DIFFERENTIAL GEOMETRY, D COURSE

Gabriel P. Paternain

Department of Pure Mathematics and Mathematical Statistics, University of Cambridge, Cambridge CB3 0WB, England

E-mail address: `g.p.paternain@dpmms.cam.ac.uk`

Contents

CHAPTER 1

Smooth manifolds and smooth maps

1.1. Definitions

Definition 1.1. Let U be an open set of $\mathbb{R}^n$. A map $f : U \to \mathbb{R}^m$ is called *smooth* if it has continuous partial derivatives of all orders. A map $f : X \to \mathbb{R}^m$ defined on an arbitrary subset X of $\mathbb{R}^n$ is called *smooth* if for each $x \in X$ there is an open set $U \subset \mathbb{R}^n$ containing x and a smooth map $F : U \to \mathbb{R}^m$ such that F equals f on $U \cap X$.

We will use the shorthand term *local* when we wish to refer to behaviour only in a neighbourhood of a point. Smoothness is a local property. (The term *global* refers to the whole space X.)

Definition 1.2. A smooth map $f : X \to Y$ between subsets of Euclidean space is a *diffeomorphism* if it is a bijection, and if the inverse $f^{-1} : Y \to X$ is also smooth. X and Y are *diffeomorphic* if such an f exists.

Differential Topology is about properties of a set $X \subset \mathbb{R}^n$ which are invariant under diffeomorphisms.

Arbitrary sets of $\mathbb{R}^n$ may be too wild. We would like to have a class of sets on which can do locally the same as in Euclidean space.

Definition 1.3. Let X be a subset of $\mathbb{R}^N$. We say that X is a *k-dimensional manifold* if each point possesses a neighbourhood V in X which is diffeomorphic to an open set of $\mathbb{R}^k$. A diffeomorphism $\phi : U \to V$, where U is an open set of $\mathbb{R}^k$, is called a *parametrization* of the neighbourhood V. The inverse diffeomorphism $\phi^{-1} : V \to U$ is called a *coordinate system* or a *chart* on V.

If we write $\phi^{-1} = (x_1, \ldots, x_k)$ the k smooth functions $x_1, \ldots, x_k$ on V are called *coordinate functions*. The dimension k of X is written as dim X.

Example 1.4. The unit sphere S^2 given by all $(x, y, z) \in \mathbb{R}^3$ with $x^2+y^2+z^2 = 1$ is a smooth manifold of dimension 2. The diffeomorphism

$$(x, y) \mapsto (x, y, \sqrt{1 - x^2 - y^2}),$$

for $x^2+y^2 < 1$, parametrizes the region $z > 0$ of S^2. By interchanging the roles of x, y and z and changing the sign of the variables, we obtain similar parametrizations of the regions $x > 0$, $y > 0$, $x < 0$, $y < 0$ and $z < 0$. Since these cover S^2, it follows that S^2 is a smooth manifold. Similarly $S^n := \{x \in \mathbb{R}^{n+1} : \ |x| = 1\}$ is a smooth manifold of dimension n.

Exercise 1.5. If X and Y are manifolds, so is $X \times Y$, and dim $X \times Y =$ dim $X +$ dim Y.

DEFINITION 1.6. If X and Z are both manifolds in $\mathbb{R}^N$ and $Z \subset X$, then Z is a *submanifold* of X. In particular, X itself is a submanifold of $\mathbb{R}^N$. Any open set of X is a submanifold of X. The *codimension* of Z in X is dim $X -$ dim Z.

1.2. Tangent spaces and derivatives

For any open set $U \subset \mathbb{R}^n$ and $x \in U$, *the tangent space* to U at x (denoted T_xU) is defined to be $\mathbb{R}^n$. Recall from Analysis II that for any smooth map $f : U \to \mathbb{R}^m$, the derivative of f at $x \in U$ is the linear map

$$df_x : \mathbb{R}^n \to \mathbb{R}^m$$

given by

$$df_x(h) = \lim_{t \to 0} \frac{f(x+th) - f(x)}{t}$$

for $h \in \mathbb{R}^n$.

Now let us define *the tangent space* T_xX of an arbitrary smooth manifold $X \subset \mathbb{R}^N$. Choose a parametrization $\phi : U \to X$ around x where U is an open set of $\mathbb{R}^k$. Without loss of generality assume that $\phi(0) = x$. Think of ϕ as a map from U to $\mathbb{R}^N$, so $d\phi_0 : \mathbb{R}^k \to \mathbb{R}^N$ is defined. Set

$$T_xX := d\phi_0(\mathbb{R}^k).$$

LEMMA 1.7. *T_xX does not depend on ϕ and* dim $T_xX = k$.

PROOF. Suppose $\psi : V \to X$ is another choice, with $\psi(0) = x$. By shrinking both U and V, we may assume that $\phi(U) = \psi(V)$. Then the map $h = \psi^{-1} \circ \phi : U \to V$ is a diffeomorphism. Now we write $\phi = \psi \circ h$ and we differentiate. Using the chain rule we have: $d\phi_0 = d\psi_0 \circ dh_0$. Since dh_0 is an invertible linear map, it follows at once that $d\phi_0$ and $d\psi_0$ have the same image.

Since $\phi^{-1} : \phi(U) \to U$ is a smooth map, we can choose an open set W in $\mathbb{R}^N$ containing x and a smooth map $\Phi : W \to \mathbb{R}^k$ that extends ϕ^{-1}. Then $\Phi \circ \phi$ is the identity map of U, so the chain rule implies that the composition of linear maps

$$\mathbb{R}^k \xrightarrow{d\phi_0} T_xX \xrightarrow{d\Phi_x} \mathbb{R}^k$$

is the identity map of $\mathbb{R}^k$. Thus $d\phi_0 : \mathbb{R}^k \to T_xX$ is an isomorphism and dim $T_xX = k$.

□

We can now define the derivative of a smooth map $f : X \to Y$ between arbitrary manifolds. Let x be a point in X and set $y = f(x)$. The derivative must be a linear map $df_x : T_xX \to T_yY$ which gives us back the usual derivative when X and Y are open sets in Euclidean space and it must also satisfy the chain rule.

Keeping this in mind, let $\phi : U \to X$ and $\psi : V \to Y$ be parametrizations around x and y respectively, where U is an open set of $\mathbb{R}^k$ and V is an open set of $\mathbb{R}^l$. As before suppose $\phi(0) = x$ and $\psi(0) = y$. If U is small enough we have the diagram:

$$\begin{array}{ccc} X & \xrightarrow{f} & Y \\ \phi \uparrow & & \uparrow \psi \\ U & \xrightarrow{h = \psi^{-1} \circ f \circ \phi} & V \end{array}$$

Hence our definition of df_x must be so that the following diagram commutes:

$$\begin{array}{ccc} T_xX & \xrightarrow{df_x} & T_yY \\ {\scriptstyle d\phi_0}\uparrow & & \uparrow{\scriptstyle d\psi_0} \\ \mathbb{R}^k & \xrightarrow{dh_0} & \mathbb{R}^l \end{array}$$

Since $d\phi_0$ is an isomorphism, we must have

$$df_x := d\psi_0 \circ dh_0 \circ d\phi_0^{-1}.$$

We must verify that this definition does not depend on the choices of ϕ and ψ.

Exercise 1.8. Prove that this is indeed the case.

Chain Rule. If $X \xrightarrow{f} Y \xrightarrow{g} Z$ are smooth maps of manifolds, then

$$d(g \circ f)_x = dg_{f(x)} \circ df_x.$$

1.2.1. The inverse function theorem. Let $f : X \to Y$ be a smooth map between manifolds. We say that f is a *local diffeomorphism* at x if f maps a neighbourhood of x diffeomorphically onto a neighbourhood of $f(x)$.

Theorem 1.9 (The inverse function theorem). *Suppose that $f : X \to Y$ is a smooth map whose derivative df_x at the point x is an isomorphism. Then f is a local diffeomorphism at x.*

Exercise 1.10. Prove the theorem assuming the inverse function theorem for smooth functions between open sets of Euclidean space, as you have seen it in Analysis II.

1.3. Regular values and Sard's theorem

Let $f : X \to Y$ be a smooth map between manifolds. Let C be the set of all points $x \in X$ such that $df_x : T_xX \to T_{f(x)}Y$ is not surjective.

Definition 1.11. A point in C will be called a *critical point*. A point in $f(C)$ will be called a *critical value*. A point in the complement of $f(C)$ will be called a *regular value*.

Remark 1.12. Note that if $\dim X < \dim Y$, then $C = X$ and the preimage of a regular value is the empty set.

Theorem 1.13 (Preimage theorem). *Let y be a regular value of $f : X \to Y$ with* $\dim X \geq \dim Y$. *Then the set $f^{-1}(y)$ is a submanifold of X with* $\dim f^{-1}(y) = \dim X - \dim Y$.

Proof. Let $x \in f^{-1}(y)$. Since y is a regular value, the derivative df_x maps T_xX onto T_yY. The kernel of df_x is a subspace K of T_xX of dimension $p := \dim X - \dim Y$. Suppose $X \subset \mathbb{R}^N$ and let $T : \mathbb{R}^N \to \mathbb{R}^p$ be any linear map such that $\mathrm{Ker}(T) \cap K = \{0\}$. Consider the map $F : X \to Y \times \mathbb{R}^p$ given by

$$F(z) = (f(z), T(z)).$$

The derivative of F is given by

$$dF_x(v) = (df_x(v), T(v))$$

which is clearly nonsingular by our choice of T. By the inverse function theorem, F is a local diffeomorphism at x, i.e. F maps some neighbourhood U of x diffeomorphically onto a neighbourhood V of $(y, T(x))$. Hence F maps $f^{-1}(y) \cap U$ diffeomorphically onto $(\{y\} \times \mathbb{R}^p) \cap V$ which proves that $f^{-1}(y)$ is a manifold with $\dim f^{-1}(y) = p$.

□

The case $\dim X = \dim Y$ is particularly important. The theorem says that if y is a regular value of f, then $f^{-1}(y)$ is a 0-dimensional manifold, i.e. collection of points. If X is compact, this collection must be finite (why?) so we obtain:

Corollary 1.14. *Let $f : X \to Y$ be a smooth map between manifolds of the same dimension. If X is compact and y is a regular value, $f^{-1}(y)$ consists of a finite set of points.*

We can actually say a bit more:

Theorem 1.15 (Stack of records theorem). *Let $f : X \to Y$ be a smooth map between manifolds of the same dimension with X compact. Let y be a regular value of f and write $f^{-1}(y) = \{x_1, \dots, x_k\}$. Then there exists a neighbourhood U of y in Y such that $f^{-1}(U)$ is a disjoint union $V_1 \cup \cdots \cup V_k$, where V_i is an open neighbourhood of x_i and f maps each V_i diffeomorphically onto U.*

Proof. By the inverse function theorem we can pick disjoint neighbourhoods W_i of x_i such that f maps W_i diffeomorphically onto a neighbourhood of y. Observe that $f(X - \cup_i W_i)$ is a compact set which does not contain y. Now take

$$U := \cap_i f(W_i) - f(X - \cup_i W_i).$$

□

If we let $\# f^{-1}(y)$ be the cardinality of $f^{-1}(y)$, the theorem implies that the function $y \mapsto \# f^{-1}(y)$ is locally constant as y ranges over regular values of f.

The Preimage theorem gives a particularly nice way of producing manifolds. For example, if we consider the map $f : \mathbb{R}^{n+1} \to \mathbb{R}$ given by $f(x) = |x|^2$ we can check easily that 1 is a regular value of f (do it!). Hence $S^n = f^{-1}(1)$ is a smooth manifold of dimension n. By switching our mental wavelength a little we can produce quite interesting examples as follows:

Example 1.16 (Orthogonal group). Let $O(n)$ be the group of orthogonal matrices of size $n \times n$, i.e., a matrix $A \in O(n)$ if and only if $A\,A^t = I$, where I is the identity matrix. The space of all $n \times n$

matrices $M(n)$ is just $\mathbb{R}^{n^2}$, so we can think of $O(n)$ as living inside $\mathbb{R}^{n^2}$. We will show that $O(n)$ is a manifold of dimension $\frac{n(n-1)}{2}$.

Let $S(n) \subset M(n)$ be the space of all symmetric matrices. Since it is a vector space, is clearly a submanifold of $M(n)$ and it has dimension $\frac{n(n+1)}{2}$.

Let $f : M(n) \to S(n)$ be the smooth map $f(A) = A\,A^t$. Since $O(n) = f^{-1}(I)$ it suffices to show that I is a regular value of f. We compute:

$$\begin{aligned} df_A(H) &= \lim_{s \to 0} \frac{f(A + s\,H) - f(A)}{s} \\ &= \lim_{s \to 0} \frac{(A + s\,H)(A + s\,H)^t - A\,A^t}{s} \\ &= H\,A^t + A\,H^t. \end{aligned}$$

Fix A with $A\,A^t = I$. We must prove that given any $C \in S(n) = T_I S(n)$ there is $H \in M(n) = T_A M(n)$ such that

$$df_A(H) = H\,A^t + A\,H^t = C.$$

Since C is symmetric, we can write $C = \frac{1}{2}C + \frac{1}{2}C^t$ so if we can solve $H\,A^t = \frac{1}{2}C$ we are done. Multiplying by A on the right and using $A\,A^t = I$ we obtain $H = \frac{1}{2}C\,A$ which solves $df_A(H) = C$. Thus I is a regular value of f.

Note that $O(n)$ is *both* a group and a manifold. In fact, the group operations are smooth, that is, the maps $(A, B) \mapsto A\,B$ and $A \mapsto A^{-1} = A^t$ are smooth. (Why?) A group that is manifold, and whose group operations are smooth, is called a *Lie group*.

1.3.1. Sard's theorem. The Preimage theorem raises the question: is it easy to find regular values? How are abundant are they?

Recall that an arbitrary set A in $\mathbb{R}^n$ has *measure zero* if it can be covered by a countable number of rectangular solids with arbitrary small total volume. In other words given $\varepsilon > 0$, there exists a countable collection $\{R_1, R_2, \dots\}$ of rectangular solids in $\mathbb{R}^n$, such that A is contained in $\cup_i R_i$ and

$$\sum_i \mathrm{Vol}(R_i) < \varepsilon.$$

Let X be manifold. An arbitrary subset $A \subset X$ has *measure zero* if, for every local parametrization ϕ of X, $\phi^{-1}(A)$ has measure zero in Euclidean space.

The following deep theorem, tells us that there are plenty of regular values.

Theorem 1.17 (Sard's theorem, 1942). *The set of critical values of a smooth map $f : X \to Y$ has measure zero.*

Since a set of measure zero cannot contain a non-empty open set we obtain:

Corollary 1.18. *The regular values of any smooth map $f : X \to Y$ are dense in Y.*

1.3.2. Morse functions. See the first example sheet.

1.4. Transversality

We know that the solutions of the equation $f(x) = y$ form a smooth manifold, provided that y is a regular value of $f : X \to Y$. Suppose now that we replace y by a submanifold $Z \subset Y$ and we ask, when is $f^{-1}(Z)$ a submanifold of X?

Definition 1.19. A smooth map $f : X \to Y$ is said to be *transversal* to a submanifold $Z \subset Y$ if for every $x \in f^{-1}(Z)$ we have

$$\mathrm{Image}(df_x) + T_{f(x)}Z = T_{f(x)}Y.$$

We write $f \pitchfork Z$.

Note that if $Z = \{y\}$, the notion of transversality reduces to the notion of regular value.

With this definition, we can now state a more general version of the Preimage theorem.

Theorem 1.20. *If the smooth map $f : X \to Y$ is transversal to a submanifold $Z \subset Y$, then $f^{-1}(Z)$ is submanifold of X. Moreover, $f^{-1}(Z)$ and Z have the same codimension.*

PROOF. (Non-examinable, sketch only.) It is not hard to see that Z can be written in a neighbourhood of a point $y = f(x)$ as the zero set of a collection of functions $h_1, \ldots, h_r$, where r is the codimension of Z in Y. Let $H := (h_1, \ldots, h_r)$. Then near x, $f^{-1}(Z)$ is the zero set of the function $H \circ f$. Thus if $0 \in \mathbb{R}^r$ is a regular value of $H \circ f$ we are done. But $dH_y \circ df_x$ is surjective if and only if

$$\text{Image}(df_x) + T_{f(x)}Z = T_{f(x)}Y$$

since $dH_y : T_yY \to \mathbb{R}^r$ is onto with kernel T_yZ. □

EXAMPLE 1.21. Consider $f : \mathbb{R} \to \mathbb{R}^2$ given by $f(t) = (0, t)$ and let Z be the x axis in $\mathbb{R}^2$. Then f is transversal to Z, but for example the map $h(t) = (t, t^2)$ is not.

An important special case occurs when f is the inclusion of a submanifold X of Y and Z is another submanifold of Y. In this case the condition of transversality reduces to

$$T_xX + T_xZ = T_xY$$

for every $x \in X \cap Z$. This condition is quite easy to visualize and when it holds we say that X and Z are *transversal* (we write $X \pitchfork Z$). We now have:

THEOREM 1.22. *The intersection of two transversal submanifolds of Y is a submanifold of codimension given by*

$$\text{codim}(X \cap Z) = \text{codim}\, X + \text{codim}\, Z.$$

One of the main virtues of the notion of transversality is its *stability* i.e. it survives after small perturbations of the map f. You can convince yourself of this property by looking at pictures of transversal submanifolds of Euclidean space. Another very important virtue is *genericity*: smooth maps may be deformed by arbitrary small amounts into a map that is transversal to Z. (This is non-obvious and we refer to Chapter 2 in the book by Guillemin and Pollack for details.)

1.5. Manifolds with boundary

Consider the closed half-space

$$\mathbb{H}^k := \{(x_1, \ldots, x_k) \in \mathbb{R}^k : \ x_k \geq 0\}.$$

The *boundary* $\partial\mathbb{H}^n$is defined to be the hyperplane $x_k = 0$ in $\mathbb{R}^k$.

DEFINITION 1.23. A subset $X \subset \mathbb{R}^N$ is called a smooth k-manifold with boundary if each $x \in X$ has a neighbourhood diffeomorphic to an open set in $\mathbb{H}^k$. As before, such a diffeomorphism is called a chart on X. The *boundary* of X, denoted ∂X, is given by the set of points that belong to the image of $\partial\mathbb{H}^k$ under some local parametrization. Its complement is called the interior of X, $\text{Int}(X) = X - \partial X$.

REMARK 1.24. Warning: do not confuse the boundary or interior of X as defined above with the topological notions of interior and boundary as a subset of $\mathbb{R}^N$.

The tangent space is defined as before, so that T_xX is a k-dimensional vector space even for points $x \in \partial X$. The interior of X is a k-manifold without boundary and ∂X is a manifold without boundary of dimension $k-1$ (this requires a proof!).

Here is an easy way of generating examples.

LEMMA 1.25. *Let X be a manifold without boundary and let $f : X \to \mathbb{R}$ be a smooth function with 0 as a regular value. Then the subset $\{x \in X : f(x) \geq 0\}$ is a smooth manifold with boundary equal to $f^{-1}(0)$.*

PROOF. The set where $f > 0$ is open in X and is therefore a submanifold of the same dimension as X. For a point $x \in X$ with $f(x) = 0$, the same proof of the Preimage Theorem 1.13 shows that x has a neighbourhood diffeomorphic to a neighbourhood of a point in $\mathbb{H}^k$.

□

As an easy application of the lemma, consider the unit ball B^k given by all $x \in \mathbb{R}^k$ such that $|x| \leq 1$. By considering the function $f(x) = 1 - |x|^2$ it follows that B^k is a smooth manifold with boundary S^{k-1}.

THEOREM 1.26. *Let $f : X \to Y$ be a smooth map from an m-manifold with boundary to an n-manifold, where $m > n$. If y is a regular value, both for f and for the restriction of f to ∂X, then $f^{-1}(y)$ is a smooth $(m-n)$-manifold with boundary equal to $f^{-1}(y) \cap \partial X$.*

PROOF. Recall that being a submanifold is a local property so without loss of generality we can suppose that $f : \mathbb{H}^m \to \mathbb{R}^n$ with $y \in \mathbb{R}^n$ a regular value. Consider $z \in f^{-1}(y)$. If z belongs to the interior of $\mathbb{H}^m$, then as in the Preimage theorem 1.13, $f^{-1}(y)$ is a smooth manifold near z.

Let z be now in $\partial\mathbb{H}^m$. Since f is smooth, there is a neighbourhood U of z in $\mathbb{R}^m$ and a smooth map $F : U \to \mathbb{R}^n$ such that F restricted to $U \cap \mathbb{H}^m$ is f. By shrinking U if necessary we can assume that F has no critical points in U (why?). Hence $F^{-1}(y)$ is a smooth manifold of dimension $m - n$.

Let $\pi : F^{-1}(y) \to \mathbb{R}$ be the projection $(x_1, \ldots, x_m) \mapsto x_m$. Observe that the tangent space of $F^{-1}(y)$ at a point $x \in \pi^{-1}(0)$ is equal to the kernel of

$$dF_x = df_x : \mathbb{R}^m \to \mathbb{R}^n.$$

Hence 0 must be a regular value of π since we are assuming that y is a regular value of f restricted to $\partial\mathbb{H}^m$.

But $F^{-1}(y) \cap \mathbb{H}^m = f^{-1}(y) \cap U$ is the set of all $x \in F^{-1}(y)$ with $\pi(x) \geq 0$ and by Lemma 1.25, is a smooth manifold with boundary equal to $\pi^{-1}(0)$.

□

It is not hard to guess what is the appropriate version of this theorem for the more general case of a map $f : X \to Y$ and a submanifold Z of Y. Suppose X has boundary, but Y and Z are boundaryless. The next theorem is stated without proof.

THEOREM 1.27. *Suppose that both $f : X \to Y$ and $f|_{\partial X} : \partial X \to Y$ are transversal to Z. Then $f^{-1}(Z)$ is a manifold with boundary given by $f^{-1}(Z) \cap \partial X$ and codimension equal to the codimension of Z.*

1.6. Degree modulo 2

Let X be a smooth boundaryless manifold. Then $X \times [0, 1]$ is a manifold with boundary $\partial X = X \times \{0\} \cup X \times \{1\}$.

DEFINITION 1.28. Two maps $f, g : X \to Y$ are called *smoothly homotopic* if there exists a smooth map $F : X \times [0, 1] \to Y$ with $F(x, 0) = f(x)$ and $F(x, 1) = g(x)$ for all $x \in X$. The map F is called a *smooth homotopy* between f and g.

The relation of smooth homotopy is an equivalence relation. (Why?) The equivalence class to which a map belongs is its *homotopy class*. Let $f_t : X \to Y$ be the 1-parameter family of maps given by $f_t(x) = F(x,t)$.

DEFINITION 1.29. The diffeomorphism f is *smoothly isotopic* to g if there exists a smooth homotopy $F : X \times [0,1] \to Y$ from f to g such that for each $t \in [0,1]$, f_t maps X diffeomorphically onto Y.

LEMMA 1.30 (Homotopy Lemma). *Let $f, g : X \to Y$ be smooth maps which are smoothly homotopic. Suppose X is compact, has the same dimension as Y and $\partial X = \emptyset$. If y is a regular value for both f and g, then*

$$\# f^{-1}(y) = \# g^{-1}(y) \pmod 2.$$

PROOF. The proof relies on the following important fact.

THEOREM 1.31 (Classification of 1-manifolds). *Every compact connected 1-manifold is diffeomorphic to $[0,1]$ or S^1.*

As every compact manifold is the disjoint union of finitely many compact connected manifolds we have:

COROLLARY 1.32. *The boundary of any compact 1-manifold consists of an even number of points.*

Let $F : X \times [0,1] \to Y$ be a smooth homotopy between f and g. Assume for the time being that y is also a regular value for F. Then $F^{-1}(y)$ is a compact 1-manifold with boundary equal to

$$F^{-1}(y) \cap (X \times \{0\} \cup X \times \{1\}) = f^{-1}(y) \times \{0\} \cup g^{-1}(y) \times \{1\}.$$

Thus the cardinality of the boundary of $F^{-1}(y)$ is just $\# f^{-1}(y) + \# g^{-1}(y)$. By the corollary above

$$\# f^{-1}(y) = \# g^{-1}(y) \pmod 2.$$

If y is not a regular value of F we proceed as follows. From the Stack of records theorem 1.15 we know that $w \mapsto \# f^{-1}(w)$, $w \mapsto \# g^{-1}(w)$ are locally constant as w ranges over regular values. Thus there are neighbourhoods V and W of y, consisting of regular values of f and g respectively for which

$$\# f^{-1}(w) = \# f^{-1}(y)$$

for all $w \in V$, and

$$\# g^{-1}(w) = \# g^{-1}(y)$$

for all $w \in W$. By Sard's theorem we can choose a regular value z of F in $V \cap W$. Then

$$\# f^{-1}(y) = \# f^{-1}(z) = \# g^{-1}(z) = \# g^{-1}(y) \pmod 2$$

as desired.

□

LEMMA 1.33 (Homogeneity Lemma). *Let X be a smooth connected manifold, possibly with boundary. Let y and z be points in $\mathrm{Int}(X)$. Then there exists a diffeomorphism $h : X \to X$ smoothly isotopic to the identity such that $h(y) = z$.*

PROOF. Let us call two points y and z "isotopic" if there exists a diffeomorphism h isotopic to the identity that maps y to z. It is evident that this is an equivalence relation. Since X is connected, it suffices to show that each equivalence class is a open set.

To prove that equivalence classes are open we will construct an isotopy h_t of $\mathbb{R}^k$ such that h_0 is the identity, each h_t is the identity outside the open unit ball around the origin, and $h_1(0)$ is any specified point in the open unit ball. This will suffice, since we can now parametrize a small neighbourhood of y in $\mathrm{Int}(X)$ and use h_t to construct an isotopy that will move y to any point nearby.

Let $\varphi : \mathbb{R}^k \to \mathbb{R}$ be a smooth function such that $\varphi(x) > 0$ for $|x| < 1$ and $\varphi(x) = 0$ for $|x| \geq 1$ (why do they exist?). Given a unit vector $u \in \mathbb{R}^k$ consider the ordinary differential equation in $\mathbb{R}^k$ given by

$$\frac{dx}{dt} = u\,\varphi(x).$$

Let $F_t : \mathbb{R}^k \to \mathbb{R}^k$ be the flow of this differential equation, i.e., for each $x \in \mathbb{R}^k$, the curve $t \mapsto F_t(x)$ is the unique solution passing through x. Standard theorems on differential equations tell us that:

(1) F_t is defined for all $x \in \mathbb{R}^k$ and $t \in \mathbb{R}$ and smooth;
(2) F_0 is the identity;
(3) $F_{t+s} = F_t \circ F_s$.

Clearly for each t, F_t leaves all points outside the unit ball fixed. For appropriate choices of u and t, F_t will map the origin to any point in the open unit ball.

(I may replace this proof by the one given by Guillemin and Pollack on page 142 of their book which does not use flows.)

□

In what follows suppose that X is compact and without boundary and Y is connected and with the same dimension as X. Let $f : X \to Y$ be a smooth map.

THEOREM 1.34 (Degree mod 2). *If y and z are regular values of f then*

$$\#\, f^{-1}(y) = \#\, f^{-1}(z) \pmod 2.$$

This common residue class is called the degree mod 2 of f (denoted $\deg_2(f)$*) and only depends on the homotopy class of f.*

PROOF. Given y and z, let h be the diffeomorphism smoothly isotopic to the identity such that $h(y) = z$ given by the Homogeneity Lemma. Observe that z is also a regular value of $h \circ f$. Since $h \circ f$ is homotopic to f, the Homotopy Lemma tells us that

$$\#\,(h \circ f)^{-1}(z) = \#\, f^{-1}(z) \pmod 2.$$

But $(h \circ f)^{-1}(z) = f^{-1}(h^{-1}(z)) = f^{-1}(y)$ and therefore

$$\#\, f^{-1}(y) = \#\, f^{-1}(z) \pmod 2$$

as desired. Let g be smoothly homotopic to f. By Sard's theorem, there is a point $y \in Y$ which is a regular value for both f and g (why?). By the Homotopy Lemma

$$\deg_2(f) = \#\, f^{-1}(y) \pmod 2 = \#\, g^{-1}(y) \pmod 2 = \deg_2(g)$$

which completes the proof.

□

EXAMPLE 1.35. The identity map of a compact boundaryless manifold X has $\deg_2 = 1$ and the constant map has $\deg_2 = 0$. Therefore they are never homotopic. When $X = S^n$, this implies that there is no smooth map $f : B^{k+1} \to S^k$ which restricts to the identity on S^k (i.e. there is no *retraction*). Indeed, if such a map exists it would give rise to a homotopy $F : S^k \times [0,1] \to S^k$, $F(x,t) = f(tx)$ which is a homotopy between the constant map and the identity.

The previous example yields the smooth version of a famous fixed point theorem.

THEOREM 1.36 (Smooth Brouwer fixed point theorem). *Any smooth map $f : B^k \to B^k$ has a fixed point.*

PROOF. Suppose f has no fixed point. We will construct a map $g : B^k \to S^{k-1}$ which restricts to the identity on S^{k-1} which contradicts Example 1.35. For $x \in B^k$, let $g(x) \in S^{k-1}$ be the point where the line segment starting at $f(x)$ and passing through x hits the boundary. One can write a formula for g to show smoothness. $\square$

In fact, any continuous map $f : B^k \to B^k$ has a fixed point (Brouwer fixed point theorem). One can prove this using the smooth version approximating f by polynomials. Indeed, by the Weierstrass approximation theorem, given $\varepsilon > 0$, there is a polynomial P such that $|f(x) - P(x)| < \varepsilon$ for all $x \in B^k$. Set $Q(x) = P(x)/(1+\varepsilon)$. Now Q maps B^k into B^k and $|f(x) - Q(x)| < 2\varepsilon$ for all $x \in B^k$. Suppose $f(x) \neq x$ for all x. Then the continuous function $|f(x) - x|$ must take a positive minimum τ on B^k. If now let $\varepsilon = \tau/2$ the Q from above does not have a fixed point, which contradicts the smooth Brouwer fixed point theorem.

1.6.1. Intersection numbers modulo 2. What we said above about degree modulo 2 can be generalized to the case of a smooth map $f : X \to Y$ and Z a submanifold of Y. Suppose that:

(1) X is compact without boundary;
(2) Z is closed and without boundary;
(3) $f \pitchfork Z$;
(4) $\dim X + \dim Z = \dim Y$.

Under these conditions $f^{-1}(Z)$ is a closed 0-dimensional submanifold of X and hence it consists of finitely many points. Define the *mod 2 intersection number* of the map f with Z, $I_2(f,Z)$, to be the cardinality of $f^{-1}(Z)$ modulo 2. An analogue of the Homotopy Lemma (for intersection numbers) also holds here: if f_0 and f_1 are maps transversal to Z and homotopic, then $I_2(f_0,Z) = I_2(f_1,Z)$. In fact, we can define $I_2(f,Z)$ for a map f which is not necessarily transversal to Z. Using that transversality is generic, we can find a map g homotopic to f such that $g \pitchfork Z$ and we now set $I_2(f,Z) := I_2(g,Z)$. It does not matter which g we choose as long as g is homotopic to f, thanks to the Homotopy Lemma for intersection numbers.

Let us have a closer look at the important special case in which X itself is a compact submanifold of Y and Z a closed submanifold of complementary dimension. In this case f is the inclusion map $X \hookrightarrow Y$, and if $X \pitchfork Z$, $I_2(X,Z) := I_2(f,Z)$ is just $\#(X \cap Z) \bmod 2$. If $I_2(X,Z) \neq 0$, it means that no matter how X is deformed, we cannot move it away completely from Z.

EXAMPLE 1.37. Let Y be the 2-torus $S^1 \times S^1$. Let $X = S^1 \times \{1\}$ and $Z = \{1\} \times S^1$. Then $I_2(X,Z) = 1$ and we cannot smoothly pull the circles apart.

When $\dim X = \frac{1}{2}\dim Y$, we can consider $I_2(X, X)$ the *self-intersection number modulo 2*. If X is the central curve in a Möbius band, then $I_2(X, X) = 1$.

1.7. Abstract manifolds and Whitney's embedding theorem

One can actually define manifolds without making any reference to the ambient space $\mathbb{R}^N$. You will study manifolds in this more abstract setting in Part III and in the Riemann Surfaces course. In any case, here is the definition:

DEFINITION 1.38. An n-dimensional smooth manifold is a second countable Hausdorff space X together with a collection of maps called *charts* such that:

(1) a chart is a homeomorphism $\phi : U \to \phi(U)$, where U is open in X and $\phi(U)$ is open in $\mathbb{R}^n$;
(2) each point $x \in X$ belongs to the domain of some chart;
(3) for charts $\phi : U \to \phi(U) \subset \mathbb{R}^n$ and $\psi : V \to \phi(V) \subset \mathbb{R}^n$, the map $\phi \circ \psi^{-1} : \psi(U \cap V) \to \phi(U \cap V)$ is smooth;
(4) the collection of charts is maximal with respect to the properties above.

A set of charts satisfying the first three properties is called an *atlas*. An atlas can always be enlarged uniquely to give a maximal atlas as in the definition. If in the definition we require the maps $\phi \circ \psi^{-1}$ just to be of class C^k ($k \geq 1$) then we say that we have a manifold of class C^k. (Recall that a map is of class C^k if it has continuous partial derivatives up to order k.)

The definition is set up so that it is plain how to define a smooth map between manifolds. An *immersion* is a smooth map $f : X \to Y$ such that df_x is injective for all $x \in X$. A *submersion* is a smooth map $f : X \to Y$ such that df_x is surjective for all $x \in X$. An *embedding* is a smooth map $f : X \to Y$ which is an immersion and a homeomorphism onto its image.

The next theorem tells us that we did not lose much by restricting out attention to manifolds as subsets of Euclidean space.

THEOREM 1.39 (Whitney's embedding theorem). *A smooth n-manifold X can be embedded into $\mathbb{R}^{2n+1}$.*

The proof of this theorem can be found in most books about manifolds. In fact, Whitney proved the much harder result that X can be embedded in $\mathbb{R}^{2n}$.

CHAPTER 2

Length, area and curvature

2.1. Arc-length, curvature and torsion of curves

DEFINITION 2.1. Let $I \subset \mathbb{R}$ be an interval and let X be a manifold. A *curve* in X is a smooth map $\alpha : I \to X$. The curve is said to be *regular* if α is an immersion, i.e., if the *velocity vector* $\dot{\alpha}(t) := d\alpha_t(1) \in T_{\alpha(t)}X$ is never zero.

If $X = \mathbb{R}^2, \mathbb{R}^3$ (or any "Riemannian manifold") we can talk about the *arc-length* of a regular curve. By definition, given $t \in I$, the arc-length of $\alpha : I \to \mathbb{R}^3$ from the point t_0 is given by

$$s(t) := \int_{t_0}^{t} |\dot{\alpha}(\tau)|\, d\tau.$$

If the interval I has endpoints a and b, $a < b$, the *length* of α is

$$\ell(\alpha) := \int_{a}^{b} |\dot{\alpha}(t)|\, dt.$$

The curve is said to be parametrized by arc-length if $|\dot{\alpha}(t)| = 1$ for all $t \in I$. Since the curve is regular, the function $t \mapsto s(t)$ is strictly increasing, and therefore there exists a smooth inverse $t = t(s)$. The new curve $s \mapsto \alpha(t(s))$ is now parametrized by arc-length and has the same image and the same length as α.

From now on we will assume that curves are parametrized by arc-length.

DEFINITION 2.2. Let $\alpha : I \to \mathbb{R}^3$ be a curve parametrized by arc-length. The *curvature* of α at $s \in I$ is the number $k(s) := |\ddot{\alpha}(s)|$. If $k(s) \neq 0$, the unit vector $n(s)$ in the direction of $\ddot{\alpha}(s)$ is well defined by the equation $\ddot{\alpha}(s) = k(s)\, n(s)$. Differentiating $|\dot{\alpha}|^2 = 1$ we see that $\langle \dot{\alpha}, \ddot{\alpha} \rangle = 0$ and thus $n(s)$ is normal to $\dot{\alpha}(s)$ and is called the *normal vector* at s. The plane determined by $\dot{\alpha}(s)$ and $n(s)$ is called the *osculating plane* at s.

Traditionally one denotes $\dot{\alpha}$ by $t(s)$. The unit vector $b(s) := t(s) \wedge n(s)$, where $\wedge$ denotes the usual cross product in $\mathbb{R}^3$, is normal to the osculating plane and is called the *binormal vector* at s. Note

$$\dot{b}(s) = \dot{t}(s) \wedge n(s) + t(s) \wedge \dot{n}(s) = t(s) \wedge \dot{n}(s),$$

thus we may write

$$\dot{b}(s) = \tau(s)\, n(s)$$

for some function $\tau(s)$. (Note that $\langle b, \dot{b} \rangle = 0$ since b has unit norm.)

DEFINITION 2.3. The number $\tau(s)$ is called the *torsion* of the curve at s.

For each point s for which $k(s) \neq 0$, we have three orthonormal vectors $t(s), n(s)$ and $b(s)$ which form the so called *Frenet trihedron* at s. Observe that since $n = b \wedge t$, differentiating with respect to s we obtain:

$$\dot{n}(s) = \dot{b}(s) \wedge t(s) + b(s) \wedge \dot{t}(s) = -\tau(s)b(s) - k(s)t(s).$$

Thus we have obtained the following three equations, called the *Frenet formulas*:

$$\begin{aligned} \dot{t} &= k\, n, \\ \dot{n} &= -k\, t - \tau\, b, \\ \dot{b} &= \tau\, n. \end{aligned}$$

We can think of a curve in $\mathbb{R}^3$ as being obtained from a straight line by bending (curvature) and twisting (torsion). It is not hard to guess that curvature and torsion determine the curve locally. Indeed we have:

THEOREM 2.4 (Fundamental theorem of the local theory of curves). *Given smooth functions $k(s) > 0$ and $\tau(s)$, $s \in I$, there exists a regular curve $\alpha : I \to \mathbb{R}^3$ such that s is arc-length, $k(s)$ is the curvature, and $\tau(s)$ is the torsion of α. Moreover any other curve $\overline{\alpha}$, satisfying the same conditions, differs from α by an isometry; that is, there exists an orthogonal linear map T, with positive determinant, and a vector a such that $\overline{\alpha} = T \circ \alpha + a$.*

You are invited to give a proof by using the Frenet formulas and the theorem on existence and uniqueness of solutions of ordinary differential equations that you saw in Analysis II.

REMARK 2.5. If α is a *plane curve*, that is, $\alpha(I)$ is contained in a plane, then $\tau \equiv 0$. In this case the proof of the fundamental theorem is actually quite simple.

For plane curves $\alpha : I \to \mathbb{R}^2$, one can give a sign to the curvature.

Let $\{e_1, e_2\}$ be the standard basis of $\mathbb{R}^2$ and define the normal vector $n(s)$, $s \in I$, by requiring the basis $\{t(s), n(s)\}$ to have the same orientation as $\{e_1, e_2\}$. The curvature k is then defined by the equation:

$$\dot{t}(s) = k(s)\, n(s)$$

which might be either positive or negative. The absolute value of k coincides with the previous definition.

2.2. The isoperimetric inequality in the plane

There are many proofs of the isoperimetric inequality. The one we present here uses a bit of Fourier analysis and is due to A. Hurwitz (1901). This is our first example of a *global result*, and its extensions and ramifications are of central importance in Differential Geometry.

Let $\Omega \subset \mathbb{R}^2$ be a *domain*, that is, a connected open set. We will assume that Ω has compact closure and that its boundary $\partial\Omega$ is a connected 1-manifold of class C^1. We will denote by $A(\Omega)$ the area of Ω.

THEOREM 2.6 (The isoperimetric inequality in the plane). *Let Ω be as above. Then*

$$\ell^2(\partial\Omega) \geq 4\pi\, A(\Omega),$$

with equality if and only if Ω is a disk.

We will need a preliminary lemma, which is an exercise in Fourier series.

LEMMA 2.7 (Wirtinger's Inequality). *Let $f : \mathbb{R} \to \mathbb{R}$ be a C^1 function which is periodic with period L. Suppose*

$$\int_0^L f(t)\,dt = 0.$$

Then

$$\int_0^L |f'|^2(t)\,dt \geq \frac{4\pi^2}{L^2}\int_0^L |f|^2(t)\,dt,$$

with equality if and only if there exist constants a_{-1} and a_1 such that

$$f(t) = a_{-1}\,e^{-2\pi it/L} + a_1\,e^{2\pi it/L}.$$

PROOF. Consider the Fourier series expansions of f and f':

$$f(t) = \sum_{k=-\infty}^{\infty} a_k\,e^{2\pi ikt/L},$$

$$f'(t) = \sum_{k=-\infty}^{\infty} b_k\,e^{2\pi ikt/L}$$

where

$$a_k = \frac{1}{L}\int_0^L f(t)e^{-2\pi ikt/L}\,dt$$

$$b_k = \frac{1}{L}\int_0^L f'(t)e^{-2\pi ikt/L}\,dt.$$

The hypothesis implies $a_0 = 0$. Also note that

$$b_0 = \frac{1}{L}\int_0^L f'(t)\,dt = \frac{1}{L}(f(L) - f(0)) = 0.$$

Integration by parts gives

$$b_k = \frac{2\pi ik}{L}\,a_k$$

for all $|k| \geq 1$. Parseval's identity gives:

$$\begin{aligned}\int_0^L |f'|^2\,dt = L\sum_{k\neq 0}|b_k|^2 &= \frac{4\pi^2}{L}\sum_{k\neq 0} k^2\,|a_k|^2\\ &\geq \frac{4\pi^2}{L}\sum_{k\neq 0}|a_k|^2\\ &= \frac{4\pi^2}{L^2}\int_0^L |f|^2\,dt.\end{aligned}$$

Equality holds if and only if a_k vanishes for all k with $|k| > 1$.

□

PROOF OF THEOREM 2.6. By translating Ω if necessary, we can assume that

$$\int_{\partial\Omega} X\,ds = 0$$

where $X(x,y) = (x,y) \in \mathbb{R}^2$. Let us apply the 2-dimensional divergence theorem to the vector field X in the domain Ω. Let n be the outward unit normal vector field along $\partial\Omega$. The divergence theorem says

$$\int_\Omega \mathrm{div} X \, dA = \int_{\partial\Omega} \langle X, n\rangle \, ds.$$

Since $\mathrm{div} X = 2$, we have

$$2\,A(\Omega) = \int_{\partial\Omega} \langle X, n\rangle \, ds.$$

The Cauchy-Schwarz inequality for the inner product of $\mathbb{R}^2$ gives

$$\langle X, n\rangle \leq |X|$$

and thus

$$2\,A(\Omega) \leq \int_{\partial\Omega} |X| \, ds. \tag{2.1}$$

Now the integral Cauchy-Schwarz inequality yields

$$2\,A(\Omega) \leq \left(\int_{\partial\Omega} |X|^2 \, ds\right)^{1/2} \left(\int_{\partial\Omega} 1^2 \, ds\right)^{1/2} = \ell^{1/2}(\partial\Omega) \left(\int_{\partial\Omega} |X|^2 \, ds\right)^{1/2}. \tag{2.2}$$

Since we are parametrizing $\partial\Omega$ by arc-length, the components of $X(s) = (x(s), y(s))$ along $\partial\Omega$ are C^1 periodic functions of period $L = \ell(\partial\Omega)$. By Wirtinger's inequality 2.7 (applied to each component of $X(s)$) we have

$$\left(\int_{\partial\Omega} |X|^2 \, ds\right)^{1/2} \leq \left(\frac{\ell^2(\partial\Omega)}{4\pi^2} \int_{\partial\Omega} |X'|^2 \, ds\right)^{1/2} = \ell^{3/2}(\partial\Omega)/2\pi \tag{2.3}$$

Combining (2.2) and (2.3) yields:

$$2\,A(\Omega) \leq \ell^2(\partial\Omega)/2\pi,$$

with equality if and only if we have equality in (2.1), (2.2) and (2.3). But if we have equality in (2.2), the characterization of the equality case in the Cauchy-Schwarz inequality implies that $s \mapsto |X(s)|$ is constant, i.e. Ω must be a disk.

□

2.3. First fundamental form and Area

A *surface* is a 2-dimensional manifold S. In this section we begin the study of surfaces which live inside $\mathbb{R}^3$. As in the case of curves, the inner product of $\mathbb{R}^3$ will induce an important geometric structure on the surface: *a Riemannian metric*. As we move along you will be able to guess that many of the ideas that we present actually hold with much greater generality. A good grasp of these ideas will make Part III courses in Geometry/Topology much easier to understand. The old fashion term for Riemannian metric in this context is *First fundamental form*.

Definition 2.8. Let $S \subset \mathbb{R}^3$ be a surface. The quadratic form I_p on T_pS given by

$$I_p(w) := \langle w, w\rangle = |w|^2$$

is called the first fundamental form of the surface at p.

In general, if X is a manifold, a Riemannian metric on X, is a smooth map that asigns to each $p \in X$ an inner product g_p in T_pX. If $f : X \to Y$ is an immersion and Y has a Riemannian metric g, then using df we can endow X with a Riemannian metric h defined by

$$h_p(w) := g_{f(p)}(df_p(w)).$$

The first fundamental form is the Riemannian metric on $S = X$ determined by the inner product in $\mathbb{R}^3 = Y$, where f is the inclusion map $S \hookrightarrow \mathbb{R}^3$.

DEFINITION 2.9. Two surfaces S_1 and S_2 are said to be *isometric* if there exists a diffeomorphism $f : S_1 \to S_2$ such that for all $p \in S_1$, df_p is a linear isometry between T_pS_1 and $T_{f(p)}S_2$.

A Riemannian metric will allow us to do various measurements on the surface, like length and area. Before we go into that, let us express a Riemannian metric locally.

Let $\phi : U \subset \mathbb{R}^2 \to S \subset \mathbb{R}^3$ be a parametrization of a neighbourhood of a point $p \in S$. We will denote by (u, v) points in U and let

$$\phi_u(u,v) := \frac{\partial \phi}{\partial u}(u,v) \in T_{\phi(u,v)}S,$$
$$\phi_v(u,v) := \frac{\partial \phi}{\partial v}(u,v) \in T_{\phi(u,v)}S.$$

(These vectors are linearly independent, why?) Set

$$\begin{aligned} E(u,v) &:= \langle \phi_u, \phi_u \rangle_{\phi(u,v)}, \\ F(u,v) &:= \langle \phi_u, \phi_v \rangle_{\phi(u,v)}, \\ G(u,v) &:= \langle \phi_v, \phi_v \rangle_{\phi(u,v)}. \end{aligned}$$

Since a tangent vector $w \in T_pS$ is the tangent vector of a curve $\alpha(t) = \phi(u(t), v(t))$, $t \in (-\varepsilon, \varepsilon)$, with $p = \alpha(0) = \phi(u_0, v_0)$ we have

$$\begin{aligned} I_p(\dot\alpha(0)) &= \langle \dot\alpha(0), \dot\alpha(0) \rangle_p \\ &= \langle \phi_u, \phi_u \rangle_p (\dot u)^2 + 2\langle \phi_u, \phi_v \rangle_p \, \dot u \, \dot v + \langle \phi_v, \phi_v \rangle_p (\dot v)^2 \\ &= E(\dot u)^2 + 2\, F \, \dot u \, \dot v + G(\dot v)^2. \end{aligned}$$

EXAMPLE 2.10. Let S be the torus obtained by rotating a circle of radius r about a straight line belonging to the plane of the circle and at a distance $a > r$ away from the centre of the circle. A parametrization which covers all of the torus except for a meridian and a parallel is given by

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

where $u, v \in (0, 2\pi)$. Now we can easily compute E, F and G:

$$E = r^2, \quad F = 0, \quad G = (r\, \cos u + a)^2.$$

The local coefficients E, F and G are useful when we wish to compute the length of a curve in S. If we can express the curve in local coordinates we just have to integrate $\sqrt{E(\dot u)^2 + 2\, F\, \dot u\, \dot v + G(\dot v)^2}$.

EXERCISE 2.11. Show that

$$|\phi_u \wedge \phi_v| = \sqrt{EG - F^2}.$$

We now define the *area* of a bounded domain in a surface. A *domain* Ω of S is an open and connected subset of S. It is bounded if it is contained in some ball in $\mathbb{R}^3$. Suppose that Ω is contained in the image of a parametrization $\phi : U \to S$.

LEMMA 2.12. *The integral*

$$\int_{\phi^{-1}(\Omega)} |\phi_u \wedge \phi_v| \, du \, dv$$

does not depend on the parametrization.

PROOF. We have to show that if $\psi : \bar{U} \to S$ is another parametrization with $\Omega \subset \psi(\bar{U})$, then

$$\int_{\phi^{-1}(\Omega)} |\phi_u \wedge \phi_v| \, du \, dv = \int_{\psi^{-1}(\Omega)} |\psi_{\bar{u}} \wedge \psi_{\bar{v}}| \, d\bar{u} \, d\bar{v}.$$

Let $J(\bar{u}, \bar{v})$ be the Jacobian of $h := \phi^{-1} \circ \psi$. Since

$$|\psi_{\bar{u}} \wedge \psi_{\bar{v}}| = |J| \, |\phi_u \wedge \phi_v| \circ h$$

the lemma follows from the formula for the change of variables of multiple integrals that you have seen in the Vector Calculus course (h is the change of variables). □

DEFINITION 2.13. Let $\Omega \subset S$ be a bounded domain contained in the image of a parametrization $\phi : U \to S$. The positive number

$$A(\Omega) := \int_{\phi^{-1}(\Omega)} |\phi_u \wedge \phi_v| \, du \, dv$$

is called the *area* of Ω.

It is possible to define the area of open sets which are not contained in the image of a parametrization. This is done using a *partition of unity*, which is technical gadget designed to glue smooth objects in a manifold. You will see this in Part III. For practical purposes one never uses partitions of unity to compute area. Usually a clever choice of parametrization covers the set that we are interesed in except for some curves (i.e. sets of measure zero) which do not contribute to the area. In fact, using the *exponential map* it is possible to produce always such parametrizations. Once you have the notion of area for open sets, you can extend it to a measure defined on the Borel sigma algebra generated by the open sets. In this form we obtain the *Riemannian measure*. Different Riemannian metrics give rise to different Riemannian measures, but they all have the same measure zero sets. We will denote the Riemannian measure by dA and we will have the chance to use it when we discuss the Gauss-Bonnet theorem. If $f : S \to \mathbb{R}$ is a continuous function and $\phi : U \to S$ is a parametrization that covers S up to a set of measure zero, then

$$\int_S f \, dA = \int_U f(u,v) \sqrt{EG - F^2} \, du dv.$$

EXAMPLE 2.14. Let us compute the area A of the torus from Example 2.10. Recall that

$$|\phi_u \wedge \phi_v| = \sqrt{EG - F^2}.$$

For our example

$$\sqrt{EG - F^2} = r(r \cos u + a)$$

thus

$$A = \int_{(0,2\pi)\times(0,2\pi)} r(r\cos u + a)\, du\, dv = 4\pi^2 ra.$$

2.4. The Gauss map

Given a parametrization $\phi : U \subset \mathbb{R}^2 \to S \subset \mathbb{R}^3$ around a point $p \in S$, we can choose a unit normal vector at each point of $\phi(U)$ by setting

$$N(q) := \frac{\phi_u \wedge \phi_v}{|\phi_u \wedge \phi_v|}(q).$$

Thus we have smooth map $N : \phi(U) \to \mathbb{R}^3$, which in fact takes values in $S^2 \subset \mathbb{R}^3$.

Definition 2.15. *A smooth field of unit normal vectors* on a surface S is a smooth map $N : S \to \; S^2 \subset \mathbb{R}^3$ such that for every $p \in S$, $N(p)$ is orthogonal to T_pS. A surface $S \subset \mathbb{R}^3$ is *orientable* if it admits a smooth field of unit normal vectors. The choice of such a field is called an *orientation*.

The Möbius band is an example of a non-orientable surface (why?). Orientability is definitely a global property. Observe that an orientation as we have defined it induces an orientation on every tangent plane T_pS by declaring a basis $\{u, v\}$ *positively oriented* if $\{u, v, N\}$ is a positively oriented basis of $\mathbb{R}^3$.

Definition 2.16. Let S be an oriented surface and $N : S \to S^2$ the smooth field of unit normal vectors defining the orientation. The map N is called *the Gauss map* of S.

The Gauss map contains lots of geometric information about S. To start unpacking it we look at its derivative $dN_p : T_pS \to T_{N(p)}S^2$. Since T_pS and $T_{N(p)}S^2$ are parallel planes, we will regard dN_p as a linear map $dN_p : T_pS \to T_pS$. We state and prove at once the key property of the Gauss map:

Proposition 2.17. *The linear map $dN_p : T_pS \to T_pS$ is selfadjoint.*

Proof. Let $\phi : U \to S$ be a parametrization around p. If $\alpha(t) = \phi(u(t), v(t))$ is a curve in $\phi(U)$ with $\alpha(0) = p$ we have

$$\begin{aligned} dN_p(\dot\alpha(0)) &= dN_p(\dot u(0)\phi_u + \dot v(0)\phi_v) \\ &= \frac{d}{dt}\,|_{t=0}\, N(u(t), v(t)) \\ &= \dot u(0)N_u + \dot v(0)N_v. \end{aligned}$$

In particular $dN_p(\phi_u) = N_u$ and $dN_p(\phi_v) = N_v$ and since $\{\phi_u, \phi_v\}$ is a basis of the tangent plane, we only have to prove that

$$\langle N_u, \phi_v\rangle = \langle N_v, \phi_u\rangle.$$

To prove the last equality, observe that $\langle N, \phi_u\rangle = \langle N, \phi_v\rangle = 0$. Take derivatives with respect to v and u to obtain:

$$\langle N_v, \phi_u\rangle + \langle N, \phi_{uv}\rangle = 0,$$
$$\langle N_u, \phi_v\rangle + \langle N, \phi_{vu}\rangle = 0,$$

which gives the desired equality.

$\square$

DEFINITION 2.18. The quadratic form defined on T_pS by $II_p(w) := -\langle dN_p(w), w\rangle$ is called *the second fundamental form* of S at p.

Let $\alpha : (-\varepsilon, \varepsilon) \to S$ be a curve parametrized by arc-length with $\alpha(0) = p$. If we let $N(s)$ be $N \circ \alpha$ we have $\langle N(s), \dot\alpha(s)\rangle = 0$ for all s. If we differentiate we obtain:

$$\langle N(s), \ddot\alpha(s)\rangle = -\langle \dot N(s), \dot\alpha(s)\rangle.$$

But $II_p(\dot\alpha(0)) = -\langle \dot N(0), \dot\alpha(0)\rangle$ and thus

$$II_p(\dot\alpha(0)) = \langle N(0), \ddot\alpha(0)\rangle = \langle N, k\, n\rangle(p)$$

where k is the curvature of α and n its unit normal. The expression $\langle N, k\, n\rangle(p)$ is called the *normal curvature* of α at p and denoted by $k_n(p)$. Observe that $k_n(p)$ only depends on the tangent vector $\dot\alpha(0)$.

Recall from Linear Maths that any selfadjoint linear map can be diagonalized in an orthonormal basis. Hence there exist an orthonormal basis $\{e_1, e_2\}$ of T_pS and real numbers k_1 and k_2 with $k_1 \geq k_2$ such that $dN_p(e_i) = -k_i\, e_i$ for $i = 1, 2$. Moreover, the numbers k_1 and k_2 are the maximum and minimum of II_p restricted to the set of unit vectors on T_pS. (They are extreme values of the normal curvature at p.)

DEFINITION 2.19. The numbers k_1 and k_2 are called the *principal curvatures* at p and e_1 and e_2 are called the *principal directions* at p.

DEFINITION 2.20. The determinant of dN_p is the *Gaussian curvature* $K(p)$ of S at p. Minus half of the trace of dN_p is called the *mean curvature* $H(p)$ of S at p.

Clearly $K = k_1\, k_2$ and $H = \frac{k_1+k_2}{2}$. A point $p \in S$ of a surface is called *elliptic* if $K(p) > 0$, *hyperbolic* if $K(p) < 0$, *parabolic* if $K(p) = 0$ and $dN_p \neq 0$, and planar if $dN_p = 0$.

EXAMPLE 2.21. The points of a round sphere are elliptic points. The point $(0, 0, 0)$ of the paraboloid $z = x^2 + y^2$ is also elliptic. The point $(0, 0, 0)$ of the hyperboloid $z = x^2 - y^2$ is a hyperbolic point. The points of a cylinder are parabolic points and the points of a plane are planar.

A point $p \in S$ is called *umbilical* if $k_1 = k_2$.

PROPOSITION 2.22. *If all points on a connected surface are umbilical, then S is either contained in a sphere or a plane.*

PROOF. See Example sheet 2. □

2.5. The second fundamental form in local coordinates

Let $\phi : U \to S$ be a parametrization around a point $p \in S$. Let us express the second fundamental form in the basis $\{\phi_u, \phi_v\}$. Since $\langle N, \phi_u\rangle = \langle N, \phi_v\rangle = 0$ we have

$$\begin{aligned} e &:= -\langle N_u, \phi_u\rangle = \langle N, \phi_{uu}\rangle, \\ f &:= -\langle N_v, \phi_u\rangle = \langle N, \phi_{uv}\rangle = -\langle N_u, \phi_v\rangle, \\ g &:= -\langle N_v, \phi_v\rangle = \langle N, \phi_{vv}\rangle. \end{aligned}$$

The coefficients e, f and g are sometimes also denoted by L, M and N respectively.

If α is a curve passing at $t = 0$ through p we can write:

$$II_p(\dot\alpha(0)) = -\langle dN_p(\dot\alpha(0)), \dot\alpha(0)\rangle = e(\dot u)^2 + 2\, f\, \dot u\, \dot v + g(\dot v)^2.$$

It is now a simple exercise in Linear Maths to express K and H in terms of the local coefficients E, F, G and e, f, g. Write:

$$dN_p(\phi_u) = N_u = a_{11}\phi_u + a_{21}\phi_v,$$
$$dN_p(\phi_v) = N_v = a_{12}\phi_u + a_{22}\phi_v.$$

Taking inner products of these equations with ϕ_u and ϕ_v we obtain:

$$\begin{pmatrix} E & F \\ F & G \end{pmatrix}\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} = -\begin{pmatrix} e & f \\ f & g \end{pmatrix}. \tag{2.4}$$

Check now that:

$$K = \frac{eg - f^2}{EG - F^2}, \tag{2.5}$$

$$H = \frac{eG - 2fF + gE}{2(EG - F^2)}. \tag{2.6}$$

The principal curvatures are the roots of the characteristic polynomial:

$$k^2 - 2Hk + K = 0.$$

Example 2.23. Let us compute the Gaussian curvature of the torus from Example 2.10. We have already calculated E, F and G. To compute e, f and g we use the formulas above to obtain

$$e = r, \quad f = 0, \quad g = \cos u(a + r\cos u).$$

Thus

$$K = \frac{\cos u}{r(a + r\cos u)}.$$

We see that $K = 0$ along the parallels $u = \pi/2$ and $u = 3\pi/2$ (parabolic points). In the region $\pi/2 < u < 3\pi/2$, K is negative (hyperbolic points) and in the region $0 < u < \pi/2$ or $3\pi/2 < u < 2\pi$, K is positive (elliptic points).

2.6. Theorema Egregium

Theorem 2.24 (Theorema Egregium, Gauss 1827). *The Gaussian curvature K of a surface is invariant under isometries.*

Proof. Since by definition isometries preserve the first fundamental form it suffices to prove that K can be expressed in local coordinates purely in terms of the coefficients E, F and G of the first fundamental form and their derivatives.

Let $\phi : U \to S$ be a parametrization. On each point of $\phi(U)$ we have a basis of $\mathbb{R}^3$ given by $\{\phi_u, \phi_v, N\}$.

We can now express the derivatives of the vectors ϕ_u, ϕ_v in this basis to obtain (recall the definitions of the coefficients e, f and g):

$$\phi_{uu} = \Gamma_{11}^1\phi_u + \Gamma_{11}^2\phi_v + eN, \tag{2.7}$$
$$\phi_{uv} = \Gamma_{12}^1\phi_u + \Gamma_{12}^2\phi_v + fN, \tag{2.8}$$
$$\phi_{vu} = \Gamma_{21}^1\phi_u + \Gamma_{21}^2\phi_v + fN, \tag{2.9}$$
$$\phi_{vv} = \Gamma_{22}^1\phi_u + \Gamma_{22}^2\phi_v + gN, \tag{2.10}$$

The coefficients Γ_{ij}^k are called the *Christoffel symbols*. Since $\phi_{uv} = \phi_{vu}$ we have $\Gamma_{12}^1 = \Gamma_{21}^1$ and $\Gamma_{12}^2 = \Gamma_{21}^2$.

If in the first equation we take inner products with ϕ_u and ϕ_v we obtain:

$$\Gamma_{11}^1 E + \Gamma_{11}^2 F = \langle \phi_{uu}, \phi_u \rangle = \frac{1}{2} E_u, \tag{2.11}$$

$$\Gamma_{11}^1 F + \Gamma_{11}^2 G = \langle \phi_{uu}, \phi_v \rangle = F_u - \frac{1}{2} E_v. \tag{2.12}$$

Since $EG - F^2 \neq 0$ we can solve for Γ_{11}^1 and Γ_{11}^2 and arguing in a similar fashion with the other equations we conclude that we can express the Christoffel symbols in terms of E, F and G and their first derivatives.

Consider the identity:

$$\phi_{uuv} = \phi_{uvu}.$$

Using (2.7) and (2.8) we get:

$$\begin{aligned} &\Gamma_{11}^1 \phi_{uv} + \Gamma_{11}^2 \phi_{vv} + e N_v + \left(\Gamma_{11}^1\right)_v \phi_u + \left(\Gamma_{11}^2\right)_v \phi_v + e_v N \\ &\qquad = \Gamma_{12}^1 \phi_{uu} + \Gamma_{12}^2 \phi_{vu} + f N_u + \left(\Gamma_{12}^1\right)_u \phi_u + \left(\Gamma_{12}^2\right)_u \phi_v + f_u N. \end{aligned}$$

Using equations (2.7), (2.8) and (2.10) and equating the coefficients of ϕ_v one obtains after some manipulations using (2.4) from the previous section:

$$\begin{aligned} &\left(\Gamma_{12}^2\right)_u - \left(\Gamma_{11}^2\right)_v + \Gamma_{12}^1 \Gamma_{11}^2 + \Gamma_{12}^2 \Gamma_{12}^2 - \Gamma_{11}^2 \Gamma_{22}^2 - \Gamma_{11}^1 \Gamma_{12}^2 \\ &\qquad = -f a_{21} + e a_{22} = -E \frac{eg - f^2}{EG - F^2} = -EK. \end{aligned} \tag{2.13}$$

This formula shows that K can be expressed solely in terms of the coefficients of the first fundamental form and their derivatives.

□

Remark 2.25. Formula (2.13) is known as *Gauss formula*. If ϕ is an orthogonal parametrization, i.e. $F = 0$, calculations simplify quite a bit and the Gauss formula yields:

$$K = -\frac{1}{2\sqrt{EG}} \left\{ \left(\frac{E_v}{\sqrt{EG}} \right)_v + \left(\frac{G_u}{\sqrt{EG}} \right)_u \right\}. \tag{2.14}$$

Definition 2.26. A parametrization is said to be *isothermal* if $E = G = \lambda^2(u, v)$ and $F = 0$.

For isothermal parametrizations equation (2.14) simplifies even further to give:

$$K = -\frac{1}{\lambda^2} \Delta(\log \lambda), \tag{2.15}$$

where Δ is the Laplacian in (u, v)-coordinates.

CHAPTER 3

Critical points of length and area

In this chapter we consider the extremals of the functionals length and area. As you have seen in the Methods course the extremals of these functionals will satisfy the corresponding *Euler-Lagrange equations.* The solutions of these equations will be *geodesics* (length) and *minimal surfaces* (area). In the case of geodesics, the Euler-Lagrange equations are just ODEs and existence and uniqueness of solutions will be a fairly straightforward consequence of the results you have seen in Analysis II. In the case of minimal surfaces what we obtain is a PDE and existence and uniqueness is a fairly delicate problem.

3.1. Geodesics

Let $S \subset \mathbb{R}^3$ be a surface and p, q two points in S. Let $\Omega(p, q)$ be the set of all curves $\alpha : [0, 1] \to S$ with $\alpha(0) = p$ and $\alpha(1) = q$. Let $\ell : \Omega(p, q) \to \mathbb{R}$ be the functional given by the length of α, i.e.

$$\ell(\alpha) = \int_0^1 |\dot{\alpha}|\, dt.$$

A closely related functional, which is quite useful for variational purposes, is the *energy* $E : \Omega(p, q) \to \mathbb{R}$ given by

$$E(\alpha) = \frac{1}{2}\int_0^1 |\dot{\alpha}|^2\, dt.$$

One of the advantages of energy over length is that energy is sensitive to reparametrizations. The next simple exercise, which is a consequence of the Cauchy-Schwarz inequality, describes the relationship between the two functionals.

Exercise 3.1. $\ell(\alpha) \leq \sqrt{2E(\alpha)}$ with equality if and only if α is parametrized by a parameter proportional to arc-length.

Thus a curve minimizing E will have parameter proportional to arc-length.

Let us try to find the extremals of E. For this consider a smooth 1-parameter family of curves $\alpha_s \in \Omega(p, q)$ for $s \in (-\varepsilon, \varepsilon)$ with $\alpha_0 = \alpha$. Let $E(s) := E(\alpha_s)$. We have

$$\left.\frac{dE}{ds}\right|_{s=0} = \int_0^1 \left\langle \frac{\partial}{\partial s}\frac{\partial \alpha_s}{\partial t}, \frac{\partial \alpha_s}{\partial t} \right\rangle dt.$$

Integrating by parts we find:

$$\left.\frac{dE}{ds}\right|_{s=0} = \langle W(1), \dot{\alpha}(1)\rangle - \langle W(0), \dot{\alpha}(0)\rangle - \int_0^1 \langle W(t)\, , \ddot{\alpha}(t)\rangle\, dt$$

where

$$W(t) := \left.\frac{\partial \alpha_s(t)}{\partial s}\right|_{s=0}.$$

Since $\alpha_s \in \Omega(p,q)$, $W(0) = W(1) = 0$ and thus:

$$\left.\frac{dE}{ds}\right|_{s=0} = -\int_0^1 \langle W(t)\,, \ddot{\alpha}(t)\rangle \; dt.$$

Note that for each $t \in [0,1]$, $W(t) \in T_{\alpha(t)}S$, so if α has the property that $\ddot{\alpha}(t)$ is orthogonal to $T_{\alpha(t)}S$ at the point $\alpha(t)$ for all t, then α is an extremal point of E. This motivates the following:

Definition 3.2. A curve $\alpha : I \to S$ is said to be a *geodesic* if for all $t \in I$, $\ddot{\alpha}(t)$ is orthogonal to $T_{\alpha(t)}S$ at the point $\alpha(t)$.

In other words we can say that α is a geodesic if its acceleration has no component in the direction of the surface. Thus, in physical terms, a geodesic describes the motion of a *free* particle on the surface and what makes the motion interesting is the bending of the surface.

3.2. Covariant derivative and parallel transport

Let $\alpha : I \to S$ be a curve. *A vector field* V *along* α is a smooth map $V : I \to \mathbb{R}^3$ such that for all t, $V(t) \in T_{\alpha(t)}S$.

Definition 3.3. The vector obtained by the normal projection of $\frac{dV}{dt}(t)$ onto the plane $T_{\alpha(t)}S$ is called the *covariant derivative* of V at t and is denoted by $\frac{DV}{dt}(t)$.

We can now say that a geodesic is a curve α for which $\frac{D\dot{\alpha}}{dt} = 0$.

Definition 3.4. A vector field V along a curve $\alpha : I \to S$ is said to be *parallel* if $\frac{DV}{dt}(t) = 0$ for every $t \in I$.

Proposition 3.5. *Let* V *and* W *be parallel vector fields along* $\alpha : I \to S$*. Then* $\langle W(t), V(t)\rangle$ *is constant.*

Proof.

$$\frac{d}{dt}\langle W(t), V(t)\rangle = \left\langle \frac{dW}{dt}(t), V(t)\right\rangle + \left\langle W(t), \frac{dV}{dt}(t)\right\rangle$$

But the last two terms are zero because of the definition of parallel vector field and the fact that V and W are tangent to S. □

This proposition implies that if α is a geodesic, then $|\dot{\alpha}|$ is constant, so geodesics are parametrized with parameter proportional to arc-length.

Let $\phi : U \to S$ be a parametrization and let $\alpha : I \to S$ be a curve such that $\alpha(I) \subset \phi(U)$. For some functions $u(t)$ and $v(t)$ we can write $\alpha(t) = \phi(u(t), v(t))$. Let V be a vector field along α. For some functions $a(t)$ and $b(t)$ we can write $V(t) = a(t)\phi_u + b(t)\phi_v$. We have

$$\frac{dV}{dt} = a(\phi_{uu}\dot{u} + \phi_{uv}\dot{v}) + b(\phi_{vu}\dot{u} + \phi_{vv}\dot{v}) + \dot{a}\phi_u + \dot{b}\phi_v.$$

Using equations (2.7), (2.8) and (2.10) and the definition of covariant derivative (just drop the normal component) we obtain:

$$\begin{aligned}\frac{DV}{dt} &= (\dot{a} + \Gamma^1_{11}a\dot{u} + \Gamma^1_{12}a\dot{v} + \Gamma^1_{12}b\dot{u} + \Gamma^1_{22}b\dot{v})\phi_u \\ &\quad + (\dot{b} + \Gamma^2_{11}a\dot{u} + \Gamma^2_{12}a\dot{v} + \Gamma^2_{12}b\dot{u} + \Gamma^2_{22}b\dot{v})\phi_v. \qquad (3.1)\end{aligned}$$

The last expression shows that the covariant derivative only depends on the first fundamental form, even though it was defined using the ambient space $\mathbb{R}^3$. From (3.1) one can also obtain easily the equations of geodesics in local coordinates. It suffices to set $a = \dot{u}$ and $b = \dot{v}$ which gives the pair of equations:

$$\ddot{u} + \Gamma^1_{11}(\dot{u})^2 + 2\Gamma^1_{12}\dot{u}\dot{v} + \Gamma^1_{22}(\dot{v})^2 = 0; \tag{3.2}$$

$$\ddot{v} + \Gamma^2_{11}(\dot{u})^2 + 2\Gamma^2_{12}\dot{u}\dot{v} + \Gamma^2_{22}(\dot{v})^2 = 0. \tag{3.3}$$

Note the nonlinearity of (3.2) and (3.3).

If we regard (3.1) as a linear ODE in a and b, we see that given $v_0 \in T_{\alpha(t_0)}S$, $t_0 \in I$, there exists a unique parallel vector field $V(t)$ along $\alpha(t)$ with $V(t_0) = v_0$. The vector $V(t_1)$, $t_1 \in I$, is called the *parallel transport* of v_0 along α at the point t_1.

Let $\alpha \in \Omega(p, q)$. Denote by $P : T_pS \to T_qS$ the map that assigns to each $v \in T_pS$ its parallel transport along α at q. The map P is a linear map (why?) and by Proposition 3.5, P is an isometry.

Finally (3.2) and (3.3) imply:

PROPOSITION 3.6. *Given a point $p \in S$ and a vector $v \in T_pS$, there exists an $\varepsilon > 0$ and a unique geodesic $\gamma : (-\varepsilon, \varepsilon) \to S$ such that $\gamma(0) = p$ and $\dot{\gamma}(0) = v$.*

3.3. Minimal surfaces

DEFINITION 3.7. A surface $S \subset \mathbb{R}^3$ is said to be *minimal* if its mean curvature vanishes everywhere.

Let $\phi : U \subset \mathbb{R}^2 \to S$ be a parametrization and let $D \subset U$ be a bounded domain with closure $\overline{D} \subset U$. Let $h : \overline{D} \to \mathbb{R}$ be a smooth function.

DEFINITION 3.8. The *normal variation* of $\phi(\overline{D})$ determined by h is the map $\rho : \overline{D} \times (-\varepsilon, \varepsilon) \to \mathbb{R}^3$ given by

$$\rho(u, v, t) = \phi(u, v) + th(u, v)N(u, v).$$

REMARK 3.9. If the closure of D has the property that its boundary is a nice set, e.g. a piecewise smooth 1-manifold, then one often demands the boundary to be fixed in the variation. This can be achieved by selecting h so that it vanishes on the boundary.

For each fixed $t \in (-\varepsilon, \varepsilon)$, consider the map $\rho^t : D \to \mathbb{R}^3$ given by $\rho^t(u, v) := \rho(u, v, t)$. Since $\rho^0 = \phi$, for ε small enough $\rho^t(D)$ is a smooth surface in $\mathbb{R}^3$ (this follows from ρ^t_u and ρ^t_v being linearly independent). Clearly

$$\rho^t_u = \phi_u + thN_u + th_uN,$$

$$\rho^t_v = \phi_v + thN_v + th_vN.$$

If we let E^t, F^t and G^t be the coefficients of the first fundamental form of $\rho^t(D)$ we get:

$$E^t = E + 2th\langle\phi_u, N_u\rangle + t^2h^2\langle N_u, N_u\rangle + t^2h_uh_u,$$

$$F^t = F + 2th\langle\phi_u, N_v\rangle + t^2h^2\langle N_u, N_v\rangle + t^2h_uh_v,$$

$$G^t = G + 2th\langle\phi_v, N_v\rangle + t^2h^2\langle N_v, N_v\rangle + t^2h_vh_v.$$

Using the definition of e, f, g and (2.6) we obtain

$$\begin{aligned} E^tG^t - (F^t)^2 &= EG - F^2 - 2th(Eg - 2Ff + Ge) + r \\ &= (EG - F^2)(1 - 4thH) + r \end{aligned}$$

where $\lim_{t\to 0} r/t = 0$. If we let $A(t)$ be the area of $\rho^t(D)$ we have:

$$\begin{aligned} A(t) &= \int_D \sqrt{E^tG^t - (F^t)^2}\, dudv \\ &= \int_D \sqrt{1 - 4thH + \bar{r}}\, \sqrt{EG - F^2}\, dudv, \end{aligned}$$

where $\bar{r} = r/(EG - F^2)$. Clearly $t \mapsto A(t)$ is smooth and

$$A'(0) = -\int_D 2hH\sqrt{EG - F^2}\, dudv. \tag{3.4}$$

Proposition 3.10. *$\phi(U)$ is minimal if and only if $A'(0) = 0$ for all bounded domains $D \subset U$ and all normal variations of $\phi(D)$.*

Proof. If H vanishes, (3.4) immediately gives $A'(0) = 0$. Suppose $H(q) \neq 0$ for some $q \in D$. Let $h : D \to \mathbb{R}$ be H. Then by (3.4), $A'(0) < 0$ for the variation determined by such h. □

Remark 3.11. The proposition explains the use of the term "minimal" for surfaces with vanishing mean curvature. It should be noted that the critical point of A may not actually be a minimum.

The *mean curvature vector* is $\mathbf{H} := HN$. A normal variation in the direction of HN always has $A'(0) < 0$ (provided H does not vanish).

Remark 3.12. Minimal surfaces are often associated with soap films. A physical argument shows that at regular points the mean curvature of the film must be zero.

3.4. The Weierstrass representation

Proposition 3.13. *Let $\phi : U \to \mathbb{R}^3$ be an isothermal parametrization (cf. Definition 2.26). Then*

$$\phi_{uu} + \phi_{vv} = 2\lambda^2\mathbf{H}.$$

Proof. Since ϕ is isothermal $E = G = \lambda^2$ and $F = 0$. By differentiating $\langle \phi_u, \phi_u\rangle = \langle \phi_v, \phi_v\rangle$ with respect to u we get

$$\langle \phi_{uu}, \phi_u\rangle = \langle \phi_{vu}, \phi_v\rangle$$

and since $\langle \phi_{vu}, \phi_v\rangle = -\langle \phi_u, \phi_{vv}\rangle$ we obtain

$$\langle \phi_{uu} + \phi_{vv}, \phi_u\rangle = 0.$$

Similarly

$$\langle \phi_{uu} + \phi_{vv}, \phi_v\rangle = 0.$$

It follows that $\phi_{uu} + \phi_{vv}$ is parallel to N. Using (2.6) we have

$$H = \frac{g + e}{2\lambda^2}.$$

Thus

$$2\lambda^2 H = g + e = \langle N, \phi_{uu} + \phi_{vv}\rangle;$$

hence $\phi_{uu} + \phi_{vv} = 2\lambda^2 \mathbf{H}$ as desired.

$\square$

The proposition immmediately implies:

COROLLARY 3.14. *Let* $\phi : U \to \mathbb{R}^3$ *be an isothermal parametrization. Then* $\phi(U)$ *is minimal if and only if* $\Delta\phi = 0$*, where* Δ *is the Laplacian in* (u, v)*-coordinates.*

EXAMPLE 3.15. *The catenoid* is given by

$$\phi(u, v) = (a \cosh v \cos u, a \cosh v \sin u, av),$$

for $u \in (0, 2\pi)$ and $v \in \mathbb{R}$ (together with an obvious additional parametrization). This surface is obtained by rotating the catenary $y = a \cosh(z/a)$ about the z-axis. An easy calculation shows that $E = G = a^2 \cosh^2 v$ and $F = 0$, so ϕ is isothermal. Moreover $\Delta\phi = 0$ and so the catenoid is a minimal surface. In fact, it is the only surface of revolution which is minimal (Can you prove this?).

EXAMPLE 3.16. *The helicoid* is given by

$$\phi(u, v) = (a \sinh v \cos u, a \sinh v \sin u, au),$$

for $u \in \mathbb{R}$ and $v \in \mathbb{R}$. As in the previous example an easy calculation shows that $E = G = a^2 \cosh^2 v$ and $F = 0$, so ϕ is isothermal and $\Delta\phi = 0$. Thus the helicoid is a minimal surface. In fact, it is only minimal surface, besides the plane, which is *ruled*, i.e. it can be obtained from a straight line sliding smoothly along a curve.

J.B. Meusnier in 1776 discovered that the helicoid and the catenoid were minimal surfaces. For a long time they were the only examples. The next proposition establishes a connection between minimal surfaces and holomorphic functions that will allow us to find more examples. Write $\phi(u, v) = (x(u, v), y(u, v), z(u, v))$.

PROPOSITION 3.17. *Consider the complex valued functions*

$$\varphi_1 := x_u - ix_v \quad \varphi_2 := y_u - iy_v \quad \varphi_3 := z_u - iz_v.$$

Then ϕ *is isothermal if and only if* $\varphi_1^2 + \varphi_2^2 + \varphi_3^2 \equiv 0$*. Suppose* ϕ *is isothermal. Then* ϕ *is minimal if and only if* φ_1*,* φ_2 *and* φ_3 *are holomorphic functions.*

PROOF. An easy calculation gives:

$$\varphi_1^2 + \varphi_2^2 + \varphi_3^2 = E - G - 2iF,$$

which immediately implies the first claim in the proposition. To see the second claim, consider the Cauchy-Riemann equations of φ_1:

$$x_{uv} = x_{vu},$$

$$x_{uu} = -x_{vv}.$$

The first equation is trivially satisfied (ϕ is smooth) and the second one is just $\Delta x = 0$.

$\square$

The next lemma will tell us the solutions of

$$\varphi_1^2 + \varphi_2^2 + \varphi_3^2 \equiv 0. \tag{3.5}$$

LEMMA 3.18. *Let D be a domain in the complex plane, $g(\zeta)$ an arbitrary meromorphic function in D and $f(\zeta)$ a holomorphic function in D having the property that at each point where $g(\zeta)$ has a pole of order k, $f(\zeta)$ has a zero of order at least $2k$. Then the functions*

$$\varphi_1 = \frac{1}{2}f(1-g^2), \quad \varphi_2 = \frac{i}{2}f(1+g^2), \quad \varphi_3 = fg \tag{3.6}$$

will be holomorphic in D and satisfy (3.5). Conversely, every triple of holomorphic functions satisfying (3.5) may be represented in the form (3.6), except for $\varphi_1 \equiv i\varphi_2$, $\varphi_3 \equiv 0$.

PROOF. A calculation shows that the functions given by (3.6) satisfy (3.5). Conversely, given a solution of (3.5), set

$$f := \varphi_1 - i\varphi_2, \quad g := \frac{\varphi_3}{\varphi_1 - i\varphi_2}. \tag{3.7}$$

We now write (3.5) in the form

$$(\varphi_1 - i\varphi_2)(\varphi_1 + i\varphi_2) = -\varphi_3^2 \tag{3.8}$$

to obtain:

$$\varphi_1 + i\varphi_2 = -\frac{\varphi_3^2}{\varphi_1 - i\varphi_2} = -fg^2. \tag{3.9}$$

Combining (3.7) and (3.9) gives (3.6). The condition relating the zeros of f with the poles of g must hold, otherwise by (3.9), $\varphi_1+i\varphi_2$ would fail to be holomorphic. This representation can only fail if the denominator in the formula for g in (3.7) vanishes identically. In this case, by (3.8) we have that $\varphi_3 \equiv 0$, which is the exceptional case mentioned.

□

Let ζ_0 be a fixed point in D, and let γ be a path in D connecting ζ_0 to $\zeta \in D$. If D is simply connected and φ_i is holomorphic, the integrals

$$\int_\gamma \varphi_i(\zeta)\, d\zeta$$

do not depend on the path γ and define holomorphic functions whose real parts are x, y and z. Hence if we take functions f and g as in Lemma 3.18 and D is simply connected, then the equations

$$x(u,v) = \Re\left(\int_{\zeta_0}^{(u,v)} \frac{1}{2}f(\zeta)(1-g^2(\zeta))\, d\zeta\right), \tag{3.10}$$

$$y(u,v) = \Re\left(\int_{\zeta_0}^{(u,v)} \frac{i}{2}f(\zeta)(1+g^2(\zeta))\, d\zeta\right), \tag{3.11}$$

$$z(u,v) = \Re\left(\int_{\zeta_0}^{(u,v)} f(\zeta)g(\zeta)\, d\zeta\right) \tag{3.12}$$

define a parametrization which gives rise to a minimal surface. The equations (3.10)-(3.12) are called the *Weierstrass representation* of the surface. Of course one has to worry about whether ϕ is an immersion or an embedding.

EXERCISE 3.19. Show that if ϕ is defined by the Weierstrass representation, then ϕ is an immersion if and only f vanishes only at the poles of g and the order of its zero at such a point is *exactly* twice the order of the pole of g.

The converse is also true: any simply connected minimal surface in $\mathbb{R}^3$ can be represented by formulas (3.10)–(3.12).

By appropriate choices of f and g one obtains various celebrated examples of minimal surfaces. More importantly, the Weierstrass representation allows one to obtain general theorems about minimal surfaces by using complex analysis.

EXAMPLE 3.20 (Enneper's surface 1864). The simplest choice of D, f and g is $D = \mathbb{C}$, $f(\zeta) = 1$ and $g(\zeta) = \zeta$. It gives the minimal immersion $\phi : \mathbb{C} \to \mathbb{R}^3$

$$\phi(u,v) = \frac{1}{2}(u - u^3/3 + uv^2, -v + v^3/3 - u^2v, u^2 - v^2).$$

Its Gaussian curvature is

$$K = -\frac{16}{(1+ \mid \zeta \mid^2)^4},$$

where $\zeta = u + iv$.

EXAMPLE 3.21 (Scherk's surface 1835). Let D be the open unit disk in the complex plane, $f(\zeta) = 4/(1-\zeta^4)$ and $g(\zeta) = \zeta$. This gives the immersion $\phi : D \to \mathbb{R}^3$

$$\phi(u,v) = \left(-\arg\frac{\zeta + i}{\zeta - i}, -\arg\frac{\zeta+1}{\zeta-1}, \log\left|\frac{\zeta^2+1}{\zeta^2-1}\right|\right).$$

From the expressions of x, y and z it is easily seen that

$$z = \log\frac{\cos y}{\cos x}.$$

EXERCISE 3.22. Find D, f and g for the catenoid and the helicoid.

There are great web sites about minimal surfaces. Take a look at this one. It contains beautiful pictures of all the surfaces we have seen and much more.

3.5. The meaning of g in the Weierstrass representation

Let $\pi : S^2 \to \mathbb{C} \cup \{\infty\}$ be the stereographic projection. Recall that

$$\pi^{-1}(z) = \left(\frac{2\Re z}{1+ \mid z \mid^2}, \frac{2\Im z}{1+ \mid z \mid^2}, \frac{\mid z \mid^2 -1}{\mid z \mid^2 +1}\right). \tag{3.13}$$

Let ϕ be a parametrization given by formulas (3.10)-(3.12). Clearly

$$\phi_u - i\phi_v = (\varphi_1, \varphi_2, \varphi_3)$$

and hence

$$\lambda^2 = E = G = \frac{1}{2}\sum \mid \varphi_k \mid^2 = \left(\frac{\mid f \mid (1+ \mid g \mid^2)}{2}\right)^2.$$

Furthermore

$$\phi_u \wedge \phi_v = \Im\{(\varphi_2\overline{\varphi}_3, \varphi_3\overline{\varphi}_1, \varphi_1\overline{\varphi}_2)\}$$

and using the expressions of φ_i in terms of f and g we derive:

$$\phi_u \wedge \phi_v = \frac{\mid f \mid^2 (1+ \mid g \mid^2)}{4}(2\Re g, 2\Im g, \mid g \mid^2 -1).$$

Note also that

$$| \phi_u \wedge \phi_v |= \left(\frac{| f | (1+ | g |^2)}{2} \right)^2 = \lambda^2.$$

Hence, if $N : D \to S^2$ is the Gauss map, we obtain:

$$N = \left(\frac{2\Re g}{1+ | g |^2}, \frac{2\Im g}{1+ | g |^2}, \frac{| g |^2 -1}{| g |^2 +1} \right). \tag{3.14}$$

From (3.13) and (3.14) we see that

$$\pi \circ N = g. \tag{3.15}$$

If you are attending the Riemann surfaces course you will notice that (3.15) implies right away the following lemma:

Lemma 3.23. *The Gauss map $N : D \to S^2$ is a holomorphic map.*

Exercise 3.24. Show that the Gaussian curvature is given by

$$K = - \left(\frac{4 | g' |}{| f | (1+ | g |^2)^2} \right)^2 .$$

Show that either $K \equiv 0$ or its zeros are isolated.

Here is a beautiful application of (3.15):

Theorem 3.25. *If $D = \mathbb{C}$, then either the surface lies on a plane, or else the Gauss map takes all directions with a most two exceptions.*

Proof. To the surface we associate the function $g(\zeta)$ which fails to be defined only if $\varphi_1 \equiv i\varphi_2$, $\varphi_3 \equiv 0$. But in this case z is constant and the surface lies in a plane. Otherwise $g(\zeta)$ is a meromorphic function on the entire complex plane, and by Picard's theorem it either takes all the values with at most two exceptions, or else is constant. By (3.15) the same alternative applies to the Gauss map, and in the latter case the surface lies on a plane. $\square$

In the catenoid, the Gauss map misses exactly two points. The following remarkable result is stated without proof (its proof is a more elaborate application of (3.15):

Theorem 3.26 (Osserman 1960). *Let S be a minimal surface in $\mathbb{R}^3$ which is not a plane and is a closed subset of $\mathbb{R}^3$. Then the image of the Gauss map is dense in S^2.*

In fact, subsequent work (Xavier 1981, Fujimoto 1988), shows that if S is a minimal surface in $\mathbb{R}^3$ which is not a plane and is a closed subset of $\mathbb{R}^3$, then the image of the Gauss map can omit at most four directions.

CHAPTER 4

Global Riemannian geometry

4.1. The exponential map and geodesic polar coordinates

Let $S \subset \mathbb{R}^3$. By Proposition (3.6) we know that given a point $p \in S$ and a vector $v \in T_pS$, there exists an $\varepsilon > 0$ and a unique geodesic $\gamma : (-\varepsilon, \varepsilon) \to S$ such that $\gamma(0) = p$ and $\dot\gamma(0) = v$. To indicate the dependence on the vector v, let us write $\gamma_v = \gamma$. Here is a simple exercise:

Exercise 4.1. Prove that if the geodesic $\gamma_v(t)$ is defined for $t \in (-\varepsilon, \varepsilon)$, then the geodesic $\gamma_{\lambda v}(t)$, $\lambda \in \mathbb{R}$, $\lambda \neq 0$, is defined for $t \in (-\varepsilon/\lambda, \varepsilon/\lambda)$, and $\gamma_{\lambda v}(t) = \gamma_v(\lambda t)$.

Definition 4.2. If $v \in T_pS$, $v \neq 0$, is such that $\gamma_{v/|v|}(|v|) = \gamma_v(1)$ is defined, we set

$$\exp_p(v) = \gamma_v(1) \ \text{ and } \ \exp_p(0) = p.$$

The important point is that $\exp_p$ is always defined and differentiable in some ball B_ε with center at the origin in T_pS. We will not prove this result here, but you should know that it is a fairly straightforward application of a well known theorem in ODEs about smooth dependence of solutions on initial conditions (as far as I know this result is not proved in the Tripos). The map $\exp_p : B_\varepsilon \to S$ is called *the exponential map at p*. If the surface S is a closed subset of $\mathbb{R}^3$, then one can show that $\exp_p$ is defined on all of T_pS and geodesics are defined for all $t \in \mathbb{R}$. The exponential map also depends smoothly on p.

We can actually prove the following important complement:

Proposition 4.3. *$\exp_p : B_\varepsilon \to S$ is a diffeomorphism onto its image in a neighbourhood $U \subset B_\varepsilon$ of the origin in T_pS.*

Proof. By the inverse function theorem it suffices to show that the differential $d(\exp_p)_0$ is nonsingular. Let $\alpha(t) := tv$, $v \in T_pS$. The curve $(\exp_p \circ \alpha)(t) = \exp_p(tv) = \gamma_v(t)$ has at $t = 0$ the tangent vector v. Thus $d(\exp_p)_0(v) = v$. □

The last proposition implies that the exponential map $\exp_p$ when restricted to a small ball is actually a parametrization of S around p. The set $V := \exp_p(U)$ is called a *normal neighbourhood of p.*

Since $\exp_p$ maps lines through the origin to geodesics in S, it gives in fact a very useful and "geometrically friendly" parametrization. For example, if we choose in T_pS cartesian coordinates, and we look at the coefficients of the first fundamental form in these coordinates, then we see right away that $E(p) = G(p) = 1$ and $F(p) = 0$, so the metric at p looks like the Euclidean metric.

If we now choose polar coordinates (r, θ) on T_pS we obtain the so called *geodesic polar coordinates* on S. Recall that polar coordinates in the plane are not defined in the closed half line ℓ that corresponds to $\theta = 0$. Set $L = \exp_p(\ell)$. Since

$\exp_p : U - \ell \to V - L$ is still a diffeomorphism, we may parametrize $V - L$ by the coordinates (r, θ). However for any (r, θ) we may still define

$$\phi(r,\theta) := \exp_p(r(\cos\theta\, e_1 + \sin\theta\, e_2)) = \exp_p(rv(\theta)) = \gamma_{v(\theta)}(r), \tag{4.1}$$

where $v(\theta) = \cos\theta\, e_1 + \sin\theta\, e_2$.

DEFINITION 4.4. The images under $\phi : U \to V$ of circles in U centered at the origin are called *geodesic circles* and the images under ϕ of the lines through the origin are called *radial geodesics*.

The coefficients of the first fundamental form and the Gaussian curvature take a very simple form in geodesic polar coordinates as the next proposition shows.

PROPOSITION 4.5. *The coefficients $E(r,\theta)$, $F(r,\theta)$ and $G(r,\theta)$ satisfy:*

$$E = 1, \qquad F = 0, \qquad G(0,\theta) = 0, \qquad (\sqrt{G})_r(0,\theta) = 1.$$

Moreover, the Gaussian curvature K can be written as:

$$K = -\frac{(\sqrt{G})_{rr}}{\sqrt{G}}.$$

PROOF. From (4.1) we see that $\phi_r = \dot\gamma_{v(\theta)}(r)$ and thus $E = 1$ since $v(\theta)$ has norm one and geodesics travel at constant speed. Let $w = \frac{dv}{d\theta}$. Also from (4.1) we see that

$$\phi_\theta = d(\exp_p)_{rv}(rw) = rd(\exp_p)_{rv}(w).$$

Hence

$$F = r\langle \dot\gamma_v(r), d(\exp_p)_{rv}(w)\rangle,$$

$$G = r^2 |d(\exp_p)_{rv}(w)|^2.$$

Obviously $F(0,\theta) = 0$ and from the last equality we see that

$$(\sqrt{G})_r(0,\theta) = |d(\exp_p)_0(w)| = |w| = 1.$$

We now compute

$$\begin{aligned} F_r &= \langle \phi_{rr}, \phi_\theta\rangle + \langle \phi_r, \phi_{\theta r}\rangle \\ &= \langle \phi_r, \phi_{\theta r}\rangle = \frac{1}{2}\frac{\partial}{\partial\theta}\langle \phi_r, \phi_r\rangle \\ &= \frac{1}{2}E_\theta = 0, \end{aligned}$$

where we used that $\phi_r = \gamma_v$ is a geodesic. It follows that $F = 0$ everywhere since $F(0,\theta) = 0$.

Finally the expression for K follows right away from (2.14).

□

4.2. The Gauss-Bonnet theorem

4.2.1. Geodesic curvature.

DEFINITION 4.6. Let W be a differentiable field of unit vectors along a curve $\alpha : I \to S$ on an oriented surface S. Since $W(t)$ has norm 1, $(dW/dt)(t)$ is normal to $W(t)$ and therefore

$$\frac{DW}{dt} = \lambda(N \wedge W).$$

The real number $\lambda = \lambda(t)$, denoted by $[DW/dt]$, is called the *algebraic value of the covariant derivative of W at t.*

REMARK 4.7. Note that the sign of $[DW/dt]$ depends on the orientation of S. Also note that

$$\left[\frac{DW}{dt}\right] = \left\langle \frac{dW}{dt}, N \wedge W \right\rangle.$$

DEFINITION 4.8. Let $\alpha : I \to S$ be a regular curve parametrized by arc-length. The algebraic value of the covariant derivative

$$k_g(s) := \left[\frac{D\dot{\alpha}}{dt}\right] = \langle \ddot{\alpha}, N \wedge \dot{\alpha} \rangle$$

is called *the geodesic curvature* of α at $\alpha(s)$.

REMARK 4.9. Note that k_g changes sign if we change the orientation of S. Also note that α is a geodesic if and only if its geodesic curvature is zero.

Let V and W be two differentiable unit vector fields along a curve $\alpha : I \to S$. Let iV be the unique unit vector field along α such that for every $t \in I$, $\{V(t), iV(t)\}$ is a positively oriented orthonormal basis of $T_{\alpha(t)}S$. Thus we can write

$$W(t) = a(t)V(t) + b(t)iV(t)$$

where a and b are smooth functions in I and $a^2 + b^2 = 1$.

EXERCISE 4.10. Let a and b be smooth functions on I with $a^2 + b^2 = 1$ and φ_0 be such that $a(t_0) = \cos\varphi_0$, $b(t_0) = \sin\varphi_0$. Then, the smooth function

$$\varphi(t) := \varphi_0 + \int_{t_0}^{t} (a\dot{b} - b\dot{a})\, dt$$

is such that $\cos\varphi(t) = a(t)$, $\sin\varphi(t) = b(t)$, $t \in I$, and $\varphi(t_0) = \varphi_0$.

The smooth $\varphi(t)$ given by the exercise will be called a *a smooth determination of the angle from V to W.*

PROPOSITION 4.11. *Let V and W be two smooth unit vector fields along a curve $\alpha : I \to S$. Then*

$$\left[\frac{DW}{dt}\right] - \left[\frac{DV}{dt}\right] = \frac{d\varphi}{dt},$$

where φ is a smooth determination of the angle from V to W.

PROOF. By definition

$$\left[\frac{DW}{dt}\right] = \langle \dot{W}, N \wedge W \rangle,$$

$$\left[\frac{DV}{dt}\right] = \langle \dot{V}, N \wedge V \rangle = \langle \dot{V}, iV \rangle.$$

Write $W = \cos\varphi\, V + \sin\varphi\, iV$ and differentiate to obtain:

$$\dot{W} = \dot{\varphi}(-\sin\varphi\, V + \cos\varphi\, iV) + \cos\varphi\, \dot{V} + \sin\varphi \frac{d}{dt}(iV).$$

Also $N \wedge W = \cos\varphi iV - \sin\varphi V$. Thus

$$\begin{aligned}\left[\frac{DW}{dt}\right] &= \dot{\varphi} + \langle -\sin\varphi\, V + \cos\varphi\, iV, \cos\varphi\, \dot{V} + \sin\varphi \frac{d}{dt}(iV)\rangle \\ &= \dot{\varphi} + \cos^2\varphi\, \langle iV, \dot{V}\rangle - \sin^2\varphi\, \langle V, \frac{d}{dt}(iV)\rangle \\ &= \dot{\varphi} + \left[\frac{DV}{dt}\right],\end{aligned}$$

where we used that $\langle iV, \dot{V}\rangle + \langle V, \frac{d}{dt}(iV)\rangle = 0$ (differentiate $\langle V, iV\rangle = 0$).

□

REMARK 4.12. Let $\alpha : I \to S$ be a curve parametrized by arc-length and let $V(s)$ be a parallel unit vector field along α. Let φ be a smooth determination of the angle from V to $\dot{\alpha}$. Then by the proposition

$$k_g(s) = \frac{d\varphi}{ds}.$$

If the curve is a plane curve, we can take as $V(t)$ a fixed direction in the plane and we see that k_g reduces to the usual curvature of α.

PROPOSITION 4.13. *Let $\phi(u, v)$ be an orthogonal parametrization (i.e. $F = 0$) of an oriented surface S which is compatible with the orientation. Let $W(t)$ be a smooth unit vector field along the curve $\phi(u(t), v(t))$. Then*

$$\left[\frac{DW}{dt}\right] = \frac{1}{2\sqrt{EG}} \{G_u \dot{v} - E_v \dot{u}\} + \frac{d\varphi}{dt},$$

where φ is the angle from ϕ_u to W in the given orientation. In particular, if $\alpha : I \to \phi(U)$ is a curve parametrized by arc-length,

$$k_g(s) = \frac{1}{2\sqrt{EG}} \left\{ G_u \frac{dv}{ds} - E_v \frac{du}{ds} \right\} + \frac{d\varphi}{ds},$$

where φ is the angle from ϕ_u to $\dot{\alpha}$ in the given orientation.

PROOF. Let $e_1 = \phi_u/\sqrt{E}$ and $e_2 = \phi_v/\sqrt{G}$. $\{e_1, e_2\}$ is a positively oriented orthonormal basis of the tangent plane. By the previous proposition

$$\left[\frac{DW}{dt}\right] = \left[\frac{De_1}{dt}\right] + \dot{\varphi}.$$

We compute

$$\begin{aligned}\left[\frac{De_1}{dt}\right] &= \langle \dot{e}_1, N \wedge e_1 \rangle = \langle \dot{e}_1, e_2 \rangle \\ &= \langle (e_1)_u \dot{u} + (e_1)_v \dot{v}, e_2 \rangle \\ &= \langle (e_1)_u, e_2 \rangle \dot{u} + \langle (e_1)_v, e_2 \rangle \dot{v}.\end{aligned}$$

Note ($F = 0$):

$$\langle (e_1)_u, e_2\rangle = \langle (\phi_u/\sqrt{E})_u, \phi_v/\sqrt{G}\rangle = \frac{1}{\sqrt{EG}}\langle \phi_{uu}, \phi_v\rangle.$$

Also $F = 0$ implies $\langle \phi_{uu}, \phi_v\rangle = -\langle \phi_u, \phi_{uv}\rangle = -E_v/2$. Thus

$$\langle (e_1)_u, e_2\rangle = -\frac{E_v}{2\sqrt{EG}}.$$

Similarly

$$\langle (e_1)_v, e_2\rangle = \frac{G_u}{2\sqrt{EG}}.$$

□

4.2.2. Gauss theorem for small geodesic triangles.

THEOREM 4.14 (Gauss 1827). *Let T be a geodesic triangle (i.e. its sides are segments of geodesics) on a surface S. Suppose T is small enough so that it is contained in a normal neighbourhood of one of its vertices. Then*

$$\int_T K\, dA = \left(\sum_{i=1}^{3} \alpha_i\right) - \pi,$$

where K is the Gaussian curvature of S, and $0 < \alpha_i < \pi$, $i = 1, 2, 3$, are the internal angles of the geodesic triangle T.

PROOF. This was Gauss' original proof. Consider geodesic polar coordinates (r, θ) centered at one of the vertices of T. Suppose also that one of the sides of T corresponds to $\theta = 0$. Let the two other sides be given by $\theta = \theta_0$ and a geodesic segment γ. Observe that γ is given in these coordinates by $r = h(\theta)$ (why?). Then

$$\int_T K\, dA = \int_T K\sqrt{G}\, dr d\theta = \int_0^{\theta_0}\left(\lim_{\varepsilon\to 0}\int_\varepsilon^{h(\theta)} K\sqrt{G}\, dr\right) d\theta.$$

Since (cf. Proposition 4.5) $K\sqrt{G} = -(\sqrt{G})_{rr}$ and $\lim_{r\to 0}(\sqrt{G})_r = 1$, we have that

$$\left(\lim_{\varepsilon\to 0}\int_\varepsilon^{h(\theta)} K\sqrt{G}\, dr\right) = 1 - (\sqrt{G})_r(h(\theta), \theta).$$

Suppose the geodesic $\gamma(r(s), \theta(s))$ makes an angle $\varphi(s)$ with the curves $\theta =$const. in the orientation determined by the coordinates (r, θ). Then, since γ is a geodesic, Proposition 4.13 tell us that

$$(\sqrt{G})_r\frac{d\theta}{ds} + \frac{d\varphi}{ds} = 0.$$

Hence

$$\int_T K\, dA = \int_0^{\theta_0} d\theta + \int_{\varphi(0)}^{\varphi(s_0)} d\varphi = \left(\sum_{i=1}^{3}\alpha_i\right) - \pi.$$

□

Let us define the *excess of T* as $e(T) := \left(\sum_{i=1}^3 \alpha_i\right) - \pi$. Note that if K is never zero in T, $\int_T K\, dA$ is, up to a sign, the area of the image $N(T)$ of T by the Gauss map $N : S \to S^2$. Thus we can rephrase the last result by saying that the excess of a geodesic triangle T equals the area of its spherical image $N(T)$ with the appropriate sign. This was the way Gauss himself phrased his theorem.

4.2.3. Triangulations and Euler characteristic.

DEFINITION 4.15. Let S be a compact surface. *A triangulation* of S consists of a finite number of closed subsets $\{T_1, \ldots, T_n\}$ that cover S, such that each T_i is homeomorphic to an Euclidean triangle in the plane. The subsets T_i are called *triangles*. Moreover, we require that any two distinct triangles either be disjoint, have a single vertex in common, or have an entire edge in common.

It is a theorem that every compact surface (or compact manifold) admits a triangulation. In fact one can choose the T_i's to be diffeomorphic to Euclidean triangles. By considering barycentric subdivisions of the triangles we can produce triangulations with very small triangles. Smooth triangulations exist even if S has piecewise smooth boundary. Finally, one can actually choose the edges to be geodesic segments, so that the triangles will be geodesic triangles.

Given a triangulation of S, let F be the number of triangles (faces), let E be the number of edges and let V be the number of vertices. The number

$$\chi(S) := F - E + V$$

is called *the Euler characteristic* of the surface. It is an important topological result that $\chi(S)$ does not depend on the triangulation, so that $\chi(S)$ is a topological invariant of the surface. You will meet this invariant in the Algebraic Topology course and the Riemann Surfaces course.

Compact orientable surfaces can be completely classified using χ. Any such S will be diffeomorphic to a sphere with g handles and $\chi(S) = 2 - 2g$. The sphere has $\chi = 2$ (no handle), the torus (sphere with one handle) has $\chi = 0$ and the double torus (2 handles) has $\chi = -2$. The number g is called *the genus* of S.

REMARK 4.16. Compact boundaryless surfaces in $\mathbb{R}^3$ are orientable.

4.2.4. The global Gauss-Bonnet theorem.

THEOREM 4.17 (Global Gauss-Bonnet theorem). *Let S be a compact surface with empty boundary; then*

$$\int_S K\, dA = 2\pi\chi(S).$$

PROOF. Consider a triangulation with geodesic triangles T_i, $i = 1, \ldots, F$. We will assume that each T_i is small enough so that it is contained in a normal neighbourhood of one of its vertices.

Let α_i, β_i, γ_i be the interior angles of T_i. By Theorem 4.14

$$\int_{T_i} K\, dA = \alpha_i + \beta_i + \gamma_i - \pi.$$

Summation over i yields

$$\int_S K\, dA = \sum_{i=1}^{F} (\alpha_i + \beta_i + \gamma_i) - \pi F.$$

Note that the sum of the angles at every vertex of the triangulation is 2π, hence

$$\sum_{i=1}^{F} (\alpha_i + \beta_i + \gamma_i) = 2\pi V.$$

On the other hand, every edge of the triangulation belongs to two triangles, which implies

$$2E = 3F.$$

Putting everything together we obtain:

$$\int_S K\, dA = \pi(2V - F) = \pi(2V - 2E + 2F) = 2\pi\chi(S).$$

$\square$

Now we would like to generalize Theorem 4.14 to allow for boundary curves which are not necessarily geodesics. This will bring in an extra term involving the geodesic curvature of the boundary curves (this was actually Bonnet's contribution).

THEOREM 4.18 (Gauss-Bonnet Theorem (local)). *Let $\phi : U \to S$ be an orthogonal parametrization of an oriented surface S, where U is a disk in $\mathbb{R}^2$ and ϕ is compatible with the orientation of S. Let $\alpha : I \to \phi(U)$ be a smooth simple closed curve enclosing a domain R. Assume that α is positively oriented and parametrized by arc-length. Then*

$$\int_I k_g(s)\, ds + \int_R K\, dA = 2\pi,$$

where k_g is the geodesic curvature of α.

PROOF. By Proposition 4.13 we may write

$$k_g(s) = \frac{1}{2\sqrt{EG}} \left\{ G_u \frac{dv}{ds} - E_v \frac{du}{ds} \right\} + \frac{d\varphi}{ds},$$

where φ is the angle from ϕ_u to $\dot{\alpha}$ in the given orientation. Set $I = [0, l]$. Integrating the equality we obtain

$$\int_I k_g(s)\, ds = \int_I \frac{1}{2\sqrt{EG}} \left\{ G_u \frac{dv}{ds} - E_v \frac{du}{ds} \right\} ds + \varphi(l) - \varphi(0).$$

By Green's theorem

$$\int_I k_g(s)\, ds = \int_{\phi^{-1}(R)} \left\{ \left(\frac{G_u}{2\sqrt{EG}} \right)_u + \left(\frac{E_v}{2\sqrt{EG}} \right)_v \right\} dudv + \varphi(l) - \varphi(0).$$

Using (2.14) gives:

$$\int_I k_g(s)\, ds = -\int_{\phi^{-1}(R)} K\sqrt{EG}\, dudv + \varphi(l) - \varphi(0).$$

Since $\varphi(l) - \varphi(0) = 2\pi$ (Theorem of the turning tangents) we obtain :

$$\int_I k_g(s)\, ds = -\int_R K\, dA + 2\pi.$$

$\square$

REMARK 4.19. If α is piecewise smooth, then φ jumps by the external angle θ_i at each vertex, so the integral of $\frac{d\varphi}{ds}$ which is 2π in the proof of the theorem is replaced by

$$2\pi - \sum_i \theta_i,$$

so one gets:

$$\int_I k_g(s)\,ds + \int_R K\,dA = \sum_i \delta_i - (p-2)\pi$$

where δ_i are the internal angles at the vertices and p is the number of vertices. If there are 3 vertices and α is a geodesic on each smooth portion, we recover Theorem 4.14.

REMARK 4.20. A very pretty proof of the theorem of the turning tangents due to Hopf (1935) may be found in do Carmo's book or in Volume II of Spivak. It is based on the invariance under homotopies of the degree of a continuous map from S^1 to S^1.

Now we can state:

THEOREM 4.21 (Global Gauss-Bonnet Theorem including boundary terms). *Let $R \subset S$ be a domain of an oriented surface. Suppose R has compact closure and its boundary consists of n piecewise smooth simple closed curves $\alpha_i : I_i \to S$, $i = 1, \dots, n$ whose images do not intersect. Suppose the α_i's are parametrized by arc length and are positively oriented. Let θ_i, $i = 1, \dots, p$, be the external angles of the vertices of these curves. Then*

$$\sum_{i=1}^{n} \int_{I_i} k_g(s)\,ds + \int_R K\,dA + \sum_{i=1}^{p} \theta_i = 2\pi\chi(R).$$

PROOF. I will not do this proof in lectures. The idea is the same as the proof of Theorem 4.17: consider a triangulation, apply the local theorem and put everything together. An important point to note is that each "interior edge" of the triangulation is described twice in opposite orientations, hence from all the contributions of the integral of k_g only the boundary ones survive. Writing a full detailed proof is an excellent exercise (if you get stuck pick on do Carmo's book). □

4.3. Applications

THEOREM 4.22. *A compact oriented surface S with positive Gaussian curvature is diffeomorphic to S^2. Moreover, if there exist two simple closed geodesics, they must intersect.*

PROOF. The Global Gauss-Bonnet tells us right away that $\chi(S) > 0$ and since the sphere is the only orientable surface with positive Euler characteristic the first claim in the theorem follows.

To see the second claim, suppose that there are two simple closed geodesics γ_1 and γ_2 whose images do not intersect. Then the set formed by their images is the boundary of a domain R with $\chi(R) = 0$. By the Gauss-Bonnet theorem with boundary terms

$$\int_R K\,dA = 0,$$

which is a contradiction since $K > 0$.

□

In fact, the first claim in the theorem can be proved without using Gauss-Bonnet and we can actually give an explicit diffeomorphism. If $K > 0$ everywhere, every point in S^2 is a regular value of the Gauss map $N : S \to S^2$ (why?). Hence by the Stacks of records theorem 1.15 N is a covering map (making contact again with Algebraic Topology). Since S^2 is simply connected, N must be a bijection and therefore a diffeomorphism.

THEOREM 4.23. *Let S be a surface homeomorphic to a cylinder. Suppose S has negative Gaussian curvature. Then S has at most one simple closed geodesic.*

PROOF. Example sheet 4. □

4.4. Fenchel's theorem

DEFINITION 4.24. A plane regular closed curve $\alpha : [0, L] \to \mathbb{R}^2$ is *convex* if, for each $t \in [0, L]$, the curve lies in one of the closed half-planes determined by the tangent line at t.

DEFINITION 4.25. Let $\alpha : [0, L] \to \mathbb{R}^3$ be a curve parametrized by arc-length and let $k(s) \geq 0$ be its curvature. *The total curvature of* α is the number

$$\int_0^L k(s)\, ds.$$

THEOREM 4.26 (Fenchel's theorem). *The total curvature of a simple closed curve is $\geq 2\pi$, and equality holds if and only if the curve is a plane convex curve.*

We will not prove the convexity statement and we will assume from now on that the curvature is positive everywhere (or equivalently, that it is nowhere vanishing).

The proof of this theorem (and the next) will be given as a sequence of guided exercises with hints. The two main ingredients are the *tube* T around α and the *area formula*.

Recall from the example sheets that the tube is given by:

$$\phi(s, v) = \alpha(s) + r(n \cos v + b \sin v),$$

where r is sufficiently small, n is the normal to α, b the binormal and $s \in [0, L]$, $v \in [0, 2\pi]$ (α is parametrized by arc-length and has length L).

The area formula applied to the Gauss map $N : S \to S^2$ of a compact surface S (perhaps with non-empty boundary) states that

$$\int_S |K|\, dA = \int_{S^2} \#N^{-1}(y)\, dy. \tag{4.2}$$

Try to prove it using the Stack of records theorem 1.15; it is a very challenging exercise!

We now start the proof.

(1) We orient the tube T using ϕ. Show that the Gauss map is given by $N = -(n \cos v + b \sin v)$.
(2) Show that for the tube we have $EG - F^2 = r^2(1 - rk \cos v)^2$ and
$$K = \frac{-k \cos v}{r(1 - kr \cos v)}.$$
Conclude that $K = 0$ iff the line through b is orthogonal to the tube.
(3) Let $R := \{p \in T : \ K(p) \geq 0\}$. Show that
$$\int_R K\, dA = 2 \int_0^L k(s)\, ds.$$
(4) Show that the Gauss map $N : R \to S^2$ is surjective.

(5) Use the previous item and the area formula (4.2) applied to R to conclude that

$$2\int_0^L k(s)\,ds = \int_R K\,dA = \int_{S^2} \#\left(N|_R\right)^{-1}(y)\,dy \geq 4\pi.$$

(6) We now need to show that if

$$\int_R K\,dA = 4\pi, \tag{4.3}$$

then α is a plane curve. (If you get stuck with this part look at Spivak's Volume III, page 289.) Show first that if equality in (4.3) holds then given a point $p \in T$ with $K(p) > 0$, the tube T must lie on one side of the affine tangent plane at p. Now fix a circle $s = s_0$ and show that T lies between the parallel affine tangent planes P_1 and P_2 determined by the two points on the circle $s = s_0$ where the Gaussian curvature vanishes. The final step consists in showing that α is completely contained in the plane midway between P_1 and P_2. For this argue by contradiction and consider a point $\alpha(s_1)$ on α which is at a maximum distance from P (but not in P). Use that the tangent vector $\dot{\alpha}(s_1)$ must be parallel to P to derive the contradiction.

(7) Complete the proof of Fenchel's theorem by showing that if α is a plane curve, then its total curvature is 2π (apply Gauss-Bonnet to the flat region enclosed by α).

4.5. The Fáry-Milnor theorem

Think of S^1 as the standard unit circle in $\mathbb{R}^2$. Let $\alpha : S^1 \to \mathbb{R}^3$ be a continuous simple curve (i.e. α is injective). We say that α is *unknotted* if there exists a homotopy $H : S^1 \times [0,1] \to \mathbb{R}^3$ such that if we let $\alpha_t(s) = H(s,t)$, then $\alpha_0(s) = (s,0) \in \mathbb{R}^2 \times \mathbb{R} = \mathbb{R}^3$, $\alpha_1 = \alpha$ and α_t is a simple closed curve for all $t \in [0,1]$. Such a homotopy is called an *isotopy*.

Alternatively we can say that α is unknotted if $\alpha(S^1)$ is the boundary of an embedded disk.

If α is not unknotted, then it is said to be *knotted*.

Theorem 4.27 (Fáry-Milnor). *The total curvature of a knotted simple closed curve is greater than 4π.*

Proof. We will prove that the total curvature is $\geq 4\pi$ if α is knotted (Fáry, 1949). Milnor (1950) proved in fact the sharper statement that the total curvature is $> 4\pi$ and that the infimum over all simple closed curves is 4π. As in the case of Fenchel's theorem we will assume that the curvature k is nowhere vanishing (so we can use the tube as defined).

(1) Use the area formula (4.2) to show that

$$4\int_0^L k(s)\,ds = \int_{S^2} \#N^{-1}(y)\,dy.$$

(2) Suppose that $\int_0^L k(s)\,ds < 4\pi$. Using that $b([0,L]) \subset S^2$ and $-b([0,L]) \subset S^2$ have measure zero show that there exists $y \in S^2 \setminus \pm b([0,L])$ for which $\#N^{-1}(y) \leq 3$.

(3) For y as in the previous item, let $h(s) := \langle \alpha(s), y \rangle$. Show that h has at most 3 critical points which must all be strict local maxima or minima. Conclude that h has only two critical points which are the global maximum and minimum of h.
(4) Without loss of generality we may suppose $y = (0, 0, 1)$. Think of h as a height function. Show that α consists of two arcs joining the lowest and highest points, each arc having monotonically increasing heights.
(5) Each plane orthogonal to y and with height between the lowest and highest points intersects α in two points. Join each such pair of points by a line segment to obtain an embedded disk whose boundary is α. Conclude that α is unknotted.

□

If you are interested in seeing how the hypothesis of having nowhere vanishing curvature may be removed in Fenchel's theorem and the Fáry-Milnor theorem take a look at Spivak's volume III.

In 1953 Milnor (Math. Scand. **1** (1953) 289–296) proved the inequality

$$\int_0^L \sqrt{k^2 + \tau^2}\, ds \geq 4\pi$$

for closed space curves with nowhere zero torsion τ. Much later Burt Totaro rediscovered this result (Internat. J. Math. **1** (1990) 109–117). (I thank Burt for pointing this out to me.)