there exists a countable family of open sets U_i, such that every open set is a union of some of the U_i.

Note that subspaces of Hausdorff and second countable spaces are also Hausdorff and second countable. In particular, Euclidean space $\mathbb{R}^n$ is Hausdorff (as $\mathbb{R}^n$ is a metric space) and second countable (consider the set of balls $D(p, q)$ for points p with rational coordinates, and rational radii q). Hence, any subspace of $\mathbb{R}^n$ is implicitly Hausdorff and second countable. These topological requirements are typically not the purpose of considering topological spaces, but they are occasionally technical requirements to prove interesting theorems.

Example. $\mathbb{R}^2$ is a topological surface. Any open subset of $\mathbb{R}^2$ is also a topological surface. For example, $\mathbb{R}^2 \setminus \{0\}$ and $\mathbb{R}^2 \setminus \{(0,0)\} \cup \left\{ \left(0, \frac{1}{n}\right) : n = 1, 2, \dots \right\}$ are topological surfaces.

Example. Let $f : \mathbb{R}^2 \to \mathbb{R}$ be a continuous function. The graph of f, denoted Γ_f, is defined by

$$\Gamma_f = \left\{ (x, y, f(x, y)) : (x, y) \in \mathbb{R}^2 \right\}$$

with the subspace topology when embedded in $\mathbb{R}^3$. Recall that a product topology $X \times Y$ has the feature that $f : Z \to X \times Y$ is continuous if and only if $\pi_x \circ f : Z \to X$ and $\pi_y \circ f : Z \to Y$ are continuous. Hence, any graph $\Gamma \subseteq X \times Y$ is homeomorphic to X if f is continuous. Indeed, the projection π_x projects each point in the graph onto the domain. The function $s : x \mapsto (x, f(x))$ is continuous by the above. In particular, in our case, the graph Γ_f is homeomorphic to $\mathbb{R}^2$, which we know is a surface.

Remark. As a topological surface, Γ_f is independent of the function f. However, we will later introduce more ways to describe topological spaces that will ascribe new properties to Γ_f which do depend on f.

Example. The sphere:
$$S^2 = \left\{ (x, y, z) \in \mathbb{R}^3 : x^2 + y^2 + z^2 = 1 \right\}$$

is a topological surface, when using the subspace topology in $\mathbb{R}^3$. Consider the stereographic projection of S^2 onto $\mathbb{R}^2$ from the north pole $(0, 0, 1)$. The projection satisfies $\pi_+ : S^2 \setminus \{(0,0,1)\}$ and

$$(x, y, z) \mapsto \left(\frac{x}{1-z}, \frac{y}{1-z} \right)$$

Certainly, π_+ is continuous, since we do not consider the point $(0, 0, 1)$ to be in its domain. The inverse map is given by

$$(u, v) \mapsto \left(\frac{2u}{u^2 + v^2 + 1}, \frac{2v}{u^2 + v^2 + 1}, \frac{u^2 + v^2 - 1}{u^2 + v^2 + 1} \right)$$

This is also a continuous function. Hence π_+ is a homeomorphism. Similarly, we can construct the stereographic projection from the south pole, π_-. This is a homeomorphism. Hence, every point in S^2 lies either in the domain of π_+ or π_-, and hence sits in an open set $S^2 \setminus \{(0, 0, 1)\}$ or $S^2 \setminus \{(0, 0, -1)\}$ which is homeomorphic to $\mathbb{R}^2$.

Remark. S^2 is compact by the Heine–Borel theorem; it is a closed bounded set in $\mathbb{R}^3$.

Example. The real projective plane is a topological surface. The group $\mathbb{Z}/2$ acts on S^2 by homeomorphisms via the *antipodal map* $a \colon S^2 \to S^2$, mapping $x \mapsto -x$. There exists a homeomorphism $\mathbb{Z}/2$ to the group $\mathrm{Homeo}(\mathbb{S}^2)$ of homeomorphisms of S^2, by mapping $1 + \mathbb{Z} \mapsto a$. We now define the real projective plane to be the quotient of S^2 given by identifying every point x with its image $-x$ under a.

$$\mathbb{RP}^2 = {}^{S^2}\!/_{\mathbb{Z}/2} = {}^{S^2}\!/_{\sim}; \quad x \sim a(x)$$

> **Lemma.** $\mathbb{RP}^2$ bijects with the set of straight lines in $\mathbb{R}^3$ through the origin.

Proof. Any line through the origin intersects S^2 exactly in a pair of antipodal points $x, -x$. Similarly, pairs of antipodal points uniquely define a line through the origin. $\square$

> **Lemma.** $\mathbb{RP}^2$ is a topological surface.

Proof. We must check that $\mathbb{RP}^2$ is Hausdorff since it is constructed by a quotient, not a subspace. If X is a space and $q \colon X \to Y$ is a quotient map, then by definition $V \subset Y$ is open if and only if $q^{-1}(V) \subset X$ is open. If $[p] \neq [q] \in \mathbb{RP}^2$, then $\pm p, \pm q \in S^2$ are distinct antipodal pairs. We can therefore construct distinct open discs around p, q in S^2, and their antipodal images. These uniquely define open neighbourhoods of $[p], [q]$, which are disjoint.

Similarly, we can check that $\mathbb{RP}^2$ is second countable. We know that S^2 is second countable, so let $\mathcal{U}$ be a countable base for the topology on S^2. Without loss of generality, we can assert that for all sets $U \in \mathcal{U}$, we have $-U \in \mathcal{U}$. Let $\overline{\mathcal{U}}$ be the family of open sets in $\mathbb{RP}^2$ of the form $q(U) \cup q(-U)$ for $U \in \mathcal{U}$, where q is the quotient map. Now, if $V \subseteq \mathbb{RP}^2$ is open, then by definition $q^{-1}(V)$ is open in S^2 hence $q^{-1}(V)$ contains some $U \in \mathcal{U}$ and hence contains $U \cup (-U)$. Hence $\overline{\mathcal{U}}$ is a countable base for the quotient topology on $\mathbb{RP}^2$.

Finally, let $p \in S^2$ and $[p] \in \mathbb{RP}^2$ its image. Let $\overline{D}$ be a small (contained in an open hemisphere) closed disc, which is a neighbourhood of $p \in S^2$. The quotient map restricted to $\overline{D}$, written $q|_{\overline{D}} \colon \overline{D} \to q(\overline{D}) \subset \mathbb{RP}^2$, is a continuous function from a compact space to a Hausdorff space. Further, q is injective on $\overline{D}$ since the disc was contained entirely in a single hemisphere. The topological inverse function theorem states that a continuous bijection from a compact space to a Hausdorff space is a homeomorphism. So $q|_{\overline{D}}$ is a homeomorphism from $\overline{D}$ to $q(\overline{D})$. This then induces the homeomorphism $q|_D$ from the open disc $D = \overline{D}^{\circ}$ to $q(D)$. So by construction, $[p] \in q(D)$; it has an open neighbourhood in $\mathbb{RP}^2$ which is homeomorphic to an open disc, concluding the proof. $\square$

Example. Let S^1 be the unit circle in $\mathbb{C}$, and then we define the torus to be the product space $S^1 \times S^1$, with the subspace topology from $\mathbb{C}^2$ (which is identical to the product topology).

Lemma. The torus is a topological surface.

*Proof.* Consider the map $e : \mathbb{R}^2 \to S^1 \times S^1$ defined by

$$(s, t) \mapsto \left(e^{2\pi i s}, e^{2\pi i t}\right)$$

Note that this induces a map $\hat{e}$ from $\mathbb{R}^2 \big/ \mathbb{Z}^2$, since e is constant under translations by $\mathbb{Z}^2$.

$$
\begin{array}{ccc}
\mathbb{R}^2 & \xrightarrow{\;e\;} & S^1 \times S^1 \\
{\scriptstyle q}\big\downarrow & \nearrow{\scriptstyle\hat{e}} & \\
\mathbb{R}^2 \big/ \mathbb{Z}^2 & &
\end{array}
$$

Under the quotient topology given by the quotient map q, $\mathbb{R}^2 \big/ \mathbb{Z}^2$ is a topological space. The map $[0, 1]^2 \to \mathbb{R}^2 \to \mathbb{R}^2 \big/ \mathbb{Z}^2$ is surjective, so $\mathbb{R}^2 \big/ \mathbb{Z}^2$ is compact. So $\hat{e}$ is a continuous map from a compact space to a Hausdorff space, and $\hat{e}$ is bijective, so $\hat{e}$ is a homeomorphism. We already have that $S^1 \times S^1$ is compact and Hausdorff (as a closed and bounded set in $\mathbb{C}^2$), so it suffices to show it is locally homeomorphic to $\mathbb{R}^2$. Let $[p] = q(p) \in S^1 \times S^1$, then we can choose a small disc $\overline{D}(p)$ such that $\overline{D}(p) \cap \left(\overline{D}(p) + (n, m)\right) = \varnothing$ for nonzero $(n, m) \in \mathbb{Z}^2$. Hence $e|_{\overline{D}(p)}$ is injective and $q|_{\overline{D}(p)}$ is injective. Now, restricting to the open disc as before, we can find an open disc neighbourhood of $[p]$. Since $[p]$ was chosen arbitrarily, $S^1 \times S^1$ is a topological surface. $\qquad\square$

Example. Let P be a planar Euclidean polygon, with oriented edges. We will pair the edges, and without loss of generality we will assume that paired edges have the same Euclidean length.
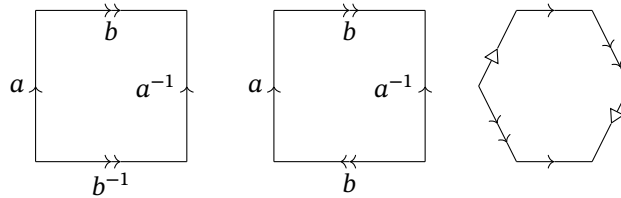
We can assign letter names to each edge pair, and denote a polygon by the sequence of edges found when traversing in a clockwise orientation. The edge pair name is inverted if the edge is traversed in the reverse direction. Note the difference between the annotations on the first two shapes above, due to the reversed direction of the edge. If two edges $e, \hat{e}$ are paired, this defines a unique Euclidean isometry from e to $\hat{e}$ respecting the orientation, which will be written $f_{e\hat{e}} : e \to \hat{e}$. The set of all such functions generate an equivalence relation on the polygon, identifying paired edges with each other.

Lemma. $P \big/ \sim$, with the quotient topology, is a topological surface.

Example. Consider the torus, defined here as $T^2 = [0, 1]^2 \big/ \sim$. Let P be the polygon $[0, 1]^2$. If p is in the interior of P, then construct a sufficiently small disc that lies entirely within the interior. The quotient map is injective on the closure of the disc and is a homeomorphism on its interior.

Let p be on an edge, but not a vertex. Let us say without loss of generality that $p = (0, y_0) \sim (1, y_0)$. Let δ be sufficiently small that the closed half-discs U, V centred on p with radius δ do not intersect any vertices. Then we define a map from the union of the two half-discs to the disc $B(0, \delta) \subseteq \mathbb{R}^2$ via $(x, y) \mapsto (x, y - y_0)$ or $(x, y) \mapsto (x - 1, y - y_0)$, which will be a bijective map. Recall the gluing lemma from Analysis and Topology: that if $X = A \cup B$ is a union of closed subspaces, and $f : A \to Y$, $g : B \to Y$ are continuous and $f|_{A \cap B} = g|_{A \cap B}$, they define a continuous map on X. Let f_U, f_V be the maps on the half-discs U, V. By the definition of the quotient topology, $q \circ f_U$ and $q \circ f_V$ are also continuous. On the overlapping area, the functions $q \circ f_U$ and $q \circ f_V$ agree. Hence, by the gluing lemma, we can construct a function $f : U \cup V \to B(0, \delta)$. We can show that this is a homeomorphism using the usual process: pass to the closed disc, apply the topological inverse function theorem, then apply the result to the interior. If $[p] \in T^2$ lies on the image of an edge in $[0, 1]^2$, it has indeed a neighbourhood homeomorphic to a disc.

Now it suffices to consider points p on a vertex. All four vertices of the square are identified to the same point in the torus. A neighbourhood of each vertex can be identified with a quarter-disc in $\mathbb{R}^2$. We can repeatedly apply the gluing lemma to construct the whole disc $B(0, \delta) \subseteq \mathbb{R}^2$ and complete the argument as before.

Thus, $[0, 1]^2 \big/ {\sim}$ is a topological surface.

We can generalise this proof to an arbitrary planar Euclidean polygon P, such as the hexagon above. The equivalence relation $x \sim f_{e\hat{e}}(x)$ induces an equivalence relation on the vertices of P, by considering the images of the vertices under all $f_{e\hat{e}}$. However, it is not necessarily the case that an equivalence class of vertices contains exactly four vertices, so quarter-discs are not necessarily applicable. Again, there are three types of point:

- interior points, for which a neighbourhood not intersecting the boundary is chosen;

- points on edges, for which a corresponding point exists and two half-discs can be glued to form the neighbourhood; and

- points on vertices. For this case, all vertices of the polygon have a neighbourhood which is a sector of a circle. Let there be r vertices in a given equivalence class. Let α be the sum of the angles of the sectors in a given class. Any sector can be identified with a given sector in the disc $B(0, \delta) \subseteq \mathbb{R}^2$, which we will choose to have angle α/r. Then, we can glue each sector together in $\mathbb{R}^2$, compatibly with the orientations of the edges and arrows, inducing a neighbourhood which is locally homeomorphic to a disc. If $r = 1$, we have an equivalence class comprising a single vertex, which gives a single sector. For r to be one, the two edges attached to this vertex must be paired and have the same direction (either both inwards or outwards from the vertex). This quotient space is simply a cone, which is homeomorphic to a disc as required.

We can also show that the quotient space is Hausdorff and second countable. By construction, two distinct points in the quotient space can be separated by open neighbourhoods by selecting a sufficiently small radius such that the discs considered in the derivation above are disjoint. For second countability, consider

- discs in the interior of P with rational centres and radii;

- for each edge of P, consider an isometry $e \to [0, \ell]$ where ℓ is the length of e, taking discs on e which are centred at rational values in $[0, \ell]$; and

- for each vertex, consider discs centred at these vertices with rational radii.

Example. Given topological surfaces Σ_1, Σ_2 we can remove an open disc from each and glue the resulting circles. Explicitly, we form a quotient relation on the disjoint union of the surfaces with the discs removed. This process is known as forming the *connect sum* of the surfaces, written $\Sigma_1 \# \Sigma_2$. Typically, the information about where the discs were removed from is discarded when considering the connect sum. The connect sum of two topological surfaces is a topological surface.

Example. Consider the following octagon.
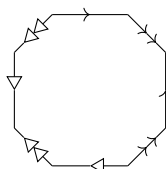
The associated quotient space $P/_\sim$ can be seen to be homeomorphic to a surface with two holes, known as a double torus. All vertices are identified as the same vertex in the quotient space. We can cut the octagon along a diagonal, leaving two topological surfaces which are homeomorphic to a torus.
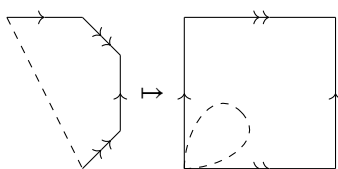
Thus, the connect sum of the two half-octagons are the connect sum of two toruses.

Example. Consider the following square.
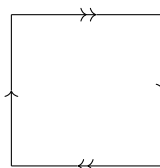
This is homeomorphic to the real projective plane $\mathbb{RP}^2$. This is because we identify points on the boundary with their antipodes, when interpreting the square as the closed disc $B(0, 1)$. The real projective plane was constructed by identifying points on the unit sphere with their antipodes. Thus, we can construct a homeomorphism by considering only points in the upper hemisphere (taking antipodes as required), and then orthographically projecting onto the xy plane. Under this transformation, points on the boundary are identified with their antipodes as required.

1.2 Subdivisions

Definition. A *subdivision* of a compact topological surface Σ comprises
  (i) a finite subset $V \subseteq \Sigma$ of vertices;
 (ii) a finite subset $E = \{e_i : [0, 1] \to \Sigma\}$ which are continuous injections and pairwise disjoint except perhaps at the endpoints;
(iii) such that each connected component of the complement of $V \cup E$ in Σ is homeomorphic to an open disc, and each such component will be called a face. In particular, the bound-

7

ary of each face has boundary inside the union of the edges and the vertices.

We say that a subdivision is a *triangulation* if each closed face (closure of a face) contains exactly three edges, and two closed faces meet either at exactly one edge or not at any edges.

Example. A cube displays a subdivision of S^2. A tetrahedron displays a triangulation of S^2.

Example. We can display subdivisions of surfaces constructed from polygons.
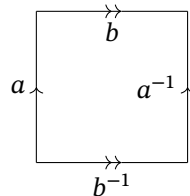
This is a subdivision of a torus with one edge, two edges, and one face. We can construct additional subdivisions of a torus, for example:
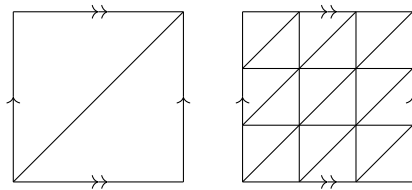
The first of these examples is not a triangulation, since the two faces meet in more than one edge. The second is a triangulation.

Remark. The following is a very degenerate subdivision of S^2.
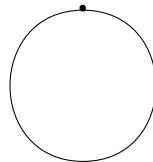
This has one vertex, no edges, and one face.

1.3   Euler classification

Definition. The *Euler characteristic* of a subdivision is

$$\#V - \#E + \#F$$

Theorem.   (i) Every compact topological surface has a subdivision (and indeed triangulations).

  (ii) The Euler characteristic is invariant under choice of subdivision, and is topologically invariant.

Hence, we might say that a surface has a particular Euler characteristic, without referring to subdivisions. We write this $\chi(\Sigma)$.

No proof will be given.

Example. The Euler characteristic of S^2 is $\chi(S^2) = 2$. For the torus, $\chi(T^2) = 0$. If Σ_1, Σ_2 are compact surfaces, then the connect sum $\Sigma_1 \# \Sigma_2$ can be constructed by removing a face of a triangulation, then gluing together the boundary circles (three edges) in a way that matches the edges. Then the connect sum inherits a subdivision, and we can find that it has Euler characteristic $\chi(\Sigma_1 \# \Sigma_2) = \chi(\Sigma_1) + \chi(\Sigma_2) - 2$, where the remaining term corresponds to the two faces that were removed; the changes of three vertices and three edges cancel each other. In particular, a surface Σ_g with g holes can be written $\#_{i=1}^{g} T^2$, so $\chi(\Sigma_g) = 2 - 2g$. We call g the *genus* of Σ.

Remark. It is not trivial to prove part (i). For part (ii), note that subdivisions can be converted into triangulations by constructing triangle fans.

Triangulations can be related by local moves, such as

It is easy to check that both of these moves do not change the Euler characteristic. However, it is hard to make this argument rigorous, and it does not give much explanation for why the result is true. In Part II Algebraic Topology, a more advanced definition of the Euler characteristic is given, which admits a more elegant proof.

2 Smooth surfaces

2.1 Charts and atlases

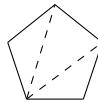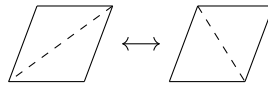Recall that if Σ is a topological surface, any point lies in an open neighbourhood homeomorphic to a disc.

Definition. A pair (U, φ), where U is an open set in Σ and $\varphi : U \to V$ is a homeomorphism to an open set $V \subseteq \mathbb{R}^2$, is called a *chart* for Σ. If $p \in U$, we might say that (U, φ) is a chart for Σ *at* p. A collection of charts whose domains cover Σ is known as an *atlas* for Σ. The inverse $\sigma = \varphi^{-1} : V \to U$ is known as a *local parametrisation* for the surface.

Example. If $Z \subseteq \mathbb{R}^2$ is closed, $\mathbb{R}^2 \setminus Z$ is a topological surface with an atlas containing one chart, $(\mathbb{R}^2 \setminus Z, \phi = \mathrm{id})$.

For S^2, there is an atlas with two charts, which are the two stereographic projections from the poles. We could consider alternative charts, for instance the projection to the yz plane, but this would be insufficient for describing the poles.

Definition. Let (U_i, φ_i) be charts containing the point $p \in \Sigma$, for $i = 1, 2$. Then the map

$$* : \varphi_1(U_1 \cap U_2) \to \varphi_2(U_1 \cap U_2); \quad * = \varphi_2 \circ \varphi_1^{-1}\Big|_{\varphi_1(U_1 \cap U_2)}$$

converts between the corresponding charts, and is called a *transition map*. This is a homeo-morphism of open sets in $\mathbb{R}^2$.

Recall from Analysis and Topology that if $V \subseteq \mathbb{R}^n$ and $V' \subseteq \mathbb{R}^m$ are open, then a continuous map $f : V \to V'$ is called *smooth* if it is infinitely differentiable. Equivalently, it is smooth if partial derivatives of all orders in all variables exist at all points. If $n = m$, then in particular the homeomorphism $f : V \to V'$ is called a *diffeomorphism* if it is smooth and has smooth inverse.

Definition. An *abstract smooth surface* is a topological space Σ together with an atlas of charts (U_i, φ_i) such that all transition maps $\varphi_i \circ \varphi_j^{-1} : \varphi_j(U_i \cap U_j) \to \varphi_i(U_i \cap U_j)$ are diffeomorphisms.

Remark. We could not simply consider a smoothness condition for Σ itself without appealing to atlases, since Σ is an arbitrary topological space and could have almost any topology.

Example. The atlas of two charts with stereographic projections gives S^2 the structure of an abstract smooth surface.

Example. For the torus $T^2 = \mathbb{R}^2 / \mathbb{Z}^2$, we can find charts of all points by choosing sufficiently small discs in $\mathbb{R}^2$ such that they do not intersect any of their non-trivial integer translates. The transition maps for this atlas are all translations of $\mathbb{R}^2$. Hence T^2 inherits the structure of an abstract smooth surface. Explicitly, let us define $e : \mathbb{R}^2 \to T^2$ by $(t, s) \mapsto (e^{2\pi i t}, e^{2\pi i s})$, then consider the atlas

$$\{(e(D_\varepsilon(x, y)), e^{-1} \text{ on this image})\}$$

for $\varepsilon < \frac{1}{3}$. These are charts on T^2, and the transition maps are (restricted to appropriate domains) translations in $\mathbb{R}^2$. Hence T^2, via this atlas, has the structure of an abstract smooth surface.

Remark. The definition of a topological surface is a notion of structure. One can observe a topological space and determine whether it is a topological surface. Conversely, to be an abstract smooth surface is to have a specific set of data; that is, we must provide charts for the surface in order to see that it is indeed an abstract smooth surface.

Definition. Let Σ be an abstract smooth surface, and $f : \Sigma \to \mathbb{R}^n$ be a continuous map. We say that f is *smooth* at $p \in \Sigma$ if, for all charts (U, φ) of p belonging to the smooth atlas for Σ, the map

$$f \circ \varphi^{-1} : \varphi(U) \to \mathbb{R}^n$$

is smooth at $\varphi(p) \in \mathbb{R}^2$.

Remark. Note that the choice of chart and atlas was arbitrary, but smoothness of f at p is independent of the choice of chart, since the transition maps between two such charts are diffeomorphisms.

Definition. Let Σ_1, Σ_2 be abstract smooth surfaces. Then a map $f : \Sigma_1 \to \Sigma_2$ is *smooth* if it is 'smooth in the local charts'. Given a chart (U, φ) at p and a chart (U', ψ) at $f(p)$, both mapping to open subsets of $\mathbb{R}^2$, the map $\psi \circ f \circ \varphi^{-1}$ is smooth at $\varphi(p)$. Smoothness of f does not depend on the choice of chart, provided that the charts all belong to the same atlas.

Definition. Two surfaces Σ_1, Σ_2 are *diffeomorphic* if there exists a homeomorphism $f : \Sigma_1 \to \Sigma_2$ which is smooth and has smooth inverse.

Remark. Often, we convert from a given smooth atlas for an abstract smooth surface Σ to the *maximal compatible* smooth atlas. That is, we consider the atlas with the maximal possible set of charts, all of which have transition maps that are diffeomorphisms. This can be accomplished formally by use of Zorn's lemma.

3 Smooth surfaces in $\mathbb{R}^3$

3.1 Definitions and equivalent characterisations

Recall that if $V \subseteq \mathbb{R}^n$ and $V' \subseteq \mathbb{R}^m$, then $f : V \to V'$ is smooth if it is infinitely differentiable.

Definition. If Z is an arbitrary subset of $\mathbb{R}^n$, we say that a continuous function $f : Z \to \mathbb{R}^m$ is smooth at $p \in Z$ if there exists an open ball $p \in B \subseteq \mathbb{R}^n$ and a smooth map $F : B \to \mathbb{R}^m$ which extends f such that they agree on $B \cap Z$. In other words, f is locally the restriction of a smooth map defined on an open set.

Definition. Let $X \subseteq \mathbb{R}^n$ and $Y \subseteq \mathbb{R}^m$. We say that X and Y are *diffeomorphic* if there exists a continuous function $f : X \to Y$ such that f is a smooth homeomorphism with smooth inverse.

Definition. A *smooth surface in* $\mathbb{R}^3$ is a subspace of $\mathbb{R}^3$ such that for all points $p \in \Sigma$, there exists an open subset $p \in U \subseteq \Sigma$ that is diffeomorphic to an open set in $\mathbb{R}^2$. In other words, for all $p \in \Sigma$, there exists an open ball $p \in B \subseteq \mathbb{R}^3$ such that if $U = B \cap \Sigma$ and there exists a map $f : B \to V \subseteq \mathbb{R}^2$ such that $f|_U : U \to V$ is a homeomorphism, and the inverse map $V \to U \subseteq \Sigma \subseteq \mathbb{R}^3$ is smooth.

Definition. Let $\sigma : V \to U$ where $V \subseteq \mathbb{R}^2$ is open and $U \subseteq \Sigma \subseteq \mathbb{R}^3$ is open in Σ, such that σ is a smooth homeomorphism and $D\,\sigma|_x$ has rank 2 for all $x \in V$. Then σ is called an *allowable parametrisation*. If $\sigma(0) = p$, we say that σ is an allowable parametrisation *near* p.

Theorem. For a subset $\Sigma \subseteq \mathbb{R}^3$, the following are equivalent.
   (a) Σ is a smooth surface in $\mathbb{R}^3$;
   (b) Σ is locally the graph of a smooth function, over one of the three coordinate planes: for all $p \in \Sigma$ there exists an open ball $p \in B \subseteq \mathbb{R}^3$ and an open set $V \subseteq \mathbb{R}^2$ such that

$$\Sigma \cap B = \{(x, y, g(x, y)) : g : V \to \mathbb{R} \text{ smooth}\}$$

   or one of the other coordinate planes;
   (c) Σ is locally cut out by a smooth function: for all $p \in \Sigma$ there exists an open ball $p \in$

$B \subseteq \mathbb{R}^3$ and a smooth function $f : B \to \mathbb{R}$ such that

$$\Sigma \cap B = f^{-1}(0); \quad D f\Big|_x \neq 0$$

for all $x \in B$;

(d) Σ is locally the image of an allowable parametrisation near all points.

Remark. Part (b) implies that if Σ is a smooth surface in $\mathbb{R}^3$, each $p \in \Sigma$ belongs to a chart (U, φ) where φ is (the restriction of) one of the three coordinate plane projections $\pi_{xy}, \pi_{yz}, \pi_{xz}$ from $\mathbb{R}^3$ to $\mathbb{R}^2$. Consider the transition map between two such charts. If the two charts are based on the same projection such as π_{xy}, then the transition map is the identity. If they are based on different projections π_{xy} and π_{yz}, then the transition map is

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

which has inverse

$$(y, z) \mapsto (h(y, z), y, z) \mapsto (h(y, z), y)$$

Hence all of the transition maps between such charts involve projection maps and the smooth maps involved in defining Σ as a graph. This gives Σ the structure of an abstract smooth surface.

Some of the relations given in the above theorem are easy to prove, but others come as a result of the inverse function theorem.

3.2   Inverse and implicit function theorems

Theorem (inverse function theorem). Let $U \subseteq \mathbb{R}^n$ be open, and $f : U \to \mathbb{R}^n$ be continuously differentiable. Let $p \in U$ and $f(p) = q$. Suppose $D f\big|_p$ is invertible. Then there is an open neighbourhood V of q and a differentiable map $g : V \to \mathbb{R}^n$ and $g(q) = p$ with image an open neighbourhood $U' \subseteq U$ of p such that $f \circ g = \text{id}_V$. If f is smooth, then g is also.

Remark. The chain rule then implies that $D g\big|_q = \left(D f\big|_p\right)^{-1}$. The inverse function theorem concerns functions $\mathbb{R}^n \to \mathbb{R}^n$, where $D f\big|_p$ is an isomorphism. If we have a map $\mathbb{R}^n \to \mathbb{R}^m$ for $n > m$, then we can discuss the behaviour when $D f\big|_p$ is surjective. The derivative $D f\big|_p$ is an $n \times m$ matrix, so if it has full rank, up to the permutation of coordinates we have that the last m columns are linearly independent.

Theorem (implicit function theorem). Let $p = (x_0, y_0)$ be a point in an open set $U \subset \mathbb{R}^k \times \mathbb{R}^\ell$. Let $f : U \to \mathbb{R}^\ell$ such that $p \mapsto 0$ and $\left(\dfrac{\partial f_i}{\partial y_j}\right)_{\ell \times \ell}$ is an isomorphism. Then there is an open neighbourhood V of x_0 in $\mathbb{R}^k$ and a continuously differentiable map $g : V \to \mathbb{R}^\ell$ with $x_0 \mapsto y_0$ such that if $(x, y) \in U \cap (V \times \mathbb{R}^\ell)$, then $f(x, y) = 0 \iff y = g(x)$. If f is smooth, so is g.

Proof. Let $F : U \to \mathbb{R}^k \times \mathbb{R}^\ell$ be defined by $(x, y) \mapsto (x, f(x, y))$. Then note that

$$DF = \begin{pmatrix} I & * \\ 0 & \frac{\partial f_i}{\partial y_j} \end{pmatrix}$$

hence DF is an isomorphism at (x_0, y_0). By the inverse function theorem, F is locally invertible near $F(x_0, y_0) = (x_0, f(x_0, y_0)) = (x_0, 0)$. Consider an open neighbourhood $V \times V' \subseteq \mathbb{R}^k \times \mathbb{R}^\ell$ on which this continuously differentiable inverse $G : V \times V' \to U' \subseteq U \subseteq \mathbb{R}^k \times \mathbb{R}^\ell$ exists, such that $F \circ G = \mathrm{id}_{V \times V'}$. Then,

$$G(x, y) = (\varphi(x, y), \psi(x, y)) \implies F \circ G(x, y) = (\varphi(x, y), f(\varphi(x, y), \psi(x, y))) = (x, y)$$

Hence $\varphi(x, y) = x$. We have $f(x, \psi(x, y)) = y$ when $(x, y) \in V \times V'$. This gives $f(x, y) = 0 \iff y = \psi(x, 0)$. We then define $g : V \to \mathbb{R}^\ell$ by $x \mapsto \psi(x, 0)$. $\qquad \square$

Example. Let $f : \mathbb{R}^2 \to \mathbb{R}$ be smooth and $f(x_0, y_0) = 0$, and suppose $\frac{\partial f}{\partial y} \neq 0$ at (x_0, y_0). Then there exists a smooth map $g : (x_0 - \varepsilon, x_0 + \varepsilon) \to \mathbb{R}$ with $g(x_0) = y_0$ and $f(x, y) = 0 \iff y = g(x)$ for (x, y) in some open neighbourhood of (x_0, y_0). Since $f(x, g(x)) = 0$ in this open neighbourhood, we can differentiate that expression to find

$$g'(x) = \frac{-f_x}{f_y}$$

noting that $f_y \neq 0$ in some neighbourhood near (x_0, y_0). Note that the level set $f(x, y) = 0$ is implicitly defined by g, which is a function for which we have an integral expression.

Example. Let $f : \mathbb{R}^3 \to \mathbb{R}$ be a smooth map with $f(x_0, y_0, z_0) = 0$. Consider the level set $\Sigma = f^{-1}(0)$, assuming that $Df \neq 0$ at (x_0, y_0, z_0). Permuting coordinates if necessary, we can assume $\frac{\partial f}{\partial z} \neq 0$ at this point. Then there exists an open neighbourhood V of (x_0, y_0) and a smooth function $g : V \to \mathbb{R}$ such that $(x_0, y_0) \mapsto z_0$ with the property that for an open set $(x_0, y_0, z_0) \in U$, the set $f^{-1}(0) \cap U = \Sigma \cap U$ is the graph of the function g, which is $\{(x, y, g(x, y)) : (x, y) \in V\}$.

3.3   Conditions for smoothness

We now prove the theorem stated above, relating equivalent conditions for smoothness of a surface Σ.

Proof. First, we show that (b) implies all of the other conditions. If Σ is locally a graph $\{(x, y, g(x, y))\}$, we find a chart from the coordinate plane projection π_{xy} of that graph. Since this projection is smooth and defined on an open neighbourhood of points of Σ in its domains, this shows that Σ is a smooth surface in $\mathbb{R}^3$ (a). Further, since Σ is locally the given graph, it is cut out by the function $f(x, y, z) = z - g(x, y)$ and $\frac{\partial f}{\partial z} \neq 0$ (c). Finally, the local parametrisation $\sigma(x, y) = (x, y, g(x, y))$ is allowable; g is smooth, the partial derivatives of σ are linearly independent by considering the x and y components, and σ is injective where required (d).

Now, we show (a) implies (d). This is simply part of the definition of being a smooth surface in $\mathbb{R}^3$, being locally diffeomorphic to $\mathbb{R}^2$. In particular, at $p \in \Sigma$, Σ is locally diffeomorphic to $\mathbb{R}^2$ and the inverse of such a local diffeomorphism is an allowable parametrisation.

We have already shown (c) implies (b); this was the example of the implicit function theorem provided above.

Finally, we must prove (d) implies (a) and (b), and then the result will hold. Let $p \in \Sigma$ and V be an open set in $\mathbb{R}^2$ with an allowable parametrisation to Σ such that $\sigma(0) = p$. If $\sigma = (\sigma_1(u, v), \sigma_2(u, v), \sigma_3(u, v))$,

we have

$$D\sigma = \begin{pmatrix} \frac{\partial \sigma_1}{\partial u} & \frac{\partial \sigma_1}{\partial v} \\ \frac{\partial \sigma_2}{\partial u} & \frac{\partial \sigma_2}{\partial v} \\ \frac{\partial \sigma_3}{\partial u} & \frac{\partial \sigma_3}{\partial v} \end{pmatrix}$$

This has rank 2, hence there exist two rows defining an invertible matrix. Suppose those are the first two rows, and let $\mathrm{pr} = \pi_{xy}$ be the projection map. Consider $\mathrm{pr} \circ \sigma \colon V \to \mathbb{R}^2$. This has isomorphic derivative at zero, so we can apply the inverse function theorem. Hence Σ is locally a graph over the xy coordinate plane, so (b) holds. Moreover, let $\varphi = \mathrm{pr} \circ \sigma$, and consider the open ball $B(p, \delta) \subseteq \mathbb{R}^3$ and a map such that $(x, y, z) \mapsto \varphi^{-1}(x, y)$ in this ball. Here, $\varphi \colon W \to \Sigma$ where W is an open set in $\mathrm{pr}(B(p, \delta))$. This is a locally defined map, which is smooth on an open set in $\mathbb{R}^3$, which is a smooth inverse to σ. Hence Σ is a smooth surface in $\mathbb{R}^3$, so (a) holds. $\qquad\square$

Example. The unit sphere S^2 in $\mathbb{R}^3$ is $f^{-1}(0)$ for $f(x, y, z) = x^2 + y^2 + z^2 - 1$. For any point on S^2, $Df \neq 0$, so S^2 is a smooth surface.

Example. Let $\gamma \colon [a, b] \to \mathbb{R}^3$ be a smooth map with image in the xz plane, so

$$\gamma(t) = (f(t), 0, g(t))$$

such that γ is injective, $\gamma' \neq 0$, and $f > 0$. The *surface of revolution* of γ has allowable parametrisation

$$\sigma(u, v) = (f(u) \cos v, f(u) \sin v, g(u))$$

where $(u, v) \in (a, b) \times (\theta, \theta + 2\pi)$ for a fixed θ. Note that $\sigma_u = (f_u \cos v, f_u \sin v, g_u)$ and $\sigma_v = (-f \sin v, f \cos v, 0)$, and we can check $\|\sigma_u \times \sigma_v\| = f^2((f')^2 + (g')^2)$ which is nonzero on γ, so this really is an allowable parametrisation.

Example. The orthogonal group $O(3)$ acts on S^2 by diffeomorphisms. Indeed, any $A \in O(3)$ defines a linear (hence smooth) map $\mathbb{R}^3 \to \mathbb{R}^3$ preserving S^2. Hence, the induced map on S^2 is by a homeomorphism which is smooth according to the above definition. This is analogous to the action of the Möbius group on $S^2 = \mathbb{C} \cup \{\infty\}$.

3.4 Orientability

Definition. Let V, V' be open sets in $\mathbb{R}^2$. Let $f \colon V \to V'$ be a diffeomorphism. Then at every point $x \in V$, $D f|_x \in GL(2, \mathbb{R})$; it is invertible since f is a diffeomorphism. Let $GL^+(2, \mathbb{R})$ be the subgroup of matrices with positive determinant. We say that f is *orientation-preserving* if its derivative belongs to this subgroup for all points $x \in V$.

Definition. An abstract smooth surface Σ is *orientable* if it admits an atlas $\{(U_i, \varphi_i)\}$ where the transition maps are all orientation-preserving. A choice of such an atlas is an *orientation* of Σ; Σ can be called *oriented* when such an orientation is given.

Remark. An orientable atlas belongs to a maximal compatible oriented smooth atlas.

Lemma. If Σ_1 and Σ_2 are diffeomorphic abstract smooth surfaces, then Σ_1 is orientable if and only if Σ_2 is orientable.

Proof. Let $f : \Sigma_1 \to \Sigma_2$ be a diffeomorphism. Suppose Σ_2 is orientable and equipped with an oriented smooth atlas. Consider the atlas on Σ_1 of charts of the form $(f^{-1}(U), \phi \circ f|_{f^{-1}(U)})$, where (U, ψ) is a chart at $f(p)$ in the oriented atlas for Σ_2. Then, the transition map between two such charts is exactly a transition map between charts in the Σ_2 atlas.

In other words, in the maximal smooth atlas that exists a priori for Σ_1, we will allow charts of the form $(\widetilde{U}, \widetilde{\psi})$ when for all charts (U, ψ) at $f(p)$ in the Σ_2 atlas, the map $\psi \circ f \circ (\widetilde{\psi})^{-1}$ is orientation-preserving. Informally, if the atlas on Σ_2 was maximal as an oriented atlas, we can recover the previous set of charts. $\qquad\square$

Remark. There is no sensible classification of the set of all smooth surfaces. For instance, $\mathbb{R}^2 \setminus Z$ for a closed set Z can be shown to yield uncountably many types of homeomorphisms. However, *compact* smooth surfaces may be classified by their Euler characteristic and their orientability, up to diffeomorphism. This theorem will not be proven in this course.

There is a definition of orientation-preserving *homeomorphism* that does not rely on the determinant, but that instead relies on some algebraic topology which is not covered in this course. The Möbius band is the surface

where the dashed lines represent the absence of edges. It is provable that an abstract smooth surface is orientable if and only if it contains no subsurface homeomorphic to the Möbius band. We can therefore say that a topological surface is orientable if and only if it contains no subsurface (an open set) homeomorphic to a Möbius band.

We can define other structures on an abstract smooth surface by considering smooth atlases such that if $\varphi_1 \varphi_2^{-1}$ is a transition map, then $D(\varphi_1 \varphi_2^{-1})$ at x belongs to a specific subgroup $G \leq GL(2, \mathbb{R})$. For example, defining $G = \{e\}$ leads to Euclidean surfaces. The group $GL(1, \mathbb{C})$ identified as a subgroup of $GL(2, \mathbb{R})$ yields the Riemann surfaces.

Example. For S^2 with the atlas of two stereographic projections, we can find the transition map to be
$$(u, v) \mapsto \left(\frac{u}{u^2 + v^2}, \frac{v}{u^2 + v^2} \right)$$
on $\mathbb{R}^2 \setminus \{0\}$. This has positive determinant, so S^2 is orientable.

For the torus T^2, we previously found an atlas such that the transition maps are translations of $\mathbb{R}^2$. Hence the torus is an oriented surface, and also a Euclidean surface.

3.5   Tangent planes

Recall that an *affine* subspace of a vector space is a translate of a linear subspace.

> **Definition.** Let Σ be a smooth surface in $\mathbb{R}^3$, and $p \in \Sigma$. Let $\sigma : V \to U \subseteq \Sigma$ be an allowable parametrisation of Σ near p, so V is an open subset of $\mathbb{R}^2$ and U is open in Σ, such that $\sigma(0) = p$. The *tangent plane* $T_p\Sigma$ to p at Σ is the image of $\left(D\,\sigma|_0 \right) \subseteq \mathbb{R}^3$, which is a two-dimensional vector subspace of $\mathbb{R}^3$. The *affine tangent plane* is $p + T_p\Sigma$, which is an affine subspace of $\mathbb{R}^3$.

Remark. The affine tangent plane is the 'best' linear approximation to a surface Σ at a given point.

Lemma. $T_p\Sigma$ is independent of the choice of allowable parametrisation.

Proof (i). Suppose $\sigma : V \to U$ and $\tilde{\sigma} : \widetilde{V} \to \widetilde{U}$ are allowable parametrisations with $\sigma(0) = \tilde{\sigma}(0) = p$. There exists a transition map $\sigma^{-1} \circ \tilde{\sigma}$, which is a diffeomorphism of open sets in $\mathbb{R}^2$. Therefore,

$$\tilde{\sigma} = \sigma \circ \underbrace{(\sigma^{-1} \circ \tilde{\sigma})}_{\text{diffeomorphism}}$$

Hence $D\left(\sigma^{-1} \circ \tilde{\sigma}\right)\big|_0$ is an isomorphism. Thus, the images of $D\,\tilde{\sigma}\big|_0$ and $D\,\sigma\big|_0$ agree. $\qquad\square$

Proof (ii). Let $\gamma : (-\varepsilon, \varepsilon) \to \mathbb{R}^3$ be a smooth map such that γ has image inside Σ, and $\gamma(0) = p$. We will show that $\gamma'(0) \in T_p\Sigma$. If $\sigma : V \to U$ is an allowable parametrisation with $\sigma(0) = p$ as above, and ε is sufficiently small such that $\operatorname{Im}\gamma \subseteq U$, then $\gamma(t) = \sigma(u(t), v(t))$ for some smooth functions $u, v : (-\varepsilon, \varepsilon) \to V$. Then $\gamma'(t) = \sigma_u u'(t) + \sigma_v v'(t)$ is in the image of $D\,\sigma\big|_t$. Thus, $T_p\Sigma = \operatorname{span}\{\gamma'(0) : \gamma \text{ as above}\}$. $\qquad\square$

Definition. If Σ is a smooth surface in $\mathbb{R}^3$ and $p \in \Sigma$, the *normal direction* to Σ at p is $(T_p\Sigma)^\perp$, the Euclidean orthogonal complement to the tangent plane at p.

Remark. For all $p \in \Sigma$, there exist exactly two normalised normal vectors.

Definition. A smooth surface in $\mathbb{R}^3$ is *two-sided* if it admits a continuous global choice of unit normal vector.

Lemma. A smooth surface in $\mathbb{R}^3$ is orientable (as an abstract smooth surface) if and only if it is two-sided (as a smooth surface in $\mathbb{R}^3$).

Proof. Let $\sigma : V \to U \subseteq \Sigma$ be an allowable parametrisation. Let $\sigma(0) = p$. We will define the positive unit normal with respect to σ at p to be the normal vector $n_\sigma(p)$ with the property that $\{\sigma_u, \sigma_v, n_\sigma(p)\}$ and $\{e_1, e_2, e_3\}$ are related by a positive determinant change of basis matrix, where $\{e_1, e_2, e_3\}$ are the standard basis vectors. In other words,

$$n_\sigma(p) = \frac{\sigma_u \times \sigma_v}{\|\sigma_u \times \sigma_v\|}$$

Consider an alternative parametrisation $\tilde{\sigma} : \widetilde{V} \to \widetilde{U}$, where $\tilde{\sigma}(0) = p$, such that $\tilde{\sigma}$ belongs to the same oriented and smooth atlas as σ. Hence, $\sigma = \tilde{\sigma} \circ \varphi$ for some transition map φ. Let

$$D\,\varphi\big|_0 = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}$$

Hence,

$$\sigma_u = \alpha\tilde{\sigma}_u + \gamma\tilde{\sigma}_v; \quad \sigma_v = \beta\tilde{\sigma}_u + \delta\tilde{\sigma}_v$$

This gives

$$\sigma_u \times \sigma_v = \det\left(D\,\varphi\big|_0\right)\tilde{\sigma}_u \times \tilde{\sigma}_v \tag{$\dagger$}$$

The determinant here is positive since the charts in question belong to an oriented atlas. Thus, the positive normal depends on the orientation of Σ, but does not depend on the choice of parametrisation. The expression for $n_\sigma(p)$ is continuous since the cross product is continuous, hence Σ is two-sided.

Conversely, if Σ is two-sided and there exists a global continuous choice of normal vector, we can consider the subatlas of the natural smooth atlas with the property that we allow (U, φ) if the associated parametrisation $\sigma = \varphi^{-1}$ satisfies $\{\sigma_u, \sigma_v, n\}$ is a positive basis for $\mathbb{R}^3$, where n is the given choice of normal. By (†), the transition maps between such charts are orientation-preserving. Hence Σ is orientable. $\qquad\square$

> **Lemma.** If Σ is a smooth surface in $\mathbb{R}^3$ and $A : \mathbb{R}^3 \to \mathbb{R}^3$ is a smooth map which preserves Σ setwise, then $\left. D\, A \right|_p \in L(\mathbb{R}^3, \mathbb{R}^3)$ maps $T_p\Sigma$ to $T_{A(p)}\Sigma$ for $p \in \Sigma$.

Proof. Let $\gamma : (-\varepsilon, \varepsilon) \to \mathbb{R}^3$ be a smooth map such that its image lies on Σ, and $\gamma(0) = p$. Recall that $T_p\Sigma$ is spanned by $\gamma'(0)$ for such curves γ. Now, consider $A \circ \gamma : (-\varepsilon, \varepsilon) \to \mathbb{R}^3$, which also has image Σ, and

$$\left. D\, A \right|_{\gamma(0)} \circ \left. D\, \gamma \right|_0 = \left. D\, A \right|_p (\gamma'(0)) = \left. D\, (A \circ \gamma) \right|_0 \in T_{A(p)}\Sigma$$

$\qquad\square$

Example. Let S^2 be the unit sphere. The normal vector at p is the line through the origin and p; indeed, since SO_3 acts transitively on S^2, it suffices to check at one point, such as the north pole. We can choose the outward-facing normal vector to be the positive normal, denoted $n(p)$. S^2 is two-sided by the construction of this normal vector, hence S^2 is orientable.

Example. One embedding of the Möbius band in $\mathbb{R}^3$ is

$$\sigma(t, \theta) = \left(\left(1 - t \sin\frac{\theta}{2}\right) \cos\theta, \left(1 - t \sin\frac{\theta}{2}\right) \sin\theta, t \cos\frac{\theta}{2} \right)$$

where $(t, \theta) \in V_1 = \left\{ t \in \left(-\frac{1}{2}, \frac{1}{2}\right), \theta \in (0, 2\pi) \right\}$ or $(t, \theta) \in V_2 = \left\{ t \in \left(-\frac{1}{2}, \frac{1}{2}\right), \theta \in (-\pi, \pi) \right\}$. We begin with the unit circle $x^2 + y^2 = 1$, for $t = 0$. Then, at each point on the circle, we consider an open interval of unit length, which will rotate as we move around the circle, such that at the point θ on the circle it has rotated by $\frac{\theta}{2}$. We can check that if σ_i is σ on V_i, then σ_i is allowable. Further,

$$\sigma_t \times \sigma_\theta = \left(-\cos\theta \cos\frac{\theta}{2}, -\sin\theta \cos\frac{\theta}{2}, -\sin\frac{\theta}{2} \right) \equiv n_\theta$$

which is already normalised. As $\theta \to 0$ from above, $n_\theta \to (-1, 0, 0)$. As $\theta \to 2\pi$ from below, $n_\theta \to (1, 0, 0)$. Hence, the surface is not two-sided.

4   Geometry of surfaces in $\mathbb{R}^3$

4.1   First fundamental form

Let $\gamma : (a, b) \to \mathbb{R}^3$ be smooth. The *length* of γ is

$$L(\gamma) = \int_a^b \|\gamma'(t)\|\, \mathrm{d}t$$

This result is independent of the choice of parametrisation. Let $s \colon (A, B) \to (a, b)$ be a monotonically increasing function, and let $\tau(t) = \gamma(s(t))$. Then

$$L(\tau) = \int_A^B \|\tau'(t)\| \, \mathrm{d}t = \int_A^B \|\gamma'(s(t))\| |s'(t)| \, \mathrm{d}t = \int_a^b \|\gamma'(t')\| \, \mathrm{d}t' = L(\gamma)$$

Lemma. If $\gamma \colon (a, b) \to \mathbb{R}^3$ is continuously differentiable and $\gamma'(t) \neq 0$, then γ can be parametrised by arc length.

The proof is left as an exercise. Let Σ be a smooth surface in $\mathbb{R}^3$, and let $\sigma \colon V \to U \subseteq \Sigma$ be an allowable parametrisation. If $\gamma \colon (a, b) \to \mathbb{R}^3$ is smooth and its image is contained within U, then there exist functions $(u(t), v(t)) \colon (a, b) \to V$ such that $\gamma(t) = \sigma(u(t), v(t))$. Hence $\gamma'(t) = \sigma_u u'(t) + \sigma_v v'(t)$, giving

$$\|\gamma'(t)\|^2 = E u'(t)^2 + 2F u'(t) v'(t) + G v'(t)^2$$

for functions

$$E = \langle \sigma_u, \sigma_u \rangle ; \quad F = \langle \sigma_u, \sigma_v \rangle = \langle \sigma_v, \sigma_u \rangle ; \quad G = \langle \sigma_v, \sigma_v \rangle$$

where $\langle \, \cdot \, , \, \cdot \, \rangle$ represents the usual Euclidean inner product. Note that E, F, G depend only on σ and not on γ.

Definition. The *first fundamental form* of Σ in the parametrisation σ is the expression

$$E \, \mathrm{d}u^2 + 2F \, \mathrm{d}u \, \mathrm{d}v + G \, \mathrm{d}v^2$$

This notation is illustrative of the fact that if γ has image in the image of $\sigma(v)$, we find

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

where $\gamma(t) = \sigma(u(t), v(t))$.

Remark. The Euclidean inner product on $\mathbb{R}^3$ provides an inner product on the subspace $T_p\Sigma$. Choosing a parametrisation σ, we can say $T_p\Sigma = \operatorname{Im} D\sigma|_0 = \operatorname{span}\{\sigma_u, \sigma_v\}$ where $\sigma(0) = p$. The first fundamental form is a symmetric bilinear form on the tangent spaces $T_p\Sigma$, varying smoothly in p. However, we choose to express this in a basis coming from the parametrisation σ. In particular, we can think about the matrix expression

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

Example. The plane $\mathbb{R}^2_{xy} \subset \mathbb{R}^3$ has the parametrisation $(u, v) \mapsto (u, v, 0)$. Hence, $\sigma_u = e_1$ and $\sigma_v = e_2$, hence the first fundamental form is $\mathrm{d}u^2 + \mathrm{d}v^2$. We could also use polar coordinates, using $\sigma(r, \theta) = (r \cos \theta, r \sin \theta, 0)$. This parametrises the plane without the origin. This gives $\sigma_r = (\cos \theta, \sin \theta, 0)$ and $\sigma_\theta = (-r \sin \theta, r \cos \theta, 0)$. The first fundamental form is $\mathrm{d}r^2 + r^2 \, \mathrm{d}\theta^2$.

Definition. Let Σ, Σ' be smooth surfaces in $\mathbb{R}^3$. We say that they are *isometric* if there exists a diffeomorphism $f \colon \Sigma \to \Sigma'$ that preserves the lengths of all curves. More formally, for every smooth curve $\gamma \colon (a, b) \to \Sigma$, the length of γ is the same as the length of $f \circ \gamma$.

Example. Let $\Sigma' = f(\Sigma)$ where $f : \mathbb{R}^3 \to \mathbb{R}^3$ is a global isometry, or rigid motion, of $\mathbb{R}^3$; that is, $v \mapsto Av + b$ for an orthogonal matrix A. These isometries preserve the Euclidean inner product on $\mathbb{R}^3$, hence f preserves length. However, in the definition, we need not map all of $\mathbb{R}^3$ to itself, just $\Sigma \to \Sigma'$.

> **Definition.** We say that Σ and Σ' are *locally isometric* near points $p \in \Sigma$ and $q \in \Sigma'$ if there exist open neighbourhoods U of p and V of q such that U and V are isometric. We can also say that Σ and Σ' are locally isometric if they are locally isometric at all points; that is, each point of Σ is locally isometric to some point on Σ'.

> **Lemma.** Smooth surfaces Σ, Σ' in $\mathbb{R}^3$ are locally isometric near $p \in \Sigma$ and $q \in \Sigma'$ if and only if there exist allowable parametrisations $\sigma : V \to U \subseteq \Sigma$ and $\sigma' : V \to U' \subseteq \Sigma'$ such that the first fundamental forms are equivalent.

Proof. By definition, the first fundamental form of Σ determines the lengths of all curves on Σ that lie in U. We will now show that lengths of curves determine the first fundamental form of a parametrisation. Given $\sigma : V \to U$, without loss of generality let $V = B(0, \delta)$ for some $\delta > 0$, where $\sigma(0) = p$. Consider, for all $\varepsilon < \delta$, the curve

$$\gamma_\varepsilon : [0, \varepsilon] \to U; \quad t \mapsto \sigma(t, 0)$$

Then,

$$\frac{\mathrm{d}}{\mathrm{d}\varepsilon} L(\gamma_\varepsilon) = \frac{\mathrm{d}}{\mathrm{d}\varepsilon} \int_0^\varepsilon \sqrt{E(t, 0)} \, \mathrm{d}t = \sqrt{E(\varepsilon, 0)}$$

Hence,

$$\left. \frac{\mathrm{d}}{\mathrm{d}\varepsilon} \right|_{\varepsilon=0} L(\gamma_\varepsilon) = \sqrt{E(0, 0)}$$

So we can determine E at p by looking at lengths of curves. We can similarly consider

$$\chi_\varepsilon : [0, \varepsilon] \to U; \quad t \mapsto \sigma(0, t)$$

which determines G. Finally, consider

$$\lambda_\varepsilon : [0, \varepsilon] \to U; \quad t \mapsto \sigma(t, t)$$

which determines $\sqrt{(E + 2F + G)(0, 0)}$ which gives F implicitly. $\qquad \square$

Example. The sphere of radius a, given by $\{x^2 + y^2 + z^2 = a^2\}$, has an open set with allowable parametrisation

$$\sigma(u, v) = (a \cos u \cos v, a \cos u \sin v, a \sin u)$$

where $u \in \left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$ and $v \in (0, 2\pi)$. This parametrises the complement of a half great circle. Here,

$$\sigma_u = (-a \sin u \cos v, -a \sin u \sin v, a \cos u); \quad \sigma_v = (-a \cos u \sin v, a \cos u \cos v, 0)$$

Hence,

$$E = a^2; \quad F = 0; \quad G = a^2 \cos^2 u$$

which gives the first fundamental form as

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

Example. Consider the surface of revolution given by a curve

$$\eta(t) = (f(t), 0, g(t))$$

rotated about the z axis. The resulting surface has parametrisation

$$\sigma(u, v) = (f(u) \cos v, f(u) \sin v, g(u))$$

Hence,

$$\sigma_u = (f_u \cos v, f_u \sin v, g_u); \quad \sigma_v = (-f \sin v, f \cos v, 0)$$

which gives

$$(f_u^2 + g_u^2) \, du^2 + f^2 \, dv^2$$

Example. Consider the cone with angle $\arctan a$ to the vertical. For $u > 0$ and $v \in (0, 2\pi)$, we define

$$\sigma(u, v) = (au \cos v, au \sin v, u)$$

The first fundamental form is

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

Consider cutting the cone along the line $v = 0$ and flattening it into a plane sector. The circumference of the sector is $2\pi a$ and the radius is $\sqrt{1 + a^2}$, hence the angle traced out by the sector is $\theta_0 = \frac{2\pi a}{\sqrt{1+a^2}}$. We can parametrise this subset of the plane by

$$\sigma(r, \theta) = \left(\sqrt{1 + a^2} r \cos\left(\frac{a\theta}{\sqrt{1 + a^2}} \right), \sqrt{1 + a^2} r \sin\left(\frac{a\theta}{\sqrt{1 + a^2}} \right), 0 \right)$$

for $r > 0$ and $\theta \in (0, 2\pi)$. We can then check that the first fundamental form here is

$$(1 + a^2) \, dr^2 + r^2 a^2 \, d\theta^2$$

which matches the first fundamental form for the cone itself. Hence the cone and the plane are locally isometric. However, the cone and plane are not globally isometric, since the two topological spaces are not homeomorphic, so no diffeomorphism that preserves lengths can be constructed.

Lemma. Let Σ be a smooth surface in $\mathbb{R}^3$, and let $p \in \Sigma$. Suppose we have two allowable parametrisations $\sigma \colon V \to U$ and $\sigma' \colon V' \to U$ for the same open neighbourhood of p. The two parametrisations differ by a transition map $F = \sigma'^{-1} \circ \sigma$ which is a diffeomorphism of open subsets of $\mathbb{R}^2$. There exist first fundamental forms for both parametrisations. Then,

$$\begin{pmatrix} E & F \\ F & G \end{pmatrix} = (DF)^{\intercal} \begin{pmatrix} E' & F' \\ F' & G' \end{pmatrix} (DF)$$

Proof. By definition,

$$\begin{pmatrix} E & F \\ F & G \end{pmatrix} = \begin{pmatrix} \sigma_u \cdot \sigma_u & \sigma_u \cdot \sigma_v \\ \sigma_v \cdot \sigma_u & \sigma_v \cdot \sigma_v \end{pmatrix} = (D\sigma)^{\intercal}(D\sigma)$$

Now, $\sigma = \sigma' \circ F$ hence the result follows. $\square$

4.2 Conformality

If $v, w \in \mathbb{R}^3$, we have $v \cdot w = |v| \cdot |w| \cdot \cos\theta$. This allows us to deduce the angle θ between two vectors given their dot product and lengths. This can also be done when v, w are in the tangent plane $T_p\Sigma$, and then we can express the angle in terms of the first fundamental form. Let σ be an allowable parametrisation for Σ near p, such that $D\,\sigma|_0$ evaluates to v at v_0 and w at w_0.

$$\cos\theta = \frac{v \cdot w}{|v| \cdot |w|} = \frac{I(v_0, w_0)}{\sqrt{I(v_0, v_0)}\sqrt{I(w_0, w_0)}}$$

where I denotes the first fundamental form of σ at zero.

> **Lemma.** Let Σ be a smooth surface in $\mathbb{R}^3$, and let $\sigma : V \to U$ be an allowable parametrisation of Σ near p. Then σ is *conformal* if $E = G$ and $F = 0$ in the first fundamental form.

Proof. Consider curves $\gamma : t \mapsto (u(t), v(t))$ and $\tilde\gamma : t \mapsto (\tilde u(t), \tilde v(t))$ in V, where $\gamma(0) = \tilde\gamma(0) = 0 \in V$. Let σ be a parametrisation $V \to U \subseteq \Sigma$ such that $\sigma(0) = p \in \Sigma$. Then the curves $\sigma \circ \gamma$ and $\sigma \circ \tilde\gamma$ meet at angle θ on Σ, where

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

In particular, if σ is conformal, suppose $\gamma(t) = (t, 0)$ and $\tilde\gamma(t) = (0, t)$. Then, we have that the curves meet at $\frac{\pi}{2}$ in V, so they meet at $\frac{\pi}{2}$ in Σ, so we find that $\cos\theta = 0 \implies F = 0$. Similarly, if $\gamma(t) = (t, t)$ and $\tilde\gamma(t) = (t, -t)$, we find $\cos\theta = 0 \implies E = G$.

Conversely, suppose there exists a parametrisation σ such that $E = G$ and $F = 0$. Then, in this parametrisation, the first fundamental form is of the form $\rho(\mathrm{d}u^2 + \mathrm{d}v^2)$ for $\rho = E : V \to \mathbb{R}$. Hence, the first fundamental form is a pointwise rescaling of the Euclidean fundamental form $\mathrm{d}u^2 + \mathrm{d}v^2$. Rescaling the plane does not change angles, so σ is conformal as required. $\qquad\square$

Remark. Conformality in charts is historically important for cartography. The existence of conformal charts is closely connected to Riemann surfaces, which are topological surfaces locally modelled on $\mathbb{C}$ instead of $\mathbb{R}^2$.

4.3 Area

Recall that a parallelogram spanned by vectors v, w has area $|v \times w| = \langle v, v\rangle\langle w, w\rangle - \langle v, w\rangle^2$, where $\times$ denotes the cross product. Let $\sigma : V \to U \subseteq \Sigma$ be an allowable parametrisation with $\sigma(0) = p$, and consider $\sigma_u, \sigma_v \in T_p\Sigma$. The square of the area of the infinitesimal parallelogram spanned by σ_u, σ_v is given by

$$\langle\sigma_u, \sigma_u\rangle\langle\sigma_v, \sigma_v\rangle - \langle\sigma_u, \sigma_v\rangle^2 = EG - F^2$$

> **Definition.** Let Σ be a smooth surface in $\mathbb{R}^3$, and $\sigma : V \to U \subseteq \Sigma$ an allowable parametrisation. Then,
> $$\mathrm{area}(U) = \int_V \sqrt{EG - F^2}\,\mathrm{d}u\,\mathrm{d}v$$

Remark. This is independent of parametrisation. Indeed, suppose $\sigma \colon V \to U$ and $\tilde{\sigma} \colon \widetilde{V} \to U$ are allowable. Then $\tilde{\sigma} = \sigma \circ \varphi$ for some transition map $\varphi \colon \widetilde{V} \to V$. We know then that

$$\begin{pmatrix} \widetilde{E} & \widetilde{F} \\ \widetilde{F} & \widetilde{G} \end{pmatrix} = (D\tilde{\sigma})^{\intercal}(D\tilde{\sigma}) = (D\varphi)^{\intercal} \begin{pmatrix} E & F \\ F & G \end{pmatrix} (D\varphi)$$

Hence,

$$\sqrt{\widetilde{E}\widetilde{G} - \widetilde{F}^2} = |\det(D\varphi)|\sqrt{EG - F^2}$$

The usual change of variables formula for integration, combined with the fact that φ is a diffeomorphism, gives

$$\int_V \sqrt{EG - F^2}\, du\, dv = \int_{\widetilde{V}} \sqrt{\widetilde{E}\widetilde{G} - \widetilde{F}^2}\, du\, dv$$

Note, we can compute the area of an open set $U \subseteq \Sigma$, not necessarily lying in a single parametrisation, by covering the set by a finite amount of open subsets which lie in single charts. For instance, if Σ is compact, we can compute the area of Σ itself.

Example. Consider the graph $\Sigma = \{(u, v, f(u,v)) \colon (u,v) \in \mathbb{R}^2\}$, where $f \colon \mathbb{R}^2 \to \mathbb{R}$ is a smooth function. This has a global parametrisation $\sigma(u,v) = (u, v, f(u,v))$. Here, $\sigma_u = (1, 0, f_u)$ and $\sigma_v = (0, 1, f_v)$, hence

$$\sqrt{EG - F^2} = \sqrt{1 + f_u^2 + f_v^2}$$

Let $U_R \subseteq \Sigma$ be the part of the graph lying inside the disc $B(0, R) \subseteq \mathbb{R}^2$. Then

$$\text{area}(U_R) = \int_{B(0,R)} \sqrt{1 + f_u^2 + f_v^2}\, du\, dv \geq \pi R^2$$

with equality exactly when $f_u = f_v = 0$, so f is constant and U_R is contained inside a plane perpendicular to the z axis. Hence, the projection from Σ to $\mathbb{R}_{xy}^2$ is not area-preserving, unless Σ is a plane perpendicular to the z axis.

Example. Consider the sphere enclosed exactly by a cylinder. The cylindrically radial projection from the sphere to the cylinder is area-preserving. This is explored further in the example sheets.

4.4 Second fundamental form

Let $\sigma \colon V \to U \subseteq \Sigma$ be allowable. By using Taylor's theorem, we can write

$$\begin{aligned}
\sigma(u + h, v + \ell) &= \sigma(u, v) \\
&+ h\sigma_u(u, v) + \ell\sigma_v(u, v) \\
&+ \frac{1}{2}\big(h^2\sigma_{uu}(u, v) + 2h\ell\sigma_{uv}(u, v) + \ell^2\sigma_{vv}(u, v)\big) \\
&+ O(h^3, \ell^3)
\end{aligned}$$

where h, ℓ are small, and $(u + h, v + \ell) \in V$. Recall that if $p = \sigma(u, v)$, we have $T_p\Sigma = \langle\{\sigma_u, \sigma_v\}\rangle$. Hence, the orthogonal distance from $\sigma(u + h, v + \ell)$ to the affine tangent plane $T_p\Sigma + p$ is given by projection to the normal direction.

$$\langle n, \sigma(u + h, v + \ell) - \sigma(u, v)\rangle = \frac{1}{2}\big(\langle n, \sigma_{uu}\rangle h^2 + 2\langle n, \sigma_{uv}\rangle h\ell + \langle n, \sigma_{vv}\rangle \ell^2\big) + O(h^3, \ell^3)$$

Definition. The *second fundamental form* of Σ in the allowable parametrisation σ is the quadratic form

$$L\,du^2 + 2M\,du\,dv + N\,dv^2$$

where

$$L = \langle n, \sigma_{uu}\rangle; \quad M = \langle n, \sigma_{uv}\rangle; \quad N = \langle n, \sigma_{vv}\rangle$$

and

$$n = \frac{\sigma_u \times \sigma_v}{\|\sigma_u \times \sigma_v\|}$$

We can write this as the matrix

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

which defined a quadratic form on $T_p\Sigma$ which varies smoothly in p.

Lemma. Let V be connected and $\sigma : V \to U \subseteq \Sigma$ be an allowable parametrisation such that the second fundamental form vanishes identically with respect to σ. Then U lies in an affine plane.

Remark. The first fundamental form is a non-degenerate symmetric bilinear form on $T_p\Sigma$, whereas the second fundamental form may be degenerate.

Proof. By definition,

$$\langle n, \sigma_u\rangle = 0 = \langle n, \sigma_v\rangle$$

Hence, by differentiating, we find

$$0 = \langle n_u, \sigma_u\rangle + \langle n, \sigma_{uu}\rangle = \langle n_v, \sigma_v\rangle + \langle n, \sigma_{vv}\rangle = \langle n_v, \sigma_u\rangle + \langle n, \sigma_{uv}\rangle$$

Some of these terms appear in the definition of the second fundamental form:

$$L = \langle n, \sigma_{uu}\rangle = -\langle n_u, \sigma_u\rangle$$
$$M = \langle n, \sigma_{uv}\rangle = -\langle n_v, \sigma_u\rangle = -\langle n_u, \sigma_v\rangle$$
$$N = \langle n, \sigma_{vv}\rangle = -\langle n_v, \sigma_v\rangle$$

If the second fundamental form vanishes, then n_u is orthogonal to σ_u, σ_v, and n itself. Since σ_u, σ_v, n form a basis for $\mathbb{R}^3$, we have $n_u = 0$. Similarly, $n_v = 0$, hence n is constant by the mean value theorem. $\qquad\square$

Remark. The first fundamental form in parametrisation σ can be written $(D\sigma)^\intercal(D\sigma)$. We can similarly write the second fundamental form as

$$-(Dn)^\intercal(D\sigma) = \begin{pmatrix} L & M \\ M & N \end{pmatrix} = -\begin{pmatrix} n_u \cdot \sigma_u & n_u \cdot \sigma_v \\ n_v \cdot \sigma_u & n_v \cdot \sigma_v \end{pmatrix}$$

Hence, if $\sigma : V \to \Sigma$ and $\tilde{\sigma} : \widetilde{V} \to \Sigma$ are allowable parametrisations for an open set $U \subseteq \Sigma$ with transition map $\varphi : \widetilde{V} \to V$ given by $\varphi = \sigma^{-1} \circ \tilde{\sigma}$, we can use the above expression to find

$$\begin{pmatrix} \check{L} & \widetilde{M} \\ \widetilde{M} & \widetilde{N} \end{pmatrix} = \pm(D\varphi)^\intercal \begin{pmatrix} L & M \\ M & N \end{pmatrix} (D\varphi)$$

The change in sign depends on whether the transition map preserves or reverses orientation. If the normal vectors agree, there is no negative sign.

$$n_{\sigma\circ\varphi}\Big|_{(\tilde{u},\tilde{v})} = \pm\, n_{\sigma}\Big|_{\varphi(\tilde{u},\tilde{v})}$$

for $(\tilde{u},\tilde{v}) \in \widetilde{V}$. In particular, if $\det(D\varphi) < 0$, we arrive at a negative sign. If we assume that $V, \widetilde{V}$ are connected, the determinant $\det(D\varphi)$ does not change sign.

Example. Consider the cylinder with allowable parametrisation

$$\sigma(u,v) = (a\cos u, a\sin u, v)$$

where $u \in (0, 2\pi), v \in \mathbb{R}$. Note that $\sigma_{uv} = \sigma_{vv} = 0$, hence $M = N = 0$. We can show that the second fundamental form is given by

$$\begin{pmatrix} -a & 0 \\ 0 & 0 \end{pmatrix}; \quad -a\,du^2$$

4.5   Gauss maps

Definition. Let Σ be a smooth oriented surface in $\mathbb{R}^3$. The *Gauss map* $n : \Sigma \to \mathbb{S}^2$ is the map $p \mapsto n(p)$, where the normal vector is normalised and hence lies in the unit sphere.

Lemma. The Gauss map is smooth.

Proof. Since smoothness is a local property, it suffices to check the smoothness of the map on an arbitrary parametrised part of Σ. Let $\sigma : V \to U \subseteq \Sigma$ be allowable and compatible with a chosen orientation. Then

$$n(p) = \frac{\sigma_u \times \sigma_v}{\|\sigma_u \times \sigma_v\|}$$

Since σ is allowable, the denominator is non-vanishing. Hence, $n(p)$ is smooth as required. $\square$

Remark. If $\Sigma = F^{-1}(0)$ for some function $F : \mathbb{R}^3 \to \mathbb{R}$ with nonzero derivative DF at all points $x \in \Sigma$ (which was required for Σ to be a smooth surface in $\mathbb{R}^3$), then we can explicitly calculate the Gauss map to be

$$n(p) = \frac{\nabla F}{\|\nabla F\|}$$

Note that,

$$T_p\Sigma = T_{n(p)}S^2 = (n(p))^{\perp}$$

since the two planes are orthogonal to the same vector. More concretely, if $v \in T_p\Sigma$ is $\gamma'(0)$ where $\gamma : (-\varepsilon, \varepsilon) \to \Sigma, \gamma(0) = p$ for a smooth curve γ, we can apply the Gauss map to γ and find

$$n \circ \gamma : (-\varepsilon, \varepsilon) \to S^2; \quad (n \circ \gamma)(0) = n(p)$$

Then, by the chain rule,

$$D\,n\Big|_p (v) = (n \circ \gamma)'(0) \in T_{n(p)}S^2 = T_p\Sigma$$

Thus, the derivative of the Gauss map is $D\,n|_p : T_p\Sigma \to T_p\Sigma$. This can be viewed as an endomorphism of a fixed (with respect to parametrisation choice) two-dimensional subspace of $\mathbb{R}^3$.

To summarise, let Σ be an oriented smooth surface in $\mathbb{R}^3$. Then,

(i) The first fundamental form is a symmetric bilinear form $\langle\,\cdot\,,\,\cdot\,\rangle = \mathrm{I}_p : T_p\Sigma \times T_p\Sigma \to \mathbb{R}$, which is the restriction of the Euclidean inner product to this space $T_p\Sigma$. We can write $\mathrm{I}_p(v,w)$, where $v, w \in T_p\Sigma$.

(ii) The second fundamental form is also a symmetric bilinear form $\mathrm{I\!I}_p : T_p\Sigma \times T_p\Sigma \to \mathbb{R}$, given by

$$\mathrm{I\!I}_p(v,w) = \mathrm{I}_p\left(-D\,n\Big|_p (v), w\right)$$

where n is the Gauss map.

If we choose an allowable parametrisation (which for the second fundamental form must be correctly oriented) $\sigma : V \to U \subseteq \Sigma$ near $p \in \Sigma$, and if

$$D\,\sigma\Big|_0 (\hat{v}) = v; \quad D\,\sigma\Big|_0 (\hat{w}) = w; \quad \sigma(0) = p$$

Then,

$$\mathrm{I}_p(v,w) = \hat{v}^\intercal \begin{pmatrix} E & F \\ F & G \end{pmatrix} \hat{w}; \quad \mathrm{I\!I}_p(v,w) = \hat{v}^\intercal \begin{pmatrix} L & M \\ M & N \end{pmatrix} \hat{w}$$

where E, F, G, L, M, N depend on the choice of σ. Note that the functions I_p and $\mathrm{I\!I}_p$ are independent of σ.

Lemma. The derivative of the Gauss map is self-adjoint. More precisely, viewing the map $D\,n|_p : T_p\Sigma \to T_p\Sigma$ as an endomorphism over the inner product space with the first fundamental form, this linear map satisfies

$$\mathrm{I}_p\left(D\,n\Big|_p (v), w\right) = \mathrm{I}_p\left(v, D\,n\Big|_p (w)\right)$$

for all $v, w \in T_p\Sigma$.

Proof. From expressions for local parametrisations, we can show that I_p and $\mathrm{I\!I}_p$ are symmetric. Hence,

$$\mathrm{I}_p(D\,n\Big|_p (v), w) = -\mathrm{I\!I}_p(v,w) = -\mathrm{I\!I}_p(w,v) = \mathrm{I}_p(D\,n\Big|_p (w), v) = \mathrm{I}_p(v, D\,n\Big|_p (w))$$

$\square$

Remark. The *fundamental theorem of surfaces in $\mathbb{R}^3$* states that a smooth oriented connected surface in $\mathbb{R}^3$ is determined completely, up to rigid motion, by the two fundamental forms.

4.6 Gauss curvature

Definition. Let Σ be a smooth surface in $\mathbb{R}^3$. The *Gauss curvature* $\kappa : \Sigma \to \mathbb{R}$ of Σ is the function defined by

$$\kappa(p) = \det\left(D\, n\Big|_p \right)$$

Remark. This is always well-defined, even if Σ is not oriented. This is because Σ is always locally orientable, and the two normals differ by sign. In two dimensions, $\det(-A) = \det(A)$, so the determinant is invariant.

We can compute κ directly. Let Σ be a smooth surface in $\mathbb{R}^3$, and σ an allowable parametrisation for an open neighbourhood of a point p. Recall that

$$\mathrm{I}_p : T_p\Sigma \to T_p\Sigma; \quad (v,w) \mapsto \langle v,w\rangle; \quad \mathrm{II}_p : T_p\Sigma \to T_p\Sigma; \quad (v,w) \mapsto \mathrm{I}_p(-\,Dn\Big|_p(v), w)$$

and $Dn\big|_p : T_p\Sigma \to T_p\Sigma$. The choice of parametrisation σ for an open neighbourhood U of p provides a preferred basis $\{\sigma_u, \sigma_v\}$ for $T_p\Sigma$. We can therefore write the fundamental forms as matrices with respect to this basis. Let $A = \mathrm{I}_p, B = \mathrm{II}_p, \mathbb{S} = Dn\big|_p$ in this basis. In matrix form, we can write $\mathrm{II}_p = \mathrm{I}_p(-\,Dn\big|_p(v), w)$ as

$$B = -\mathbb{S}^\mathsf{T}A \implies \kappa(p) = \det(\mathbb{S}) = \det(-A^{-1}B) = \frac{LN - M^2}{EG - F^2}$$

If $\sigma, \tilde{\sigma}$ are allowable and $\varphi = \sigma^{-1} \circ \sigma$ is a transition map, then

$$\widetilde{A} = (D\varphi)^\mathsf{T}A(D\varphi); \quad \widetilde{B} = \pm(D\varphi)^\mathsf{T}B(D\varphi)$$

Since the sign vanishes under taking determinants, κ is intrinsic and does not depend on the choice of parametrisation.

Example. For a cylinder $\{x^2 + y^2 = 1\}$ the Gauss map $n : \Sigma \to S^2$ has image which lies in the equator. Its derivative $Dn\big|_p : T_p\Sigma \to T_p\Sigma$ has one-dimensional image, since any $\gamma : (-\varepsilon, \varepsilon) \to \Sigma$ has $n \circ \gamma \subseteq S^1$. Hence its Gauss curvature is zero.

Definition. A smooth surface in $\mathbb{R}^3$ with vanishing Gauss curvature everywhere is *flat*.

Remark. If $\sigma : V \to U$ is allowable, and n_σ is defined to be $n \circ \sigma : V \to S^2$, then

$$Dn_\sigma\Big|_0 : \sigma_u \mapsto (n_\sigma)_u; \quad \sigma_v \mapsto (n_\sigma)_v$$

In particular, $\kappa(p) = \kappa(\sigma(0))$ vanishes if and only if $(n_\sigma)_u \times (n_\sigma)_v = 0$. Usually, we will write n to denote n_σ. In this case, the condition for flatness is that $n_u \times n_v = 0$.

Example. If Σ is the graph of a smooth function f, then on the example sheets we show that

$$\kappa = \frac{f_{uu}f_{vv} - f_{uv}^2}{(1 + f_u^2 + f_v^2)^2}$$

Hence, the curvature depends on the derivative and the Hessian of f. For instance, let $f(u,v) = \sqrt{r^2 - u^2 - v^2}$. Here, the graph is a piece of a sphere of radius r. We can find

$$f_{uu}\Big|_0 = f_{vv}\Big|_0 = \frac{-1}{r}; \quad f_{uv}\Big|_0 = 0 \implies \kappa(0,0,r) = \frac{1}{r^2}$$

Since $O(3)$ acts transitively on S^2, and the fundamental forms are preserved by such global isometries, $\kappa = \frac{1}{r^2}$ everywhere on the sphere of radius r.

Example. Let Σ be the smooth surface given by $\{z = x^2 + y^2\}$. We claim that, for the inward facing choice of orientation, the image of the Gauss map is the open northern hemisphere. Note that Σ is invariant under rotations about the z axis. Also, we can show that if R is a rotation, $n \circ R = R \circ n$. Therefore, it suffices to consider an arbitrary point with $y = 0$.

Here, $\Sigma = F^{-1}(0)$ for the function $F(x,y,z) = z - x^2 - y^2$, which has nonvanishing derivative at the points $p \in \Sigma$. Hence, at $p = (x, 0, x^2)$, we have

$$n(p) = \frac{\nabla F}{\|\nabla F\|} = \frac{(-2x, 0, 1)}{\sqrt{1 + 4x^2}}$$

We can check explicitly that this map has image which an arc lying in the open northern hemisphere.

4.7 Elliptic, hyperbolic, and parabolic points

Definition. Let Σ be a smooth surface in $\mathbb{R}^3$ and $p \in \Sigma$. We say that p is
  (i) *elliptic* if $\kappa(p) > 0$;
  (ii) *hyperbolic* if $\kappa(p) < 0$;
  (iii) *parabolic* if $\kappa(p) = 0$.

Lemma. In a sufficiently small neighbourhood of an elliptic point p, Σ lies entirely on one side of $p + T_p\Sigma$. If p is hyperbolic, Σ lies on both sides of $p + T_p\Sigma$.

Proof. Let σ be a local parametrisation near p. Here,

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

The denominator is always positive, since it is the determinant of a positive definite symmetric bilinear form I_p. Hence, the sign of κ depends on the sign of $LN - M^2$. If $w = h\sigma_u + \ell\sigma_v \in T_p\Sigma$, then $\frac{1}{2}\mathrm{II}_p(w,w)$ measures the signed distance from $\sigma(h,l)$ to $p + T_p\Sigma$. If p is elliptic, then II_p has eigenvalues of the same sign, so it is either positive or negative definite at p. Since II_p varies smoothly in p, it remains positive or negative definite in a small neighbourhood of p. Hence, in such a neighbourhood, the signed distance has the same sign as required. Conversely, if p is hyperbolic, $\mathrm{II}_p(w,w)$ takes both signs in a neighbourhood of p. $\square$

Remark. We cannot conclude anything about parabolic points *a priori*. For instance, the cylinder is flat (all points are parabolic), and the surface lies on one side of the tangent plane at every point. Consider also the *monkey saddle* defined by

$$\sigma(u,v) = (u, v, u^3 - 3v^2u)$$

which has a parabolic point at the origin, but Σ lies on both sides of the tangent plane in every open neighbourhood of the origin. At $p = \sigma(0,0)$, the Gauss curvature vanishes, but the surface lies locally on both sides of the tangent plane.

Proposition. Let Σ be a compact smooth surface in $\mathbb{R}^3$. Then Σ has an elliptic point.

Proof. Since Σ is compact, it is closed and bounded as a subset of $\mathbb{R}^3$. Hence, for R' sufficiently large, Σ lies entirely within $\overline{B(0, R')}$. Let R be the minimal such R'. Up to a global isometry of $\mathbb{R}^3$, there exists a point $p = (0, 0, R) \in \Sigma$ on the sphere $S^2(R)$ of radius R. Here, $T_p\Sigma = T_pS^2$. Hence, locally near p, we can view Σ as the graph of a smooth function $f : V \to \mathbb{R}^3$ on the x, y coordinates with the property that $f - \sqrt{R^2 - u^2 - v^2} \le 0$. This expresses the fact that Σ lies underneath the sphere of radius R.

We can now consider the Taylor series of f. Note that $(0,0)$ is a maximum point of f, hence $f_u = f_v = 0$ at 0. Thus, for sufficiently small u, v,

$$\frac{1}{2}\left(f_{uu}u^2 + 2f_{uv}uv + f_{vv}v^2\right) + \frac{1}{2R}(u^2 + v^2) \le 0$$

Hence, the second fundamental form is locally negative definite near $(0, 0)$. Hence, $\kappa(p) > 0$, so p is elliptic as required. In particular, the curvature at this point is greater than that of the sphere. $\quad\square$

Theorem. Let Σ be a smooth surface in $\mathbb{R}^3$, and let $p \in \Sigma$ such that $\kappa(p) \ne 0$. Let U be an open neighbourhood of p, and a decreasing sequence $A_i \subseteq U$ of neighbourhoods that 'shrink to p', in the sense that for all $\varepsilon > 0$, $A_i \subseteq B(p, \varepsilon)$ for sufficiently large i. Then,

$$|\kappa(p)| = \lim_{i \to \infty} \frac{\text{area}_{S^2}(n(A_i))}{\text{area}_{\Sigma}(A_i)}$$

In other words, the Gauss curvature is an infinitesimal measure of how much the Gauss map n distorts area.

Remark. Around hyperbolic points, the signed area of $n(A_i)$ is reversed, since curves γ reverse direction under n. We can alternatively define the *signed area* of $n(A_i)$ to be the area of $n(A_i)$ if $\kappa > 0$ and the negation of this area if $\kappa < 0$. The above theorem holds when $\kappa = 0$, but this will not be proven.

Proof. Let σ be an allowable parametrisation near $p \in \Sigma$. Using σ, we can define the open sets $\sigma^{-1}(A_i) = V_i \subset V$. Since the A_i shrink to p, we have that $\bigcap V_i = \{(0, 0)\}$. We have

$$\text{area}_{\Sigma}(A_i) = \int_{V_1} \sqrt{EG - F^2}\, du\, dv = \int_{V_i} \|\sigma_u \times \sigma_v\|\, du\, dv$$

Recall from the chain rule applied to $n \circ \gamma$ that

$$Dn\bigg|_{(u,v)}(\sigma_u) = n_u; \quad Dn\bigg|_{(u,v)}(\sigma_v) = n_v$$

Since $\kappa(p) = \kappa(\sigma(0,0)) \ne 0$, $n \circ \sigma : V \to S^2$ has derivative of rank 2. This defines an allowable parametrisation for an open neighbourhood of $n((0,0))$ by the inverse function theorem. Therefore,

$$\text{area}_{S^2}(n(A_i)) = \int_{V_i} \|n_u \times n_v\|\, du\, dv$$

28

for sufficiently large i such that $\sigma^{-1}A_i = V_i$ lies in the open neighbourhood of $(0,0)$ where $n \circ \sigma$ is a diffeomorphism.

$$\int_{V_i} \|n_u \times n_v\| \, \mathrm{d}u \, \mathrm{d}v = \int_{V_i} \|Dn(\sigma_u) \times Dn(\sigma_v)\| \, \mathrm{d}u \, \mathrm{d}v$$

$$= \int_{V_i} |\det(Dn)| \cdot \|\sigma_u \times \sigma_v\| \, \mathrm{d}u \, \mathrm{d}v$$

$$= \int_{V_i} |\kappa(u,v)| \cdot \|\sigma_u \times \sigma_v\| \, \mathrm{d}u \, \mathrm{d}v$$

As κ is continuous, given $\varepsilon > 0$ there exists $\delta > 0$ such that $|\kappa(u,v) - \kappa(0,0)| < \varepsilon$ for all $(u,v) \in B((0,0), \delta)$. In particular, for sufficiently large i, we have

$$|\kappa(u,v)| \in (|\kappa(p)| - \varepsilon, |\kappa(p)| + \varepsilon)$$

Hence,

$$(|\kappa(p)| - \varepsilon) \int_{V_i} \|\sigma_u \times \sigma_v\| \, \mathrm{d}u \, \mathrm{d}v \leq \int_{V_i} |\kappa(u,v)| \cdot \|\sigma_u \times \sigma_v\| \, \mathrm{d}u \, \mathrm{d}v$$

$$\leq (|\kappa(p)| + \varepsilon) \int_{V_i} \|\sigma_u \times \sigma_v\| \, \mathrm{d}u \, \mathrm{d}v$$

In other words,

$$|\kappa(p)| - \varepsilon \leq \frac{\mathrm{area}_{S^2}(n(A_i))}{\mathrm{area}_\Sigma(A_i)} \leq |\kappa(p)| + \varepsilon$$

Letting $i \to \infty$ gives the result as required. $\qquad\square$

> **Theorem** (*theorema egregium*). The Gauss curvature of a smooth surface in $\mathbb{R}^3$ is isometry invariant. In other words, if $f : \Sigma_1 \to \Sigma_2$ is a diffeomorphism of surfaces in $\mathbb{R}^3$ which is an isometry, then $\kappa(p) = \kappa(f(p))$ for all p.

Remark. Isometries rely on only the first fundamental form, but Gauss curvature is defined using both fundamental forms. We can do a direct proof by simply differentiating the formula and rearranging until the result follows. This proof is given in Part II.

Alternatively, we can consider a different question: are some allowable parametrisations of a smooth surface in $\mathbb{R}^3$ 'better' than others in some way? If we have a parametrisation $\sigma : V \to U \subseteq \Sigma$, this defines certain distinguished curves, which are the images of $\sigma(t, 0)$ and $\sigma(0, t)$. In this sense, looking for a 'best' parametrisation is equivalent to looking for 'best' distinguished curves near a point. This leads to the study of geodesics. We will later show that every smooth surface in $\mathbb{R}^3$ admits local parametrisations such that the first fundamental form has form $\mathrm{d}u^2 + G \, \mathrm{d}v^2$, so $E = 1$ and $F = 0$. We will also see (on an example sheet) that if such a local parametrisation exists, then κ can be expressed as a function just of G. This allows us to approach the proof of the *theorema egregium* from a more conceptual way, since we have expressed κ in terms of the first fundamental form alone.

5 Geodesics

5.1 Definitions

Recall that we defined, for a smooth curve $\gamma : [a, b] \to \mathbb{R}^3$,

$$\mathrm{length}(\gamma) = \int_a^b \|\gamma'(t)\| \, \mathrm{d}t$$

Definition. The *energy* of γ is given by

$$E(\gamma) = \int_a^b \|\gamma'(t)\|^2 \, \mathrm{d}t$$

Definition. Let $\gamma : [a, b] \to \Sigma$, where Σ is a smooth surface in $\mathbb{R}^3$. A *one-parameter variation* (with fixed endpoints) of γ is a smooth map $\Gamma : (-\varepsilon, \varepsilon) \times [a, b] \to \Sigma$, such that if $\gamma_s = \Gamma(s, \cdot)$, then $\gamma_0(t) = \gamma(t)$, and $\gamma_s(a)$ and $\gamma_s(b)$ are independent of s.

Definition. A smooth curve $\gamma : [a, b] \to \Sigma$ is a *geodesic* if, for every variation (γ_s) of γ with fixed endpoints as above, we have $\left.\dfrac{\mathrm{d}}{\mathrm{d}s}\right|_{s=0} E(\gamma_s) = 0$. Alternatively, γ is a critical point of the energy functional on curves from $\gamma(a)$ to $\gamma(b)$.

5.2 The geodesic equations

Let γ have image contained within the image of an allowable parametrisation $\sigma : V \to U$. Then, for sufficiently small s, we can write $\gamma_s(t) = \sigma(u(s, t), v(s, t))$. Suppose that the first fundamental form, with respect to σ, is

$$E \, \mathrm{d}u^2 + 2F \, \mathrm{d}u \, \mathrm{d}v + G \, \mathrm{d}v^2$$

Let

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

By definition,

$$E(\gamma_s) = \int_a^b R \, \mathrm{d}t$$

where R depends on s. Hence,

$$\frac{\partial R}{\partial s} = \left(E_u \dot{u}^2 + 2F_u \dot{u}\dot{v} + G_u \dot{v}^2\right)\frac{\partial u}{\partial s} + \left(E_v \dot{v}^2 + 2F_v \dot{u}\dot{v} + G_v \dot{v}^2\right)\frac{\partial v}{\partial s}$$
$$+ 2(E\dot{u} + F\dot{v})\frac{\partial \dot{u}}{\partial s} + 2(F\dot{u} + G\dot{v})\frac{\partial \dot{v}}{\partial s}$$

This gives

$$\frac{\mathrm{d}}{\mathrm{d}s}E(\gamma_s) = \int_a^b \frac{\partial R}{\partial s}\,\mathrm{d}t$$

We can integrate by parts. Note that $\frac{\partial u}{\partial s}$ and $\frac{\partial v}{\partial s}$ vanish at a, b. Hence,

$$\frac{\mathrm{d}}{\mathrm{d}s}\bigg|_{s=0} E(\gamma_s) = \int_a^b \left(A\frac{\partial u}{\partial s} + B\frac{\partial v}{\partial s}\right)\mathrm{d}t$$

where

$$A = E_u \dot{u}^2 + 2F_u \dot{u}\dot{v} + G_u \dot{v}^2 - 2\frac{\partial}{\partial t}(E\dot{u} + F\dot{v})$$
$$B = E_v \dot{u}^2 + 2F_v \dot{u}\dot{v} + G_v \dot{v}^2 - 2\frac{\partial}{\partial t}(F\dot{u} + G\dot{v})$$

> **Corollary.** A smooth curve $\gamma \colon [a,b] \to \Sigma$ with image in $\mathrm{Im}\,\sigma$ is a geodesic if and only if it satisfies the *geodesic equations*:
>
> $$\frac{\mathrm{d}}{\mathrm{d}t}(E\dot{u} + F\dot{v}) = \frac{1}{2}\left(E_u \dot{u}^2 + 2F_u \dot{u}\dot{v} + G_u \dot{v}^2\right)$$
> $$\frac{\mathrm{d}}{\mathrm{d}t}(F\dot{u} + G\dot{v}) = \frac{1}{2}\left(E_v \dot{u}^2 + 2F_v \dot{u}\dot{v} + G_v \dot{v}^2\right)$$
>
> Note that these equations are evaluated at $s = 0$, so no choice of variation is required.

Remark. Solving a differential equation is a local procedure. The original definition of the geodesic seems to be a global property. However, we can always consider a sub-curve of γ to also be a geodesic, since its variations are variations of γ. So the definition can be thought of as local.

Energy is sensitive to reparametrisation. If $f, g \colon [a,b] \to \mathbb{R}$ are smooth, the Cauchy–Schwarz inequality gives that

$$\left(\int_a^b fg\,\mathrm{d}t\right)^2 \leq \int_a^b f^2\,\mathrm{d}t \cdot \int_a^b g^2\,\mathrm{d}t$$

Let us apply this to $f = \sqrt{R}, g = 1$ to find

$$\mathrm{length}(\gamma)^2 \leq E(\gamma)(b - a)$$

Since equality holds only when the two functions are proportional, we must have that $\|\gamma'(t)\|$ is constant for the equality to hold. In other words, γ must be parametrised proportional to arc length.

> **Corollary.** If γ has constant speed and locally minimises length, then it is a geodesic. Further, if γ globally minimises energy, then it must globally minimise length, and is parametrised with constant speed.

Remark. We would like geodesics to be a local property, but not necessarily global length minimisers. For example, all arcs of great circles will be shown to be geodesics, even if large arcs are not global length minimisers between fixed endpoints.

5.3   Geodesics on the plane

The plane $\mathbb{R}^2$ has parametrisation $\sigma(u, v) = (u, v, 0)$ and first fundamental form $\mathrm{d}u^2 + \mathrm{d}v^2$. The geodesic equations here are

$$\ddot{u} = 0; \quad \ddot{v} = 0$$

In particular, the geodesics on the plane are given by

$$u(t) = \alpha t + \beta; \quad v(t) = \gamma t + \delta$$

This is a straight line, parametrised at constant speed.

5.4   Geodesics on the sphere

Consider the unit sphere with parametrisation

$$\sigma(u, v) = (\cos u \cos v, \cos u \sin v, \sin u); \quad u \in \left(-\frac{\pi}{2}\right); \quad v \in (0, 2\pi)$$

This has first fundamental form

$$\mathrm{d}u^2 + \cos^2 u\, \mathrm{d}v^2 \implies E = 1;\ F = 0;\ G = \cos^2 u$$

The geodesic equations give

$$\frac{\mathrm{d}}{\mathrm{d}t}(\dot{u}) = \frac{1}{2} 2 \cos u \sin u \dot{v}^2; \quad \frac{\mathrm{d}}{\mathrm{d}t}(\cos^2 u \dot{v}) = 0$$

This gives

$$\ddot{u} + \sin u \cos u \dot{v}^2 = 0; \quad \ddot{v} - 2 \tan u \dot{u} \dot{v} = 0$$

Since geodesics are parametrised at constant speed, we can assume that it is parametrised at unit speed without loss of generality.

$$\|\gamma'(t)\| = 1 \implies \dot{u} + \cos^2 u \dot{v}^2 = 1$$

Hence,

$$\frac{\ddot{v}}{\dot{v}} = 2 \tan u \dot{u} \implies \ln \dot{v} = -2 \ln \cos u + \text{constant} \implies \dot{v} = \frac{C}{\cos^2 u}$$

Substituting into the unit speed equation,

$$\dot{u}^2 = 1 - \frac{C^2}{\cos^2 u} \implies \dot{u} = \sqrt{\frac{\cos^2 u - C^2}{\cos^2 u}}$$

Then,

$$\frac{\dot{v}}{\dot{u}} = \frac{\mathrm{d}v}{\mathrm{d}u} = \frac{C}{\cos u \sqrt{\cos^2 u - C^2}}$$

Hence,

$$v = \int \frac{\mathrm{d}v}{\mathrm{d}u} \, \mathrm{d}u = \int \frac{C \sec^2 u}{\sqrt{1 - C^2 sec^2 u}} \, \mathrm{d}u$$

Using the substitution $w = \frac{C \tan u}{\sqrt{1-C^2}}$, we find

$$v = \int \frac{w}{1 - w^2} \, \mathrm{d}w = \arcsin w + \text{constant} = \arcsin(\lambda \tan u) + \delta$$

for some constants λ, δ. Hence,

$$\sin(v - \delta) = \lambda \tan u$$

Rewriting using the angle addition formula,

$$\underbrace{(\sin v \cos u)}_{x} \cos \delta - \underbrace{(\cos v \cos u)}_{y} \sin \delta - \lambda \underbrace{\sin u}_{z} = 0$$

Hence, the geodesic γ lies on a plane through the origin, since this is a linear equation in x, y, z. Such planes intersect the sphere in great circles.

5.5   Geodesics on the torus

Consider the surface of revolution of a circle in the xz-plane centred at $(a, 0, 0)$ about the z axis, giving a torus. An allowable parametrisation for this surface is

$$\sigma(u, v) = ((a + \cos u) \cos v, (a + \cos u) \sin v, \sin u)$$

The first fundamental form is

$$\mathrm{d}u^2 + (a + \cos u)^2 \, \mathrm{d}v^2 \implies E = 1; \ F = 0; \ G = (a + \cos u)^2$$

Note that if we were to take $a = 0$, we would arrive at the unit sphere and its first fundamental form. We can follow the same procedure as above with the sphere, or formally replace $\cos u$ with $a + \cos u$ in the result.

$$\frac{\mathrm{d}v}{\mathrm{d}u} = \frac{C}{(a + \cos u)\sqrt{(a + \cos u)^2 - C^2}}$$

which cannot be integrated using classical functions. This leads to the study of elliptic functions.

5.6   Equivalent characterisation of geodesics

We have so far restricted our analysis to the first fundamental form, without considering its embedding in $\mathbb{R}^3$. Intuitively, we know that straight lines in $\mathbb{R}^2$ are not just locally shortest but also locally straightest. We would expect this to hold for other surfaces as well. We can characterise this notion via stating that the change in the tangent vector to a curve is as small as it could be, subject to the constraint that it lies on the surface.

Proposition. Let Σ be a smooth surface in $\mathbb{R}^3$. A smooth curve $\gamma : [a, b] \to \Sigma$ is a geodesic if and only if $\ddot{\gamma}(t)$ is everywhere normal to the surface Σ.

Remark. This proposition makes use of the tangent plane, a notion that exists only because we have an embedding in $\mathbb{R}^3$. Note that

$$\frac{\mathrm{d}}{\mathrm{d}t} \langle \dot{\gamma}, \dot{\gamma} \rangle = 2 \left\langle \underset{\text{tangent to } \Sigma}{\dot{\gamma}} , \underset{\text{normal to } \Sigma}{\ddot{\gamma}} \right\rangle = 0$$

Hence, $\langle \dot{\gamma}, \dot{\gamma} \rangle$ is constant, giving that geodesics are parametrised proportional to arc length.

Proof. The property of being a geodesic as we previously defined is a local property, and so is the condition in the proposition. Hence, we may work entirely within an allowable parametrisation $\sigma : V \to U$. Suppose $\gamma(t) = \sigma(u(t), v(t))$. Hence,

$$\dot{\gamma} = \sigma_u \dot{u} + \sigma_v \dot{v}$$

$\ddot{\gamma}$ is normal to Σ when it is orthogonal to the tangent plane, which is spanned by σ_u, σ_v. This is true if and only if

$$\left\langle \frac{\mathrm{d}}{\mathrm{d}t}(\sigma_u \dot{u} + \sigma_v \dot{v}), \sigma_u \right\rangle = 0 = \left\langle \frac{\mathrm{d}}{\mathrm{d}t}(\sigma_u \dot{u} + \sigma_v \dot{v}), \sigma_v \right\rangle$$

We will prove the first equality. This can be rewritten

$$\frac{\mathrm{d}}{\mathrm{d}t} \langle \sigma_u \dot{u} + \sigma_v \dot{v}, \sigma_u \rangle - \left\langle \sigma_u \dot{u} + \sigma_v \dot{v}, \frac{\mathrm{d}}{\mathrm{d}t} \sigma_u \right\rangle = 0$$

Note that $\langle \sigma_u, \sigma_u \rangle = E$ and $\langle \sigma_u, \sigma_v \rangle = F$.

$$\frac{\mathrm{d}}{\mathrm{d}t}(E\dot{u} + F\dot{v}) - \langle \sigma_u \dot{u} + \sigma_v \dot{v}, \sigma_{uu} \dot{u} + \sigma_{uv} \dot{v} \rangle = 0$$

Hence,

$$\frac{\mathrm{d}}{\mathrm{d}t}(E\dot{u} + F\dot{v}) - \left[\dot{u}^2 \langle \sigma_u, \sigma_{uu} \rangle + \dot{u}\dot{v}(\langle \sigma_u, \sigma_{uv} \rangle + \langle \sigma_v, \sigma_{uu} \rangle) + \dot{v}^2 \langle \sigma_v \sigma_{uv} \rangle \right] = 0$$

Note that $E_u = 2 \langle \sigma_u, \sigma_{uu} \rangle$, $F_u = \langle \sigma_u, \sigma_{uv} \rangle + \langle \sigma_v, \sigma_{uu} \rangle$, and $G_u = 2 \langle \sigma_v, \sigma_{uv} \rangle$. This gives

$$\frac{\mathrm{d}}{\mathrm{d}t}(E\dot{u} + F\dot{v}) = \frac{1}{2}\left(E_u \dot{u}^2 + 2F_u \dot{u}\dot{v} + G_u \dot{v}^2 \right)$$

which is the first of the geodesic equations. By symmetry, we find the second geodesic equation similarly. $\square$

5.7 Planes of symmetry

Let Σ be a smooth surface in $\mathbb{R}^3$ such that there exists a plane $\Pi \subseteq \mathbb{R}^3$ such that $\Pi \cap \Sigma$ is a smooth embedded curve $C \subseteq \Sigma$, and Σ is setwise preserved by reflection in the plane Π. We will show that C is a geodesic when parametrised at constant speed. Consider a point p on C. We can think of $\mathbb{R}^3 = \Pi \oplus \Pi^\perp$, where we change coordinates such that p is the origin. We can also write $\mathbb{R}^3 = T_p\Sigma \oplus \mathbb{R}n_p$, where $\mathbb{R}n_p$ is the vector subspace of $\mathbb{R}^3$ generated by n_p. Clearly, reflection in Π acts on Π by the identity, and on $\Pi^\perp$ by -1. Since reflection in Π fixes Σ setwise and fixes p, it must also preserve the

subspace $T_p\Sigma$. Hence it also preserves $\mathbb{R}n_p$, so $\mathbb{R}n_p \subseteq \Pi$, since Π is not the identity on $T_p\Sigma$. Now, let us parametrise C locally near p using $t \mapsto \gamma(t) \in C$ at constant speed. Since $\gamma(t) \subseteq \Pi$, we have $\dot{\gamma}(t), \ddot{\gamma}(t) \in \Pi$. γ has constant speed, so $\langle \dot{\gamma}, \ddot{\gamma} \rangle = 0$. Hence $\dot{\gamma}$ lies in $\Pi \cap T_p\Sigma$ and $\ddot{\gamma}$ is orthogonal to this and lies in Π, so lies in $\mathbb{R}n_p \subseteq \Pi$. Hence γ is indeed a geodesic.

In particular, arcs of great circles are geodesics, since they lie in planes of symmetry.

5.8 Surfaces of revolution

Consider the surface of revolution given by $\eta(u) = (f(u), 0, g(u))$ where η is smooth and injective, and $f(u) > 0$, rotated about the z axis.

> **Definition.** A circle obtained by rotating a point of η is called a *parallel*. A curve obtained by rotating η itself by a fixed angle about the z axis is called a *meridian*.

A plane in $\mathbb{R}^3$ containing the z axis is a plane of symmetry, hence meridians are geodesics by the previous discussion. Not all parallels are geodesics.

> **Lemma.** A parallel given by $u = u_0$ is a geodesic when parametrised at constant speed if and only if $f'(u_0) = 0$.

Proof. Consider the allowable parametrisation

$$\sigma(u, v) = (f(u) \cos v, f(u) \sin v, g(u))$$

where $u \in (a, b)$ and $v \in (0, 2\pi)$. The first fundamental form is

$$\left[(f')^2 + (g')^2 \right] du^2 + f^2 dv^2$$

If without loss of generality we choose to parametrise η by arc length, this becomes

$$du^2 + f^2 dv^2 \implies E = 1;\ F = 0;\ G = f^2$$

The geodesic equations are

$$\frac{d}{dt}\dot{u} = \ddot{u} = f f_u \dot{v}^2;\quad \frac{d}{dt}(f^2 \dot{v}) = 0$$

and $\dot{u}^2 + f^2\dot{v}^2$ is a nonzero constant. Given that we want to only consider parallels of the surface of revolution, we can impose the constraint that $u = u_0$ is constant. Hence, the constant speed condition gives that

$$\dot{v} = \frac{\text{constant}}{f(u_0)} = \text{constant}$$

The second equation holds automatically. The first equation is

$$0 = f(u_0) f_u(u_0) \cdot \text{constant}$$

So this holds exactly when $f_u(u_0) = 0$. $\qquad\square$

Consider a curve $\gamma(t)$ on Σ, making angle θ with the parallel of radius $\rho = f$.

> **Proposition** (Clairaut's relation). If γ is a geodesic, then $\rho \cos \theta$ is constant along γ.

Proof. Let $\gamma(t) = \sigma(u(t), v(t))$, so $\dot{\gamma} = \sigma_u \dot{u} + \sigma_v \dot{v}$. The tangent vector to the parallel is σ_v. By the earlier discussion on angles in terms of the first fundamental form,

$$\cos \theta = \frac{\langle \sigma_v, \sigma_u \dot{u} + \sigma_v \dot{v} \rangle}{\|\sigma_v\| \cdot \|\sigma_u \dot{u} + \sigma_v \dot{v}\|}$$

If γ is parametrised by arc length, $\|\dot{\gamma}\| = 1$, so $\|\sigma_u \dot{u} + \sigma_v \dot{v}\| = 1$. So, using our values for F, G above,

$$\cos \theta = |f(u)\dot{v}| = \rho \dot{v}$$

The second geodesic equation is exactly

$$\rho \cos \theta = \text{constant}$$

$\square$

Example. Usually, for a surface of revolution, we take the assumption that η never intersects the z-axis, or that f is positive. This ensures that all points on the surface are locally smooth. However, we can allow η to meet the z-axis orthogonally, as in the ellipsoid or sphere.

Consider an ellipsoid of revolution. $\rho \cos \theta$ is constant along a geodesic γ. Suppose that at some point γ intersects a parallel of radius ρ_0 at angle θ_0, and that γ is not a meridian (so $\cos \theta \neq 0$). Hence $\theta_0 \in \left[0, \frac{\pi}{2}\right)$. In particular, for $\rho \cos \theta$ to be constant, we must have that ρ is bounded below. A geodesic which is not a meridian is therefore 'trapped' between parallels corresponding to the bound on the size of ρ. In particular, any geodesic through a pole is a meridian.

5.9   Local existence of geodesics

It is difficult to solve the geodesic equations globally. We can often intead prove local results about any geodesics that may arise.

Recall Picard's theorem from Analysis and Topology. Let $I = [t_0 - a, t_0 + a] \subseteq \mathbb{R}, B = \{x : \|x - x_0\| \leq b\} \subseteq \mathbb{R}^n$, and $f : I \times B \to \mathbb{R}^n$ that is continuous, and Lipschitz in the second variable.

$$\|f(t, x_1) - f(t, x_2)\| \leq N \|x_1 - x_2\|$$

Then the differential equation $\frac{dx}{dt} = f(t, x)$ with $x(t_0) = x_0$ has a unique solution for some time interval $|t - t_0| < h$, where $h = \min\left\{a, \frac{b}{s}\right\}$ where $s = \sup \|f\|$. Further, if f is smooth in all parameters, then the solution to the differential equation is smooth and depends smoothly on the initial condition.

Recall the geodesic equations:

$$\frac{d}{dt}(E\dot{u} + F\dot{v}) = \frac{1}{2}\left(E_u \dot{u}^2 + 2F_u \dot{u}\dot{v} + G_u \dot{v}^2\right)$$
$$\frac{d}{dt}(F\dot{u} + G\dot{v}) = \frac{1}{2}\left(E_v \dot{u}^2 + 2F_v \dot{u}\dot{v} + G_v \dot{v}^2\right)$$

We can write this as

$$\begin{pmatrix} E & F \\ F & G \end{pmatrix} \begin{pmatrix} \ddot{u} \\ \ddot{v} \end{pmatrix} = R$$

where R is composed of smooth functions of u, v. The matrix on the left hand side is invertible, and the inverse map $A \mapsto A^{-1}$ on matrices is smooth. Hence, we can write the geodesic equations in the form

$$\ddot{u} = A(u, v, \dot{u}, \dot{v}); \quad \ddot{v} = B(u, v, \dot{u}, \dot{v})$$

In the usual way we can turn second-order equations into first-order equations by introducing $p = \dot{u}, q = \dot{v}$, and we find

$$\dot{u} = p; \quad \dot{v} = q; \quad \dot{p} = A(u, v, p, q); \quad \dot{q} = B(u, v, p, q)$$

This is a system of first-order ordinary differential equations as governed by Picard's theorem. Since A, B are smooth, a local bound on $\|DA\|$ and $\|DB\|$ will give the required Lipschitz condition.

Corollary. Let Σ be a smooth surface in $\mathbb{R}^3$. For $p \in \Sigma$ and $v \in T_p\Sigma$ nonzero, then there exists $\varepsilon > 0$ and a geodesic $\gamma : [0, \varepsilon) \to \Sigma$ such that

$$\gamma(0) = p; \quad \dot{\gamma}(0) = v$$

Moreover, this geodesic depends smoothly on p, v.

The local existence of geodesics gives rise to allowable parametrisations of Σ with 'nice' properties in terms of the first fundamental form. Let $p \in \Sigma$, and consider a geodesic arc γ starting at p and parametrised by arc length. At each point $\gamma(t)$ for small $t > 0$, we can consider a geodesic arc γ_t starting at $\gamma(t)$, and $\gamma_t'(0)$ is orthogonal to $\gamma'(t)$, and also parametrised by arc length. Now, we define $\sigma(u, v) = \gamma_v(u)$, which is defined for $u \in [0, \varepsilon)$ and $v \in [0, \delta)$.

Lemma. For ε, δ sufficiently small, $\sigma : (u, v) \mapsto \gamma_v(u)$ defines an allowable parametrisation of an open set in Σ, taking the interior of the domain.

Proof. Smoothness follows from the addendum to Picard's theorem above. At the origin $(0, 0)$, by construction we have σ_u, σ_v orthogonal. Hence, they stay linearly independent for sufficiently small ε, δ. So $D\sigma$ has full rank, and (on a smaller set if necessary) σ is injective. So σ is allowable. $\qquad\square$

Corollary. Any smooth surface Σ in $\mathbb{R}^3$ admits local parametrisations for which the first fundamental form has form $\mathrm{d}u^2 + G(u, v)\,\mathrm{d}v^2$, so $E = 1$ and $F = 0$.

Proof. Consider the parametrisation $\sigma(u, v) = \gamma_v(u)$. For v_0 fixed, the curve $u \mapsto \gamma_{v_0}(u)$ is a geodesic parametrised at unit speed, so $E = 1$. One of the geodesic equations is

$$\frac{\mathrm{d}}{\mathrm{d}t}(F\dot{u} + G\dot{v}) = \frac{1}{2}\left(E_v\dot{u}^2 + 2F_v\dot{u}\dot{v} + G_v\dot{v}^2\right)$$

and consider $v(t) = v_0, u(t) = t$. $E_v = \dot{v} = 0$ and $\dot{u} = 1$, so

$$\frac{\mathrm{d}}{\mathrm{d}t}F = 0 \implies F_u\dot{u} = 0 \implies F_u = 0$$

So F is independent of u. At $u = 0$, then by construction of γ_v as being orthogonal to γ at $\gamma(v)$, we see $F = 0$. $\qquad\square$

These coordinates are called *geodesic normal coordinates*. Note that by fixing u and letting v vary, the curve obtained is typically not a geodesic, except for $u = 0$ which is γ itself. In these coordinates, we can also find

$$G(0, v) = 1; \quad G_u(0, v) = 0$$

The first result holds since σ_v has unit length at $u = 0$. The second result holds because $u = 0$ yields a geodesic with arc length parametrisation, and then we can use one of the geodesic equations to find

$$\frac{\mathrm{d}}{\mathrm{d}t}(E\dot{u} + F\dot{v}) = \frac{1}{2}(E_u\dot{u}^2 + 2F_u\dot{u}\dot{v} + G_u\dot{v}^2) \implies 0 = \frac{1}{2}G_u(0, v)$$

5.10   Surfaces of constant curvature

In the example sheets, we show that for a smooth surface Σ in $\mathbb{R}^3$ with allowable parametrisation for which $E = 1$ and $F = 0$, we have the following result for the Gauss curvature.

$$\kappa = \frac{-\left(\sqrt{G}\right)_{uu}}{\sqrt{G}}$$

If $a : \mathbb{R}^3 \to \mathbb{R}^3$ is a dilation $a(x, y, z) = (ax, ay, az)$, then

$$\kappa_{a(\Sigma)} = \frac{1}{a^2}\kappa_\Sigma$$

since E, F, G rescale by a^2, and L, N, M rescale by a. This matches the results previously found for spheres of varying radii. By dilating, to understand surfaces of constant curvature it suffices to consider surfaces with constant curvature ± 1 or 0.

> **Proposition.** Let Σ be a smooth surface in $\mathbb{R}^3$. Then,
>   (i) if $\kappa \equiv 0$, then Σ is locally isometric to $(\mathbb{R}^2, \mathrm{d}u^2 + \mathrm{d}v^2)$;
>   (ii) if $\kappa \equiv 1$, then Σ is locally isometric to $(S^2, \mathrm{d}u^2 + \cos^2 u \, \mathrm{d}v^2)$.

Proof. Σ admits an allowable parametrisation with $E = 1$ and $F = 0$ by using geodesic normal coordinates, so

$$\kappa = \frac{-\sqrt{G}_{uu}}{\sqrt{G}}; \quad G(0, v) = 1; \quad G_u(0, v) = 0$$

If $\kappa \equiv 0$, we have $\sqrt{G}_{uu} = 0$, so $\sqrt{G} = A(v)u + B(v)$, and the boundary conditions give $A \equiv 0, B \equiv 1$. In particular, $G \equiv 1$. The fundamental form then is $\mathrm{d}u^2 + \mathrm{d}v^2$, which is that of $\mathbb{R}^2$.

If $\kappa \equiv 1$, we find $\left(\sqrt{G}\right)_{uu} + \sqrt{G} = 0$ so $\sqrt{G} = A(v)\sin u + B(v)\cos u$. The boundary conditions then imply that $A \equiv 0, B \equiv 1$ and hence the fundamental form is $\mathrm{d}u^2 + \cos^2 u \, \mathrm{d}v^2$. This matches the first fundamental form of a sphere with parametrisation

$$\sigma(u, v) = (\cos u \cos v, \cos u \sin v, \sin u)$$

$\qquad\square$

Remark. If $\kappa \equiv -1$, we will find the first fundamental form $\mathrm{d}u^2 + \cosh^2 u\, \mathrm{d}v^2$. There exists an object known as the tractoid, which is a smooth surface in $\mathbb{R}^3$, and has this first fundamental form. We could alternatively choose not to embed this surface in $\mathbb{R}^3$.

In fact, the change of variables $v = e^v \tanh u, w = e^v \operatorname{sech} u$ turns the fundamental form $\mathrm{d}u^2 + \cosh^2 u\, \mathrm{d}v^2$ into $\frac{\mathrm{d}V^2 + \mathrm{d}W^2}{W^2}$, which is a 'standard' presentation of the first fundamental form, which we will see more of later.

6 Riemannian metrics

6.1 Definitions

Definition. Let $V \subseteq \mathbb{R}^2$ be an open set. An *(abstract) Riemannian metric* is a smooth map from V to the set of positive definite symmetric bilinear forms, given by

$$v \mapsto \begin{pmatrix} E(v) & F(v) \\ F(v) & G(v) \end{pmatrix}$$

such that $E > 0, G > 0, EG - F^2 > 0$. The image of this map can be viewed as an open subset of $\mathbb{R}^4$.

If v is a vector at $p \in V$, we can compute its infinitesimal length by

$$\|v\|^2 = v^\mathsf{T} \begin{pmatrix} E(v) & F(v) \\ F(v) & G(v) \end{pmatrix} v$$

Thus, if $\gamma : [a, b] \to V$ is smooth,

$$\text{length}(\gamma) = \int_a^b \left(E\dot{u}^2 + 2F\dot{u}\dot{v} + G\dot{v}^2 \right)^{\frac{1}{2}} \mathrm{d}t$$

where $\gamma(t) = (u(t), v(t))$.

Definition. Let Σ be an abstract smooth surface, so $\Sigma = \bigcup_{i \in I} U_i$ for open sets U_i, with charts $\varphi_i : U_i \to V_i \subseteq \mathbb{R}^2$ which are homeomorphisms, and with smooth transition maps $\varphi_i \varphi_j^{-1} : \varphi_j(U_i \cap U_j) \to \varphi_i(U_i \cap U_j)$. A *Riemannian metric* on Σ, usually called g or $\mathrm{d}s^2$, is a choice of Riemannian metric in the above sense on each V_i, which are compatible in the following sense. Let $\sigma = \varphi_i^{-1}$ and $\tilde{\sigma} = \varphi_j^{-1}$ for some i, j, and define $f = \tilde{\sigma}^{-1} \circ \sigma$. Then we require

$$(Df)^\mathsf{T} \begin{pmatrix} \widetilde{E} & \widetilde{F} \\ \widetilde{F} & \widetilde{G} \end{pmatrix} (Df) = \begin{pmatrix} E & F \\ F & G \end{pmatrix}$$

So Df defines an isometry from an open set in the chart $(U, \varphi(U) = V)$ to one in the chart $(\widetilde{U}, \widetilde{\varphi}(\widetilde{U}) = \widetilde{V})$.

This compatibility condition is the transition law for first fundamental forms for smooth surfaces in $\mathbb{R}^3$.

Example. Recall the torus $T^2 = \mathbb{R}^2 / \mathbb{Z}^2$.
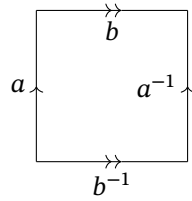
We have an atlas of charts for which the transition maps are the restrictions of translations of open subsets of $\mathbb{R}^2$. For each $V_i \subseteq \mathbb{R}^2$, we associate the natural Euclidean metric $du^2 + dv^2$. If f is a translation, Df is the identity, and so

$$(Df)^\mathsf{T} I (Df) = I$$

holds trivially. So this gives a global Riemannian metric on T^2. This metric is flat, since it is locally isometric to $\mathbb{R}^2$ at all points.

Conversely, consider the torus of revolution embedded in $\mathbb{R}^3$. As a compact smooth surface in $\mathbb{R}^3$, it must contain an elliptic point. Hence, the flat Riemannian metric described above is not the same (up to isometry) as the metric obtained by any possible embedding of the torus in $\mathbb{R}^3$.

The real projective plane $\mathbb{RP}^2$ admits a Riemannian metric with constant curvature $+1$. We have constructed a smooth atlas for $\mathbb{RP}^2$ where the charts were of the form (U, φ), with $U = q\hat{U}$ and $q : S^2 \to \mathbb{RP}^2$ the quotient map, $\hat{U} \subseteq S^2$ open and contained within an open hemisphere, and $\varphi : U : U \to V \subseteq \mathbb{R}^2$ is given by $\hat{\varphi} \circ q^{-1}|_U$ and $\hat{\varphi} : \hat{U} \to V$ a chart on S^2. The transition maps for this atlas were found to be locally the identity, or induced from the antipodal map. The antipodal map from S^2 to S^2 is an isometry, so both types of transition maps preserve the usual round metric on S^2.

In the first example sheet, we consider the Klein bottle. This has an atlas such that all transition maps are either translations or translations composed with a reflection. These preserve the flat metric in $\mathbb{R}^2$, so the Klein bottle inherits a flat Riemannian metric. The Klein bottle and $\mathbb{RP}^2$ are not embedded in $\mathbb{R}^3$, so we could not construct a 'non-abstract' Riemannian metric.

> **Definition.** Let $(\Sigma_1, g_1), (\Sigma_2, g_2)$ be abstract smooth surfaces with abstract Riemannian metrics. A diffeomorphism $f : \Sigma_1 \to \Sigma_2$ is an *isometry* if it preserves the lengths of all curves, where lengths are taken with respect to these abstract Riemannian metrics.

Example. If (Σ_2, g_2) is given, and $f : \Sigma_1 \to \Sigma_2$ is a diffeomorphism, we can equip Σ_1 with a metric known as the *pullback* metric $g_1 = f^\star g_2$ that gives that f is an isometry.

6.2 The length metric

> **Definition.** Let (Σ, g) be a connected abstract smooth surface with an abstract Riemannian metric. The *length metric* is defined by
>
> $$d_g(p, q) = \inf_\gamma L(\gamma)$$
>
> where γ varies over piecewise smooth paths in Σ from p to q, and L is length computed using g.

Proposition. Let (Σ, g) be a connected abstract smooth surface with an abstract Riemannian metric. Then d_g is indeed a metric, and d_g induces a topology on Σ that agrees with the given topology.

Proof. Let $p, q \in \Sigma$. We will show that there exists some piecewise smooth path γ from p to q, so $d_g(p, q)$ is well-defined and finite. Connected surfaces are path-connected. There exists a continuous path γ and a finite set of charts (U_i, φ_i) with associated parametrisations $\sigma_i = \varphi_i^{-1} : V_i \to U_i \subset \Sigma$ such that $\text{Im}\,\gamma \subseteq \bigcup_{i=1}^N U_i$. Consider points

$$p = x_0 \in U_1, x_1 \in U_1 \cap U_2, x_2 \in U_2 \in U_3, \dots, q = x_N \in U_N$$

Smooth paths in V_i from $\varphi_i(x_i)$ to $\varphi_{i+1}(x_{i+1})$ exist, since smooth paths between two points in $\mathbb{R}^2$ exist. Since the atlas is smooth, being a smooth path in some U_i is the same as being smooth in U_{i+1} whenever U_i and U_{i+1} intersect, since the transition maps are smooth. So $p, q \in \Sigma$ are joined by some piecewise smooth path.

For any piecewise smooth path from p to q there exists the inverse path parametrised in the opposite direction, which has the same length. We can also concatenate paths from p to q and from q to r, with length equal to the sum of the lengths. In both cases, the new paths are piecewise smooth. This then implies that d_g is symmetric, and satisfies the triangle inequality.

To show d_g is a metric, it now suffices to show that $d_g(p, q) = 0$ implies $p = q$, since the converse is trivial. Let $p \in \Sigma$ and fix a chart (U, φ) at p. Without loss of generality let $V = B(0, 1)$, and $\varphi(p) = 0$. If $q \neq p \in \Sigma$, there exists $\varepsilon > 0$ such that $q \notin \varphi^{-1}\left(\overline{B(0, \varepsilon)}\right)$. Suppose $\gamma : [0, 1] \to \Sigma$ is a piecewise smooth path from p to q. Certainly, γ must escape the disc $\varphi^{-1}\left(\overline{B(0, \varepsilon)}\right)$, since it must reach q. Length along paths is additive, so by the triangle inequality, it suffices to show that there exists $\delta > 0$ such that $d_g(p, r) > \delta$ for all $r \in \partial\varphi^{-1}\left(\overline{B(0, \varepsilon)}\right) = \varphi^{-1}\{\text{circle of radius } \varepsilon\}$. The data on the Riemannian metric g includes the non-degenerate symmetric bilinear form $\begin{pmatrix} E_z & F_z \\ F_z & G_z \end{pmatrix}$ for all $z \in \overline{B(0, \varepsilon)} \subseteq V$. We also have the usual Euclidean inner product on the disc, $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$. For all $z \in \overline{B(0, \varepsilon)}$, these matrices are positive definite. Since $\overline{B(0, \varepsilon)}$ is compact, there exists $\delta > 0$ such that $\begin{pmatrix} E_z - \delta & F_z \\ F_z & G_z - \delta \end{pmatrix}$ is still positive definite for all $z \in \overline{B(0, \varepsilon)}$. In other words, the determinant $EG - F^2 > 0$ for all $z \in \overline{B(0, \varepsilon)}$, which is compact, so it is bounded below by some positive number. Hence, $\text{length}_g(\hat{\gamma}) \geq \text{length}_{\delta \cdot \text{Euclidean}}(\hat{\gamma})$ for any $\hat{\gamma}$ contained withing $\overline{B(o, \varepsilon)}$. Taking $\hat{\gamma} = \varphi\left[\gamma \cap \varphi^{-1}\left(\overline{B(o, \varepsilon)}\right)\right]$, which is the part of γ in $\overline{B(0, \varepsilon)}$ with respect to the chart, we have that $\text{length}_{\delta \cdot \text{Euclidean}}(\hat{\gamma}) \geq \delta\varepsilon$, so $d_g(p, q) \geq \delta\varepsilon$. $\square$

Remark. The last step of the argument for the proof above, comparing the inner products $\begin{pmatrix} E_z & F_z \\ F_z & G_z \end{pmatrix}$ and $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ can be modified to show that d_g induces a topology on Σ that agrees with the given topology, which is given by local homeomorphisms to $\mathbb{R}^2$ everywhere.

6.3 The hyperbolic metric

Definition. Let
$$D = B(0,1) = \{z \in \mathbb{C} : |z| < 1\}$$

The abstract Riemannian metric g_{hyp} on D is given by

$$\frac{4(du^2 + dv^2)}{(1 - u^2 - v^2)^2} = \frac{4|dz|^2}{\left(1 - |z|^2\right)^2}$$

Since there is only one chart, this holds for all of D. In particular, if $\gamma : [0,1] \to D$ is smooth, then

$$L_{g_{\text{hyp}}}(\gamma) = 2 \int_0^1 \frac{|\dot{\gamma}(t)|}{1 - |\gamma(t)|^2} \, dt$$

If $\gamma(t) = (u(t), v(t))$, we can write

$$L(\gamma) = 2 \int_0^1 \frac{(\dot{u}^2 + \dot{v}^2)^{\frac{1}{2}}}{1 - u^2 - v^2} \, dt$$

This is very similar to a first fundamental form with $E = G = \frac{4}{(1-u^2-v^2)^2}$ and $F = 0$, but we do not claim that this fundamental form arises from an embedding in $\mathbb{R}^3$.

Note that the flat metric on $\mathbb{R}^2$ and the usual round metric on S^2 have large and transitive isometry groups. We will show that this metric also induces a large symmetry group, which is induced by the Möbius group. Recall that

$$\text{Möb} = \left\{ z \mapsto \frac{az + b}{cz + d} : \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in GL(2, \mathbb{C}) \right\} \curvearrowright \mathbb{C} \cup \{\infty\}$$

Lemma. The subgroup of the Möbius group that preserves D,

$$\text{Möb}(D) = \{T \in \text{Möb} : T(D) = D\}$$

is also given by

$$\text{Möb}(D) = \left\{ z \mapsto e^{i\theta} \frac{z - a}{1 - \bar{a}z} : |a| < 1 \right\} = \left\{ \begin{pmatrix} a & b \\ b & a \end{pmatrix} \in \text{Möb} : |a|^2 - |b|^2 = 1 \right\}$$

Proof. Note that

$$\left| \frac{z - a}{1 - \bar{a}z} \right| = 1 \iff (z - a)(\bar{z} - \bar{a}) = (1 - \bar{a}z)(1 - a\bar{z})$$
$$\iff z\bar{z} - a\bar{z} - \bar{a}z + a\bar{a} = 1 - \bar{a}z - a\bar{z} + a\bar{a}z\bar{z}$$
$$\iff |z|^2\left(1 - |a|^2\right) = 1 - |a|^2$$
$$\iff |z| = 1$$

So these maps of the form

$$z \mapsto e^{i\theta} \frac{z - a}{1 - \bar{a}z}$$

do indeed preserve the unit circle, and $a \in D$ is mapped to $0 \in D$. Hence, it preserves the entire disc. $\square$

> **Lemma.** The Riemannian metric g_{hyp} is invariant under $\mathrm{M\ddot{o}b}(D)$. In other words, the Möbius group $\mathrm{M\ddot{o}b}(D)$ acts by isometries on D.

Proof. $\mathrm{M\ddot{o}b}(D)$ is generated by $z \mapsto e^{i\theta}$ and $z \mapsto \frac{z-a}{1-\bar{a}z}$. The rotations preserve g_{hyp}, since it depends only on $|z|$ and not z itself. For the second type of transformation, let $w = \frac{z-a}{1-\bar{a}z}$. Here,

$$\mathrm{d}w = \frac{\mathrm{d}z}{1 - \bar{a}z} + \frac{z-a}{(1-\bar{a}z)^2}\bar{a}\,\mathrm{d}z = \frac{\mathrm{d}z}{(1-\bar{a}z)^2}\left(1 - |a|^2\right)$$

Then,

$$\frac{|\mathrm{d}w|}{1 - |w|^2} = \frac{|\mathrm{d}z|}{|1-\bar{a}z|^2} \cdot \frac{1 - |a|^2}{1 - \left|\frac{z-a}{1-\bar{a}z}\right|^2} = \frac{|\mathrm{d}z|\left(1 - |a|^2\right)}{|1-\bar{a}z|^2 - |z-a|^2} = \frac{|\mathrm{d}z|}{1 - |z|^2}$$

Hence the hyperbolic metric, which is a function of this $\frac{|\mathrm{d}z|}{1-|z|^2}$, is also invariant under this change of variables. $\square$

> **Lemma.** On (D, g_{hyp}),
>   (i) every pair of points in (D, g_{hyp}) is joined by a unique geodesic up to reparametrisation;
>   (ii) the geodesics are diameters of the disc and circular arcs orthogonal to the boundary ∂D.
> The *whole* geodesics (ones that are defined on $\mathbb{R}$) are called *hyperbolic lines*.

Proof. Let $a \in \mathbb{R}_+ \cap D$ and γ a smooth path from the origin to a. Let $\gamma(t) = (u(t), v(t))$. Note that $\mathrm{Re}(\gamma)(t) = (u(t), 0)$ is also a smooth path from the origin to a. By definition of the hyperbolic metric,

$$\mathrm{length}(\gamma) = \int_0^1 \frac{2|\dot{\gamma}|}{1 - |\gamma|^2}\,\mathrm{d}t = \int_0^1 \frac{2\sqrt{\dot{u}^2 + \dot{v}^2}}{1 - u^2 - v^2}\,\mathrm{d}t \geq \int_0^1 \frac{2|\dot{u}|}{1 - u^2}\,\mathrm{d}t$$

where equality holds if and only if $\dot{v} \equiv 0$, and so $v \equiv 0$.

$$\mathrm{length}(\gamma) \geq \int_0^1 \frac{2\dot{u}}{1 - u^2}\,\mathrm{d}t$$

where equality holds in this expression if and only if u is monotonic. Hence, the arc of the diameter, parametrised monotonically, is a globally length-minimising path, and hence a geodesic. We can compute this integral to be

$$\mathrm{length}(\gamma) = 2\,\mathrm{artanh}\,a$$

Now, 0 and a in $\mathbb{R}_+ \cap D$ are joined by a unique geodesic, and $\mathrm{M\ddot{o}b}(D)$ acts transitively and by isometries, and can be used to send any two points $p, q \in D$ to $0, a \in \mathbb{R}_+ \cap D$. So every pair of points must be joined by a unique geodesic. Since Möbius maps send circles to circles, and they preserve angles and hence orthogonality to the boundary, we must have that all geodesics are diameters or circular arcs orthogonal to ∂D. $\square$

Corollary. If $p, q \in D$, then the distance between them is

$$d_{\text{hyp}}(p, q) = 2 \operatorname{artanh} \left| \frac{p - q}{1 - \bar{p}q} \right|$$

6.4 The hyperbolic upper half-plane

Definition. The *hyperbolic upper half-plane* $(\mathfrak{h}, g_{\text{hyp}})$ is the set

$$\mathfrak{h} = \{z \in \mathbb{C} : \operatorname{Im} z > 0\}$$

with the abstract Riemannian metric

$$\frac{\mathrm{d}x^2 + \mathrm{d}y^2}{y^2} = \frac{|\mathrm{d}z|^2}{(\operatorname{Im} z)^2}$$

Lemma. The hyperbolic disc (D^2, g_{hyp}) and the hyperbolic upper half-plane $(\mathfrak{h}, g_{\text{hyp}})$ are isometric.

Proof. There exist maps $T : \mathfrak{h} \to D$ and $\widetilde{T} : D \to \mathfrak{h}$ given by

$$T(w) = \frac{w - i}{w + i}; \quad \widetilde{T}(z) = i\left(\frac{1 - z}{1 + z}\right)$$

which are inverse diffeomorphisms. Here,

$$T'(w) = \frac{1}{w + i} - \frac{w - i}{(w + i)^2} = \frac{2i}{(w + i)^2}$$

Considering $T(w) = z \in D$,

$$\frac{|\mathrm{d}z|}{1 - |z|^2} = \frac{|\mathrm{d}(Tw)|}{1 - |Tw|^2} = \frac{|T'(w)||\mathrm{d}w|}{1 - |Tw|^2} = \frac{2|\mathrm{d}w|}{|w + i|^2 \left(1 - \left|\frac{w-i}{w+i}\right|^2\right)} = \frac{|\mathrm{d}w|}{2 \operatorname{Im} w}$$

Hence, under this coordinate change,

$$\frac{4|\mathrm{d}z|^2}{\left(1 - |z|^2\right)^2}$$

is the metric obtained under pullback by T from $\frac{\mathrm{d}w^2}{(\operatorname{Im} w)^2}$. □

Corollary. The hyperbolic upper half-plane is globally isometric to the hyperbolic disc, so every pair of points is joined by a unique geodesic, up to reparametrisation. The geodesics are arcs of circles orthogonal to the boundary, which are vertical straight lines and semicircles centred on a point in the real axis.

Proof. The isometry between $\mathfrak{h} \to D$ is given by a Möbius map. In particular, $\mathbb{R} \cup \{\infty\} \mapsto \partial D$, and Möbius maps preserve circles and orthogonality. $\qquad\square$

Remark. When we discussed surfaces in $\mathbb{R}^3$ with constant Gauss curvature, we saw that if a surface had constant Gauss curvature, its first fundamental form in geodesic normal coordinates was of the form $du^2 + \cosh^2 dv^2$, with a change of variables taking this form to $\frac{dv^2 + dw^2}{w^2}$. This is exactly the form of the Riemannian metric on the hyperbolic upper half-plane. Gauss' *theorema egregium* implies that Gauss curvature makes sense for an abstract Riemannian metric, since it only depends on geodesics and hence the first fundamental form. We can therefore define the Gauss curvature for an abstract Riemannian metric to agree with this definition for surfaces in $\mathbb{R}^3$. Under this definition, we can show that the hyperbolic upper half-plane has constant curvature -1, and hence so does the disc.

Suppose we wanted to find a metric $d : D \times D \to \mathbb{R}_{\geq 0}$ on D^2 with the properties that it is invariant under the Möbius group $\mathrm{M\ddot{o}b}(D)$, and that the real diameter is length-minimising. By Möbius invariance, the distance between any two points is completely determined by knowing the distance from the origin to some point on the positive real axis a, which we will denote $p(a) = d(0, a)$. If $\mathbb{R}_+ \cap D$ is length-minimising, distance should be additive, so if $0 \leq a \leq b \leq 1$ we should have $d(0, a) + d(a, b) = d(0, b)$ so $d(a, b) = p\left(\frac{b-a}{1-ab}\right) = p(b) - p(a)$. If we furthermore constrain p to be differentiable, and we differentiate the above expression with respect to b and set $b = a$, we find the differential equation

$$p'(a) = \frac{p'(0)}{1 - a^2}$$

Hence, $p(a)$ is some constant multiple of $\operatorname{artanh} a$, since $p'(0)$ can be chosen freely. So, up to rescaling the length metric associated to g_{hyp} on D is the unique metric with these properties. The scale is chosen for g_{hyp} to enforce that the curvature is -1 precisely.

6.5 Isometries of hyperbolic space

We now would like to understand the full isometry group of the disc (D, g_{hyp}) or $(\mathfrak{h}, g_{\mathrm{hyp}})$. We will show that this group is precisely $\mathrm{M\ddot{o}b}(D)$ together with reflections in hyperbolic lines, which are called *inversions*.

Definition. Let $\Gamma \subseteq \widehat{\mathbb{C}}$ be a circle or line. We say that points $z, z' \in \widehat{\mathbb{C}}$ are *inverse for* Γ if every circle through z orthogonal to Γ also passes through z'.

Lemma. Such inverse points exist and are unique.

Proof. Recall that Möbius maps preserve circles in $\widehat{\mathbb{C}}$ and preserve angles. In particular, if z, z' are inverse for Γ and $T \in \mathrm{M\ddot{o}b}$, then Tz and Tz' are inverse for the circle $T(\gamma)$. If $\Gamma = \mathbb{R} \cup \{\infty\}$, then $J(z) = \bar{z}$ gives inverse points; this map satisfies the definition above. Now, if $\Gamma \subseteq \widehat{\mathbb{C}}$ is any circle, there exists $T \in \mathrm{M\ddot{o}b}$ such that $T(\mathbb{R} \cup \{\infty\}) = \Gamma$. We can therefore define inversion in Γ to be $J_\Gamma = T \circ (z \mapsto \bar{z}) \circ T^{-1}$. $\qquad\square$

Definition. The map J_Γ in the proof above, sending z to the unique inverse point z' for z with respect to Γ, is called *inversion* in Γ.

This map fixes all points of Γ, and swaps points on the interior with points on the exterior.

Example. For Γ a straight line, this is simply reflection. For the unit circle, S^1, the map J_{S^1} maps $z \mapsto \frac{1}{\bar{z}}$ and $0 \mapsto \infty$.

Remark. The composition of two inversions is a Möbius map. Let C be the conjugation map $z \mapsto \bar{z}$, which is $J_{\mathbb{R} \cup \{\infty\}}$. If $\Gamma \subseteq \widehat{\mathbb{C}}$ is any circle, we have $J_\Gamma = T \circ C \circ T^{-1}$ where T is the Möbius transformation which maps $\mathbb{R} \cup \{\infty\}$ to Γ. If Γ_1, Γ_2 are circles, and T_1, T_2 are the transformations from $\mathbb{R} \cup \{\infty\}$ to Γ_1, Γ_2 respectively, then

$$J_{\Gamma_1} \circ J_{\Gamma_2} = \left(J_{\Gamma_1} \circ C \right) \circ \left(C \circ J_{\Gamma_1} \right)$$
$$= \left(C \circ J_{\Gamma_1} \right)^{-1} \circ \left(C \circ J_{\Gamma_1} \right)$$

We have $C \circ J_\Gamma = C \circ T \circ C \circ T^{-1}$, so it suffices to show $C \circ T \circ C \in$ Möb. If $T(z) = \frac{az+b}{cz+d}$, we have

$$(C \circ T \circ C)(z) = \frac{\bar{a}z + \bar{b}}{\bar{c}z + \bar{d}} \in \text{Möb}$$

Lemma. An orientation-preserving isometry of $(\mathbb{H}^2, g_{\text{hyp}})$ is an element of Möb$(\mathbb{H})$, where $\mathbb{H}$ is D or $\mathfrak{h}$. The full isometry group is generated by inversions in hyperbolic lines.

Proof. It suffices to prove this in either model, so we will use the disc model. Inversion in the geodesic $\mathbb{R} \cap D$ is conjugation, which preserves g_{hyp}. Note that Möb$(\mathbb{H})$ acts transitively by isometries on geodesics. Hence, if inversion in one geodesic preserves the metric, so does inversion in any geodesic.

Now, suppose α is some isometry of the hyperbolic disc D under the metric g_{hyp}. We have $\alpha(0) = a \in D$, and using $z \mapsto \frac{z-a}{1-\bar{a}z}$, so there exists $T \in$ Möb(D) such that $T \circ \alpha$ fixes the origin. There exists a rotation $R \in$ Möb(D) such that $R \circ T \circ \alpha$ maps $D \cap \mathbb{R}_+$ to itself. Composing with the conjugation map C if necessary, there exists an isometry A which is an inversion composed with a Möbius map such that $A \circ \alpha$ fixes $D \cap \mathbb{R}$ pointwise and fixes $D \cap i\mathbb{R}$ pointwise. The only such isometry is the identity, since every point in D is determined by its distance to these two lines. Hence, A is the inverse of α.

If α preserves orientation and fixes $\mathbb{R} \cap D$, then it necessarily fixes $i\mathbb{R} \cap D$ pointwise, so $\alpha = (R \circ T)^{-1} \in$ Möb. In general, α was constructed from Möb$(\mathbb{H})$ and inversions in hyperbolic lines. So to show that the isometry group is generated by inversions, it suffices to show that all Möbius maps are compositions of inversions. This is presented on the example sheets. $\square$

In the upper half-plane model of hyperbolic space,

$$\text{Möb}(\mathfrak{h}) = \mathbb{P}SL(2, \mathbb{R}) = \left\{ z \mapsto \frac{az+b}{cz+d} : \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in SL(2, \mathbb{R}) \right\}; \quad d_{\text{hyp}} = 2\,\text{artanh} \left| \frac{b-a}{b-\bar{a}} \right|$$

6.6 Hyperbolic triangles

Definition. Let α be an orientation-preserving isometry of $\mathbb{H}$, which is equivalently an element of Möb($\mathbb{H}$). Suppose α is not the identity map. We say that α is
  (i) *elliptic*, if α fixes some point $p \in \mathbb{H}$ (if $p = 0 \in D$, this behaves like a rotation);
  (ii) *parabolic*, if α fixes a unique point $p \in \partial\mathbb{H}$ (if $p = \infty \in \mathfrak{h}$, this behaves like a translation);
  (iii) *hyperbolic*, if α fixes two points on $\partial\mathbb{H}$, so it fixes the unique geodesic between these two points setwise, and so α must translate points across the geodesic; it is not an inversion in the geodesic because it is not the identity map.
All elements of Möb($\mathbb{H}$) are either elliptic, parabolic, or hyperbolic.

Definition. Let ℓ, ℓ' be hyperbolic lines. Then, we say
  (i) *parallel*, if they meet at the boundary $\partial\mathbb{H}$ but never inside $\mathbb{H}$;
  (ii) *ultra-parallel*, if they never meet in $\overline{\mathbb{H}}$;
  (iii) *intersecting*, if they meet in $\mathbb{H}$.
All pairs of hyperbolic lines are either parallel, ultra-parallel, or intersecting. A *hyperbolic triangle* is a region bound by three geodesics, no two of which are ultra-parallel. Vertices that lie 'at infinity' (on $\partial\mathbb{H}$) are called *ideal vertices*.

Note that the points in $\partial\mathbb{H}$ are not contained within the hyperbolic plane, so in particular the ideal vertices are not points in $\mathbb{H}$. We typically denote side lengths by A, B, C, and denote the angles opposite these sides by α, β, γ. The vertices at α, β, γ are denoted a, b, c. The hyperbolic metric is conformal, since $E = G$ and $F = 0$. Hence, we can use Euclidean angles in place of hyperbolic angles.

Proposition (hyperbolic cosine formula). For a hyperbolic triangle,

$$\cosh C = \cosh A \cosh B - \sinh A \sinh B \cos \gamma$$

Proof. To simplify, by an isometry we can let the vertex c at γ be placed at $0 \in D$, and the vertex b at β be placed at $\mathbb{R}_+ \cap D$. Hence, the sides A, B are straight Euclidean line segments in D, and the angle between them is γ. We have

$$d_{\text{hyp}}(0, a) = 2 \operatorname{artanh} a \implies a = \tanh \frac{A}{2}; \quad b = e^{i\gamma} \tanh \frac{B}{2}; \quad \left| \frac{b - a}{1 - \bar{a}b} \right| = \tanh \frac{C}{2}$$

Recall that

$$t = \tanh \frac{\lambda}{2} \implies \cosh \lambda = \frac{1 + t^2}{1 - t^2}; \quad \sinh \lambda = \frac{2t}{1 - t^2}$$

Hence,

$$\cosh A = \frac{1 + |a|^2}{1 - |a|^2}; \quad \cosh B = \frac{1 + |b|^2}{1 - |b|^2};$$

$$\cosh C = \frac{|1 - \bar{a}b|^2 + |b - a|^2}{|1 - \bar{a}b|^2 - |b - a|^2} = \frac{\left(1 + |s|^2\right)\left(1 + |b|^2\right) - 2\left(\bar{a}b + a\bar{b}\right)}{\left(1 - |a|^2\right)\left(1 - |b|^2\right)}$$

Note that $a \in \mathbb{R}$ and $b + \bar{b} = 2 \operatorname{Re} b = 2b \cos \gamma$, so

$$\cosh C = \cosh A \cosh B - \sinh A \sinh B \cos \gamma$$

as required. $\qquad\square$

Remark. If A, B, C are small, the standard approximations to the hyperbolic sine and cosine functions give

$$C^2 \approx A^2 + B^2 - 2AB \cos \gamma$$

which is the Euclidean cosine formula. Since a dilation of a surface in $\mathbb{R}^3$ rescales curvature, at small scales we can treat any abstract smooth surface with a Riemannian metric as flat.

Since $\cos \gamma \geq -1$, we have that

$$\cosh C \leq \cosh A \cosh B + \sinh A \sinh B = \cosh(A + B)$$

The hyperbolic cosine is increasing, so $C \leq A + B$. This is a more precise variant of the hyperbolic triangle inequality.

6.7 Area of triangles

Theorem. Let $T \subseteq \mathbb{H}^2$ be a hyperbolic triangle with internal angles α, β, γ defined as before. The area of T is

$$\text{area}_{\text{hyp}}(T) = \pi - \alpha - \beta - \gamma$$

Note that α, β, γ may be zero, so T may have ideal vertices, and the internal angle is zero for such vertices.

This is a version of the Gauss–Bonnet theorem for hyperbolic triangles.

Proof. The Möbius group $\text{Möb}(\mathbb{H}^2)$ acts transitively on triples of points in the boundary with the correct cycle order. In particular, there exists a single ideal triangle (with all vertices at infinity) up to isometry. Consider the ideal triangle in the hyperbolic upper half-plane with vertices $-1, +1, \infty$. Its area is

$$\text{area}_{\text{hyp}}(T) = \int_{-1}^{1} \int_{\sqrt{1-x^2}}^{\infty} \frac{1}{y^2} \, dy \, dx$$

since $\sqrt{EG - F^2} = \frac{1}{y^2}$. We can compute this explicitly as

$$\text{area}_{\text{hyp}}(T) = \int_{-1}^{1} \frac{dx}{\sqrt{1 - x^2}} = \pi$$

Now, let $A(\alpha)$ be the area of a triangle with angles $0, 0, \alpha$. We can see that $A(\alpha)$ is decreasing in α and continuous in α, by fixing two ideal vertices in the hyperbolic disc and translating the third vertex.
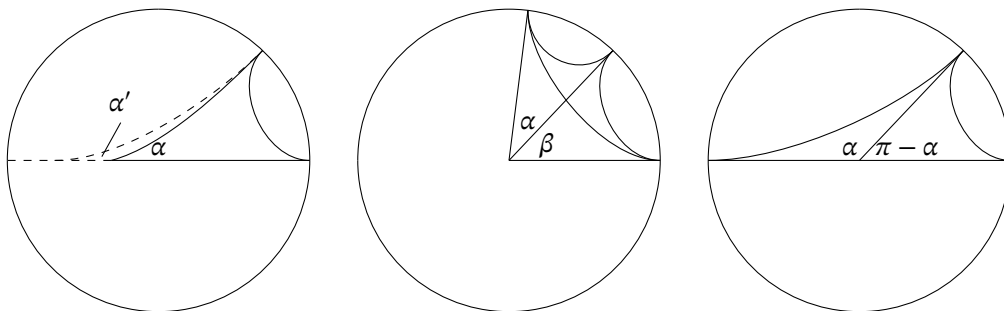
The first diagram shows that by moving the vertex α on the real line, the area must increase, since the triangle with angle $\alpha' < \alpha$ contains the triangle with angle α. From the second diagram, we see that $A(\alpha) + A(\beta) = A(\alpha+\beta) + \pi$ by comparing the different areas of triangles formed from hyperbolic lines in the diagram. By letting $F(\alpha) = \pi - A(\alpha)$, we have $F(\alpha) + F(\beta) = F(\alpha + \beta)$. Since F is continuous and increasing, we have that $F(\alpha) = \lambda\alpha$ for some fixed $\lambda > 0$. In particular, $A(\alpha) = \pi - \lambda\alpha$. Now, by considering the angles in the third diagram, we see that $A(\alpha) + A(\pi - \alpha) = \pi$. Hence, $\lambda = 1$, and so $A(\alpha) = \pi - \alpha$.

Finally, we consider the general case.
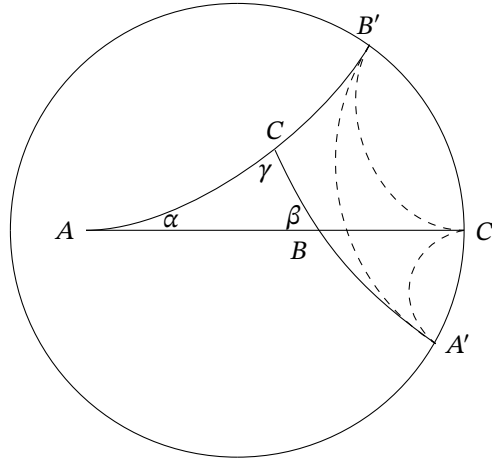
By writing ABC for $\text{area}_{\text{hyp}}(T)$ where T is the triangle with vertices A, B, C, we can see that

$$ABC + A'CB' + A'B'C' = \text{area of interior of diagram} = AB'C' + A'BC'$$

Equivalently,

$$ABC + \pi - (\pi - \gamma) + \pi = (\pi - \alpha) + (\pi - \beta) \implies ABC = \pi - \alpha - \gamma - \beta$$

as required. $\qquad\square$

Note that if G is a hyperbolic n-gon, so it is a region bound by n hyperbolic geodesics, it may be decomposed into a union of hyperbolic triangles. Since any two points in $\mathbb{H}^2$ are joined by a unique geodesic, the area of G is given by

$$\text{area}_{\text{hyp}}(G) = (n - 2)\pi - \sum_{i=1}^{n} \alpha_i$$

Lemma. If $g \geq 2$, then there exists a regular $4g$-gon in $\mathbb{H}^2$ with internal angle $\frac{2\pi}{4g} = \frac{\pi}{2g}$.

Proof. Consider an ideal $4g$-gon, whose vertices all lie at infinity, in the disc model of hyperbolic space. The ideal vertices can be placed at the $4g$-th roots of unity, such that this polygon is invariant under a rotational symmetry. By sliding each vertex radially inwards in $\mathbb{R}^2$, we obtain a continuous family of regular $4g$-gons, with areas which vary monotonically from $(4g - 2)\pi$ to zero. The internal angle of the polygon therefore varies continuously from zero to $\beta_{\min}$ such that $(4g - 2)\pi = 4g\beta_{\min}$. It therefore suffices to check that $\frac{\pi}{2g}$ lies in this interval $(0, \beta_{\min})$. $\qquad\square$

6.8 Surfaces of constant negative curvature

> **Theorem.** For each $g \geq 2$, there exists an abstract Riemannian metric on the compact surface of genus g with curvature $\kappa \equiv -1$ and locally isometric to $\mathbb{H}^2$.

Recall the the Euler characteristic of a surface of genus g is exactly $2 - 2g$. Note, if $g = 0$ we can construct a Riemannian metric with $\kappa \equiv +1$ since this is the sphere, and if $g = 1$ we can have $\kappa \equiv 0$ since this is the torus as a quotient $\mathbb{R}^2/\mathbb{Z}^2$. We will outline two proofs.

Proof. Recall that we can construct the torus and double torus by
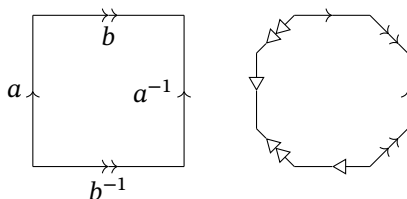
Analogously, a $4g$-gon with side labels $a_1 b_1 a_1^{-1} b_1^{-1} a_2 b_2 a_2^{-1} b_2^{-1} \ldots$ gives a surface of genus g.

We say that a *flag* comprises an oriented hyperbolic line, a point on that line, and a choice of side to that line. Given two such flags, there exists a hyperbolic isometry between them. So $\text{Möb}(\mathbb{H})$ acts transitively on flags. In particular, we can swap the side of a flag using an inversion.

Consider a regular hyperbolic $4g$-gon with internal angle $\frac{\pi}{2g}$. We label this polygon with side labels as above to give a genus g surface. For each paired set of two edges, there exists a hyperbolic isometry taking one to the other, respecting orientations and, and taking the side corresponding to the inside of the polygon to the side corresponding to the outside of the polygon. This is possible since $\text{Möb}(\mathbb{H})$ acts transitively on flags.

We can now define an atlas for Σ_g as follows.

- If p is in the interior of the polygon P, consider a small disc contained in the interior of the polygon. Then, include this disc into the hyperbolic disc D.

- If p is contained in an edge, let $\hat{p}$ be the corresponding point on the paired edge. We have an isometry γ from edge e_1 to edge e_2, exchanging sides, and mapping p to $\hat{p}$. We can use this to define the chart. Using γ, we can combine U, the intersection of P with an open neighbourhood of p, and $\widetilde{U}$, the intersection of P with an open neighbourhood of $\hat{p}$, such that the chart is an inclusion on U and is γ on $\widetilde{U}$. These agree on $U \cap \widetilde{U}$.

- All $4g$ vertices are identified to one point of Σ, and we need a chart at this point. Using a hyperbolic isometry, let one vertex v of P be at the origin in D, such that an edge e containing v is mapped to a subset of the real line. Since the polygon P has internal angle $\frac{\pi}{2g}$, the angle between $\mathbb{R}$ and the adjacent edge is $\frac{\pi}{2g}$. The fact that the internal angles sum to 2π means that we can construct hyperbolic isometries for each vertex that join them exactly, giving an open neighbourhood of zero in D in the shape of a disc. The chart is defined at $[v] \in \Sigma_g$ by this identification.

All charts are obtained from inclusion or an inclusion composed with a hyperbolic isometry, therefore the transition maps are hyperbolic isometries. In particular, hyperbolic isometries are smooth, and preserve the locally defined hyperbolic metric. $\qquad\square$

Remark. The torus can be given by $\mathbb{R}^2/\mathbb{Z}^2$. This characterisation was useful when describing the flat metric, precisely because its charts are easy to define. For Σ_g, we chose $2g$ hyperbolic isometries which paired sides. Hence, there is a group $\Gamma \leq \text{Möb}(\mathbb{H})$, generated by these isometries. In Part II Algebraic Topology, the surface Σ_g will be constructed by $\mathbb{H}/\Gamma$.

> **Lemma.** For each $\ell_\alpha, \ell_\beta, \ell_\gamma > 0$, there exists a right-angled hyperbolic hexagon with side lengths $\ell_\alpha, a, \ell_\beta, b, \ell_\gamma, c$ for some a, b, c.

Proof. Given $t > 0$, there exists a pair of ultra-parallel hyperbolic lines a distance t apart. We show on the fourth example sheet that each pair of ultra-parallel hyperbolic lines has a unique common perpendicular geodesic. Given lengths ℓ_α, ℓ_β, construct new perpendicular geodesics orthogonal to the originals, having moved lengths ℓ_α, ℓ_β from the common perpendicular (in the same direction). If t is made large, the new geodesics $\sigma, \tilde{\sigma}$ can be made ultraparallel. Hence, by making t smaller, there exists a threshold t_0 by continuity such that the new geodesics are parallel. Now, for $t \in (t_0, \infty)$, the two new geodesics are ultra-parallel. So $\sigma, \tilde{\sigma}$ have a unique common perpendicular geodesic. As t increases above t_0, the length of this line increases monotonically from zero to infinity. So there exists a value of $t > t_0$ such that the new common perpendicular has length ℓ_γ.
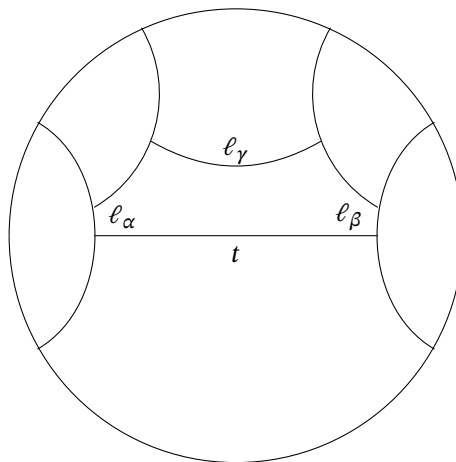
This is exactly the right-angled hyperbolic hexagon as required. $\qquad\square$

> **Definition.** A *pair of pants* is a topological space homeomorphic to the complement of three open discs in S^2.

Note that this space has a boundary. Consider two right-angled hyperbolic hexagons with side lengths $\ell_\alpha, \ell_\beta, \ell_\gamma$ arranged as above. The original configuration of two ultra-parallel geodesics of a distance t apart is unique up to isometry. So the side lengths have a correspondence, and the hexagon with side lengths $\ell_\alpha, \ell_\beta, \ell_\gamma$ is unique up to isometry. Suppose that we glue together the corresponding

unknown sides $t_{\alpha\beta}, t_{\beta\gamma}, t_{\gamma\alpha}$ with the same side identifications. Locally near ℓ_α, for instance, we arrive at a closed circle of length $2\ell_\alpha$, extended into a cylindrical shape with two seams $t_{\alpha\beta}, t_{\gamma\alpha}$. Since the hexagons were right-angled, we have constructed a hyperbolic pair of pants. The boundary circles are geodesics in the sense that, for any point on such a circle, the local neighbourhood is a point on a geodesic on a polygon in $\mathbb{H}$.

We will now construct Σ_g using a more flexible approach.

Proof. If P_1, P_2 are two hyperbolic 'surfaces' with geodesic boundaries, and if $\gamma_1 \subset P_1$ and $\gamma_2 \subset P_2$ are boundary circles of the same length (in the hyperbolic metric), we can glue P_1 and P_2 together along this common-length circle. P_1 and P_2 may be glued by any isometry of γ_1, γ_2. The result $P_1 \cup_{\gamma_1 \sim \gamma_2} P_2$ has a hyperbolic metric. For any point $p \in P_i$ not on the boundary γ_i, it already has a suitable open neighbourhood since P_i is hyperbolic. For any point $p \in \gamma_1 \sim \gamma_2$, we have a chart to a small disc in $\mathbb{H}$ using the fact that the boundary circles are geodesics. These charts are constructed analogously to the charts for points on edges of hyperbolic polygons under appropriate side identifications as seen above. Any compact surface of genus $g \geq 2$ can be built from glued pairs of pants, not necessarily uniquely.

Under this construction, we have many choices. For example, the lengths of circles in the original hyperbolic hexagons are now arbitrary. Also, the choice of 'pants decomposition' of a given surface is not unique, and the different possibilities are topologically different. $\qquad\square$

6.9 Gauss–Bonnet theorem

Recall that in a spherical triangle with internal angles α, β, γ, we have seen in the example sheets that this has area $\alpha + \beta + \gamma - \pi$, and that a hyperbolic triangle with the same internal angles has area $\pi - \alpha - \beta - \gamma$. We have seen the convex Gauss–Bonnet theorem, which states

$$\int_\Sigma \kappa \, \mathrm{d}A = 4\pi$$

where Σ bounds a convex region in $\mathbb{R}^3$ and $\kappa_\Sigma > 0$. These are special cases of a pair of theorems as shown below.

Theorem (local Gauss–Bonnet theorem)**.** Let Σ be an abstract smooth surface with abstract Riemannian metric g. Let R be an n-sided *geodesic polygon* on Σ, which is a smooth disc with boundary decomposed into n geodesic arcs. Then

$$\int_{R \subseteq \Sigma} \kappa_\Sigma \, \mathrm{d}A = \sum_{i=1}^{n} \alpha_i - (n-2)\pi$$

where the α_i are the internal angles of the polygon.

It is important that γ_i be geodesics that cut out a disc; R must be homeomorphic to $\mathbb{R}^2$, and it cannot (for example) contain any holes.

Theorem (global Gauss–Bonnet theorem)**.** Let Σ be a compact smooth surface with abstract

Riemannian metric g. Then

$$\int_\Sigma \kappa_\Sigma \, \mathrm{d}A = 2\pi\chi(\Sigma)$$

Remark. Gauss curvature can be defined using only the first fundamental form, or equivalently an abstract Riemannian metric.

For hyperbolic surfaces, we can construct Σ_g from a $4g$-gon with internal angles $\frac{\pi}{2g}$ in such a way that the total area of Σ is exactly the area of the polygon, so

$$\int_\Sigma 1 \, \mathrm{d}A = \mathrm{area(polygon)} = (4g-2)\pi - \sum_1^{4g} \frac{\pi}{2g} = (4g-4)\pi$$

Since $\kappa \equiv -1$ and $\chi(\Sigma_g) = 2 - 2g$, this agrees with the Gauss–Bonnet theorem.

A right-angled hyperbolic hexagon has area

$$4\pi - \sum_1^6 \frac{\pi}{2} = \pi$$

Each pair of pants was constructed from two such polygons, and to construct a genus g surface we required $2g - 2$ pairs of pants. So the total area is $4g - 4\pi$, which agrees with the theorem.

The Gauss–Bonnet theorem also shows that the Euler characteristic does not depend on the choice of triangulation of Σ.

Suppose Σ is a flat surface and γ is a closed geodesic, so $\gamma : \mathbb{R} \to \Sigma$ and is periodic with some period T. Then γ cannot bound a smooth disc in Σ. Conversely, on S^2, the great circle is a closed geodesic, and bounds a hemisphere. For instance, for the flat torus $\mathbb{R}^2/\mathbb{Z}^2$, if γ is a closed curve on this torus bounding a closed disc R it is not a geodesic. Indeed, if we formally add two vertices to such a geodesic, we find a geodesic 2-gon with two internal angles π, but by the Gauss–Bonnet theorem we expect

$$0 = \int_R \kappa_\Sigma \, \mathrm{d}A = \sum_1^2 \alpha_i - (n-2)\pi = 2\pi$$

We can in fact deduce the global Gauss–Bonnet theorem from the local Gauss–Bonnet theorem, utilising the following lemma.

Lemma. A compact smooth surface admits subdivisions into geodesic polygons.

The proof of this lemma considers the exponential map, discussed in Part II. Given such a subdivision on Σ, we can find

$$\sum_{\text{polygons } P} \int_P \kappa_\Sigma \, \mathrm{d}A = \int_\Sigma \kappa_\Sigma \, \mathrm{d}A$$

By the local Gauss–Bonnet theorem, the left hand side is equal to

$$\sum_n \sum_{n\text{-gons } P} \left(\sum_{i=1}^n \alpha_i(P) - (n-2)\pi \right)$$

Since the angles at each point add to 2π, and each n-gon contains two edges which each separate two polygons, this is equal to $2\pi V + 2\pi F - 2\pi E = 2\pi\chi(\Sigma)$ as required.

6.10 Green's theorem (non-examinable)

The local Gauss–Bonnet theorem is very closely related to Green's theorem in $\mathbb{R}^2$. This discussion is non-examinable.

Theorem. Let $R \subseteq \mathbb{R}^2$ be a region bound by a piecewise smooth curve γ, and P, Q be smooth real-valued functions defined on an open set $V \supset R$. Then

$$\int_\gamma P \, du + Q \, dv = \int_R (Q_u - P_v) \, du \, dv$$

We will consider a geodesic polygon on Σ which lies in the domain of some local parametrisation defined on $V \subseteq \mathbb{R}^2$. Consider an orthonormal basis for $\mathbb{R}^2$ which varies from point to point, defined by $e = \sigma_u, f = \sigma_v/\sqrt{G}$ where we use geodesic normal coordinates u, v to give $E = 1, F = 0$. Then $T_p\Sigma = \text{span}(e, f)$ if $p \in \text{Im } \sigma$. We parametrise γ by arc length and consider

$$I = \int_\gamma \langle e, \dot{f} \rangle \, dt$$

We will compute this in two ways. Note that

$$\dot{f} = f_u \dot{u} + f_v \dot{v}$$

Let $P = \langle e, f_u \rangle$ and $Q = \langle e, f_v \rangle$. Then

$$Q_u - P_v = \langle e_u, f_v \rangle - \langle f_v, e_u \rangle + \langle e, f_{uv} \rangle - \langle e, f_{uv} \rangle = \langle e_u, f_v \rangle - \langle f_u, e_v \rangle$$

which we can show to be equal to $-\left(\sqrt{G}\right)_{uu} = \kappa\sqrt{G}$. But $\sqrt{G}$ is the area element $\sqrt{EG - F^2}$, so

$$\int_R (Q_u - P_v) \, du \, dv = \int_R \kappa_\Sigma \, dA$$

Let $\theta(t)$ be the angle between $\dot{\gamma}(t)$ and $e(t)$, which is a function of t in the domain of γ. More precisely,

$$\dot{\gamma} = e \cos \theta(t) + f \sin \theta(t)$$

Thus

$$\ddot{\gamma} = \dot{e} \cos \theta + \dot{f} \sin \theta + \eta \dot{\theta}; \quad \eta = -e \sin \theta + f \cos \theta$$

γ is a piecewise geodesic, so if $\Sigma \subseteq \mathbb{R}^3$ was smooth, $\ddot{\gamma}$ is orthogonal to $T_p\Sigma = \text{span } e, f$. But $\eta \in \langle e, f \rangle$, so $\ddot{\gamma}$ is orthogonal to η. By expanding,

$$\langle \dot{e} \cos \theta + \dot{f} \sin \theta + \eta \dot{\theta}, -e \sin \theta + f \cos \theta \rangle = 0$$

Since e, f are orthogonal unit vectors, we have $\langle e, \dot{e} \rangle = 0 = \langle f, \dot{f} \rangle$ and $\langle e, \dot{f} \rangle = 0 = \langle \dot{e}, f \rangle$, so we can expand to find

$$\langle \ddot{\gamma}, \eta \rangle = 0 \implies \dot{\theta} = \langle e, \dot{f} \rangle$$

Thus,

$$I = \int_\gamma \langle e, \dot{f} \rangle \, dt = \int_\gamma \dot{\theta}(t) \, dt = 2\pi - \sum(\text{external angles of } R)$$

since γ is composed of straight lines. Since external angles and internal angles sum to π, this is exactly the local Gauss–Bonnet theorem. Green's theorem suggests the study of non-geodesic polygons.

6.11 Alternate flat toruses

We have constructed a flat metric on the torus, viewed as $\mathbb{R}^2/\mathbb{Z}^2$, or as $[0,1]^2/\sim$ for a suitably defined equivalence relation. Importantly, opposite sides of the square $[0,1]^2$ were identified by translation, which allowed us to find a smooth atlas where transition maps preserve the usual Euclidean metric on $\mathbb{R}^2$. This construction is valid for any parallelogram; any such shape $Q \subseteq \mathbb{R}^2$ defines a flat metric g_Q on T^2. If one vertex is set to zero in $\mathbb{R}^2$ and the edges of this vertex are labelled by their endpoints v_1, v_2, then $(T^2, g_Q) = \mathbb{R}^2/\mathbb{Z}v_1 \oplus \mathbb{Z}v_2$ where $\mathbb{Z}v_1 \oplus \mathbb{Z}v_2$ is a viewed as a subgroup of the group $\mathbb{R}^2$ of translations.

The area with respect to g_Q of T^2 is the Euclidean area of the parallelogram Q. In particular, if two parallelograms have different areas, the two metrics cannot be globally isometric. However, this is not the only restriction for global isometries.

> **Lemma.** Consider the torus based on $Q = [0,1]^2$ and the torus based on $\hat{Q} = [0,10] \times \left[0, \frac{1}{10}\right]$. The metrics $g_Q, g_{\hat{Q}}$ are not isometric, but both have unit total area.

Proof. Recall that geodesics in a flat torus correspond to straight lines in $\mathbb{R}^2$. By Picard's theorem, there exists a unique geodesic from a given point p for each direction in $T_p\Sigma$. We can therefore see that all geodesics through p are the images of straight lines in $\mathbb{R}^2$.

Recall that a *closed geodesic* is defined on $\mathbb{R}$ and is periodic. We can see that geodesics in $\mathbb{R}^2$ through $\hat{p} \in q^{-1}(p)$ define a closed geodesic if and only if they pass through another lift $\hat{p}' \in q^{-1}(p)$ of p; that is, the line has rational gradient. The shortest closed geodesic on the surface in metric Q is of unit length, but the shortest closed geodesic with metric $\hat{Q}$ is $\frac{1}{10}$. So the surfaces are not globally isometric. $\square$

We would like to understand all possible flat metrics on the torus T^2, up to global dilation and Euclidean isometries of Q, which lead to essentially the same geometry on the quotient torus. Given any parallelogram, we can set one vertex at zero and another at $(1,0) = 1 \in \mathbb{R}^2$ by performing dilation and a Euclidean isometry, and then the third lies at τ and the fourth at $1 + \tau$, where τ has positive y-coordinate. This provides a metric on the torus, and now the only degree of freedom is τ. Hence, this defines a map from the upper half-plane to the set of flat metrics on T^2 up to dilation.

We can pull back metrics by diffeomorphisms. Metrics allow us to measure lengths of curves by integrating lengths of tangent vectors, so a metric can be viewed as an inner product on the tangent space at each point. If $f : \Sigma \to \Sigma'$ and $p \in \Sigma$, then for two small curves γ_1, γ_2 through p, the pullback metric $f^\star g$ was defined such that

$$\langle \dot{\gamma}_1, \dot{\gamma}_2 \rangle_{p, f^\star g} = \langle f \circ \dot{\gamma}_1, f \circ \dot{\gamma}_2 \rangle_{f(p), g}$$

$SL(2, \mathbb{Z})$ acts on $\mathbb{R}^2$ preserving $\mathbb{Z}^2$, so it acts on $\mathbb{R}^2/\mathbb{Z}^2 = T^2$.

> **Lemma.** $SL(2, \mathbb{Z})$ acts by diffeomorphisms on T^2.

Proof. Clearly $A \in SL(2, \mathbb{Z})$ acts smoothly (indeed, linearly) on $\mathbb{R}^2$, and the charts for the smooth atlas are such that A then acts smoothly with respect to these. $\square$

Also, $SL(2,\mathbb{Z}) \subseteq SL(2,\mathbb{R})$ acts on the upper half-plane by Möbius maps.

> **Theorem.** The map from the upper half-plane $\mathfrak{h}$ to the set of flat metrics on T^2 modulo dilation induces a map from $\mathfrak{h}/_{SL(2,\mathbb{Z})}$ to the set of flat metrics on T^2 modulo dilation and diffeomorphism. This resulting map is a bijection. We say that $\mathfrak{h}/_{SL(2,\mathbb{Z})}$ is the *moduli space* of flat metrics on T^2.

In the above theorem, 'diffeomorphism' is taken to mean 'orientation-preserving diffeomorphism'.

Remark. The left-hand side $\mathfrak{h}/_{SL(2,\mathbb{Z})}$ is an object of hyperbolic geometry, yet the right-hand side is entirely concerned with flat metrics.

Similar results can be shown for surfaces of higher genus. The moduli space of hyperbolic metrics on Σ_g where $g \geq 2$ is perhaps the most studied space in all of geometry.

6.12   Further courses

There are four Part II courses that extend this course.

(i) Algebraic Topology. Spaces are studied through algebraic invariants, such as the Euler characteristic, and covering maps of surfaces like $S^2 \to \mathbb{RP}^2$ or $\mathbb{R}^2 \to T^2$.

(ii) Differential Geometry. While in IB Geometry the Gauss curvature $\kappa = \det(DN)$ is discussed, the trace $\mathrm{tr}(DN)$ is the *mean curvature*, discussed heavily in this course.

(iii) Riemann Surfaces. This course studies the fact that if $f : \mathbb{C} \to \mathbb{C}$ is holomorphic (or, indeed, entire) and $w \in \mathbb{C}$, then $f(z + w)$ is holomorphic, and if $f : D \to D$ is holomorphic and $A \in \mathrm{Möb}(D)$, then $f \circ A$ is holomorphic.

(iv) General Relativity. This is the theory of light as geodesics.