Analytic and Differential geometry

Mikhail G. Katz*

Department of Mathematics, Bar Ilan University, Ramat Gan 52900 Israel

E-mail address: katzmik@math.biu.ac.il

*Supported by the Israel Science Foundation (grants no. 620/00-10.0 and 84/03).

Abstract. We start with analytic geometry and the theory of conic sections. Then we treat the classical topics in differential geometry such as the geodesic equation and Gaussian curvature. Then we prove Gauss's *theorema egregium* and introduce the abstract viewpoint of modern differential geometry.

Contents

CHAPTER 1

Analytic geometry

These notes are based on a course taught at Bar Ilan University. After dealing with classical geometric preliminaries including the *theorema egregium* of Gauss, we present new geometric inequalities on Riemann surfaces, as well as their higher dimensional generalisations.

We will first review some familiar objects from classical geometry and try to point out the connection with important themes in modern mathematics.

1.1. Sphere, projective geometry

The familiar 2-sphere S^2 is a surface that can be defined as the collection of unit vectors in 3-sphace:

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

DEFINITION 1.1.1. A *great circle* is the intersection of the sphere with a plane passing through the origin.

The equator is an example of a great circle.

DEFINITION 1.1.2. The *great circle distance* on S^2 is the distance measured along the arcs of great circles.

Namely, the distance between a pair of points $p, q \in S^2$ is the length of the smaller of the two arcs of the great circle passing through p and q.

The antipodal quotient gives the *real projective plane* $\mathbb{RP}^2$, a space of fundamental importance in projective geometry.

1.2. Circle, isoperimetric inequality

The familiar unit circle in the plane, defined to be the locus of the equation

$$x^2 + y^2 = 1$$

in the (x, y)-plane. The circle solves the isoperimetic problem in the plane. Namely, consider simple (non-self-intersecting) closed curves of equal perimeter, for instance a polygon.[1]

[1] metzula

Among all such curves, the circle is the curve that encloses the largest area. In other words, the circle satisfies the boundary case of equality in the following inequality, known as the *isoperimetric inequality*.

Theorem 1.2.1 (Isoperimetric inequality). *Every Jordan curve in the plane satisfies the inequality*

$$\left(\frac{L}{2\pi}\right)^2 - \frac{A}{\pi} \geq 0,$$

with equality if and only if the curve is a round circle.

[2]

1.3. Relativity theory

In relativity theory, one uses a framework similar to classical differential geometry, with a technical difference having to do with the basic quadratic form being used. Nonetheless, some of the key concepts, such as geodesic and curvature, are common to both approaches.

In the first approximation, one can think of relativity theory as the study of 4-manifolds with a choice of a "light cone"[3] at every point.

Einstein gave a strong impetus to the development of differential geometry, as a tool in studying relativity. We will systematically use Einstein's summation convention (see below).

[2]The round circle is the subject of Gromov's filling area conjecture.

Definition 1.2.2. The *Riemannian circle* of length 2π is a great circle of the unit sphere, equipped with the great-circle distance.

The emphasis is on the fact that the distance is measured along arcs rather than chords (straight line intervals).

For all the apparent simplicity of the the Riemannian circle, it turns out that it is the subject of a still-unsolved conjecture of Gromov's, namely the filling area conjecture.

A surface with a single boundary circle will be called a *filling* of that circle. We now consider fillings of the Riemannian circles such that the ambient distance does not diminish the great-circle distance (in particular, filling by the unit disk is not allowed).

Conjecture 1.2.3 (M. Gromov). *Among all fillings of the Riemannian circles by a surface, the hemisphere is the one of least area.*

[3]konus ha'or

1.4. Linear algebra, index notation

Let $\mathbb{R}^n$ denote the Euclidean n-space. Its vectors will be denoted

$$v, w \in \mathbb{R}^n.$$

Let B be an n by n matrix. There are two ways of viewing such a matrix, either as a linear map or as a bilinear form (*cf.* Remarks 1.4.1 and 1.6). Developing suitable notation to capture this distinction helps simplify differential-geometric formulas down to readable size, and ultimately to motivate the crucial distinction between a vector and a covector.

Remark 1.4.1 (Matrix as a bilinear form $B(v, w)$). Consider a bilinear form

$$B(v, w) : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}, \tag{1.4.1}$$

sending the pair of vectors (v, w) to the real number $v^t B w$. Here v^t is the transpose of v. We write

$$B = (b_{ij})_{i=1,\dots,n;\, j=1,\dots,n}$$

so that b_{ij} is an entry while (b_{ij}) denotes the matrix.

For example, in the 2 by 2 case,

$$v = \begin{pmatrix} v^1 \\ v^2 \end{pmatrix}, w = \begin{pmatrix} w^1 \\ w^2 \end{pmatrix}, B = \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix}.$$

Then

$$Bw = \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} \begin{pmatrix} w^1 \\ w^2 \end{pmatrix} = \begin{pmatrix} b_{11}w^1 + b_{12}w^2 \\ b_{21}w^1 + b_{22}w^2 \end{pmatrix}.$$

Now the transpose $v^t = (v^1 \ v^2)$ is a row vector, so

$$\begin{aligned} B(v, w) &= v^t B w \\ &= (v^1 \ v^2) \begin{pmatrix} b_{11}w^1 + b_{12}w^2 \\ b_{21}w^1 + b_{22}w^2 \end{pmatrix} \\ &= b_{11}v^1 w^1 + b_{12}v^1 w^2 + b_{21}v^2 w^1 + b_{22}v^2 w^2, \end{aligned}$$

and therefore $v^t B w = \sum_{i=1}^2 \sum_{j=1}^2 b_{ij} v^i w^j$. We would like to simplify this notation by deleting the summation symbols "Σ", as follows.

1.5. Einstein summation convention

The following useful notational device was originally introduced by Albert Einstein.

DEFINITION 1.5.1. The rule is that whenever a product contains a symbol with a lower index and another symbol with the *same* upper index, take summation over this repeated index.

Using this notation, the bilinear form (1.4.1) defined by the matrix B can be written as follows:

$$B(v, w) = b_{ij} v^i w^j,$$

with implied summation over both indices.

DEFINITION 1.5.2. Let B be a symmetric matrix. The *associated quadratic form* is a quadratic form associated with a bilinear form $B(v, w)$ by the following rule:

$$Q(v) = B(v, v).$$

Let $v = v^i e_i$ where $\{e_i\}$ is the standard basis of $\mathbb{R}^n$.

LEMMA 1.5.3. *The associated quadratic form satisfies*

$$Q(v) = b_{ij} v^i v^j.$$

PROOF. To compute $Q(v)$, we must introduce an extra index j :

$$Q(v) = B(v, v) = B(v^i e_i, v^j e_j) = B(e_i, e_j) v^i v^j = b_{ij} v^i v^j. \tag{1.5.1}$$

□

DEFINITION 1.5.4. The *polarisation formula* is the following formula:

$$B(v, w) = \frac{1}{4}(Q(v + w) - Q(v - w)).$$

The polarisation formula allows one to reconstruct the bilinear form from the quadratic form, at least if the characteristic is not 2. For an application, see section 13.3.

1.6. Matrix as a linear map

Consider a map

$$B : \mathbb{R}^n \to \mathbb{R}^n, \quad v \mapsto Bv.$$

In order to distinguish this case from the case of the bilinear form, we will raise the first index: write B as

$$B = ({b^i}_j)_{i=1,\ldots,n;\, j=1,\ldots,n}$$

where it is important to s t a g g e r the indices, meaning that we do **not** put j under i as in

$$b^i_j,$$

but, rather, leave a blank space (in the place were j used to be), as in

$$b^i{}_j.$$

Let $v = (v^j)_{j=1,\dots,n}$ and $w = (w^i)_{i=1,\dots,n}$. Then the equation $w = Bv$ can be written as a system of n scalar equations,

$$w^i = b^i{}_j v^j \text{ for } i = 1, \dots, n$$

using the Einstein summation convention (here the repeated index is j).

Definition 1.6.1. The formula for the trace $Tr(B) = b^1{}_1 + b^2{}_2 + \cdots + b^n{}_n$ in Einstein notation becomes

$$Tr(B) = b^i{}_i$$

(here the repeated index is i).

1.7. Symmetrisation and skew-symmetrisation

Let $B = (b_{ij})$. Its symmetric part S is by definition

$$S = \frac{1}{2}(B + B^t) = \left(\frac{1}{2}(b_{ij} + b_{ji})\right)_{\substack{i=1,\dots,n \\ j=1,\dots,n}},$$

while the skew-symmetric part A is

$$A = \frac{1}{2}(B - B^t) = \left(\frac{1}{2}(b_{ij} - b_{ji})\right)_{\substack{i=1,\dots,n \\ j=1,\dots,n}}.$$

Note that $B = S + A$.

Another useful notation is that of symmetrisation

$$b_{\{ij\}} = \frac{1}{2}(b_{ij} + b_{ji}) \tag{1.7.1}$$

and antisymmetrisation

$$b_{[ij]} = \frac{1}{2}(b_{ij} - b_{ji}). \tag{1.7.2}$$

Lemma 1.7.1. *A matrix $B = (b_{ij})$ is symmetric if and only if for all indices i and j one has $b_{[ij]} = 0$.*

Proof. We have $b_{[ij]} = \frac{1}{2}(b_{ij} - b_{ji}) = 0$ since symmetry of B means $b_{ij} = b_{ji}$. □

1.8. Matrix multiplication in index notation

The usual way to define matrix multiplication is as follows. A triple of matrices $A = (a_{ij})$, $B = (b_{ij})$, and $C = (c_{ij})$ satisfy the product relation $C = AB$ if, introducing an additional dummy index k (*cf.* formula (1.5.1)), we have the relation $c_{ij} = \sum_k a_{ik}b_{kj}$.

Example 1.8.1 (Skew-symmetrisation of matrix product). By commutativity of multiplication of real numbers, $a_{ik}b_{kj} = b_{kj}a_{ik}$. Then the coefficients $c_{[ij]}$ of the skew-symmetrisation of $C = AB$ satisfy

$$c_{[ij]} = \sum_k b_{k[j}a_{i]k}.$$

Here by definition

$$b_{k[j}a_{i]k} = \frac{1}{2}(b_{kj}a_{ik} - b_{ki}a_{jk}).$$

This notational device will be particularly useful in writing down the *theorema egregium* (see Section 11.6). Given below are a few examples:

- See section 7.9, where we will use formulas of type
$$g_{mj}\Gamma^m_{ik} + g_{mi}\Gamma^m_{jk} = 2g_{m\{j}\Gamma^m_{i\}k};$$
- section 9.7 for $L^i{}_{[j}L^k{}_{l]}$;
- section 11.5 for $\Gamma^k_{i[j}\Gamma^n_{l]m}$.

The index notation we have described reflects the fact that the natural products of matrices are the ones which correspond to composition of maps. Thus, if $A = (a^i{}_j)$, $B = (b^i{}_j)$, and $C = (c^i{}_j)$ then the product relation $C = AB$ simplifies to the relation

$$c^i{}_j = a^i{}_k b^k{}_j.$$

1.9. The inverse matrix, Kronecker delta

Let $B = (b_{ij})$. The inverse matrix B^{-1} is sometimes written as

$$B^{-1} = (b^{ij})$$

(here both indices have been raised). Then the equation

$$B^{-1}B = I$$

becomes $b^{ik}b_{kj} = \delta^i{}_j$ (in Einstein notation with repeated index k), where the expression

$$\delta^i{}_j = \begin{cases} 1 & \text{if} \quad i = j \\ 0 & \text{if} \quad i \neq j \end{cases}$$

is referred to as the Kronecker delta function.

Example 1.9.1. The identity endomorphism $I = (\delta^i{}_j)$ by definition satisfies $AI = A = IA$ for all endomorphisms $A = (a^i{}_j)$, or equivalently

$$a^i{}_j \delta^j{}_k = a^i{}_k = \delta^i{}_j a^j{}_k,$$

using the Einstein summation convention.

Example 1.9.2. Let $\delta^i{}_j$ be the Kronecker delta function on $\mathbb{R}^n$, where $i, j = 1, \ldots, n$, viewed as a linear transformation $\mathbb{R}^n \to \mathbb{R}^n$.

(1) Evaluate the expression $\delta^i{}_j \delta^j{}_k$
(2) Evaluate the expression $\delta^i{}_j \delta^j{}_i$

Definition 1.9.3. Given a pair of vectors $v = v^i e_i$ and $w = w^i e_i$ in $\mathbb{R}^3$, their *vector product* is a vector $v \times w \in \mathbb{R}^3$ satisfying one of the following two equivalent conditions:

(1) we have $v \times w = \det \begin{bmatrix} e_1 & e_2 & e_3 \\ v^1 & v^2 & v^3 \\ w^1 & w^2 & w^3 \end{bmatrix}$, in other words,

$$\begin{aligned} v \times w &= (v^2 w^3 - v^3 w^2) e_1 - (v^1 w^3 - v^3 w^1) e_2 + (v^1 w^2 - v^2 w^1) e_3 \\ &= 2 \left(v^{[2} w^{3]} e_1 - v^{[1} w^{3]} e_2 + v^{[1} w^{2]} e_3 \right). \end{aligned}$$

(2) the vector $v \times w$ is perpendicular to both v and w, of length equal to the area of the parallelogram spanned by the two vectors, and furthermore satisfying the right hand rule, meaning that the 3 by 3 matrix formed by the three vectors v, w, and $v \times w$ has positive determinant.

Theorem 1.9.4. *We have an identity*

$$a \times (b \times c) = (a \cdot c) b - (a \cdot b) c$$

for every triple of vectors a, b, c in $\mathbb{R}^3$.

Properly understanding surface theory and related key concepts such as the Weingarten map (see Section 9.4) depends on linear-algebraic background related to diagonalisation of symmetric matrices or, more generally, selfadjoint endomorphisms.

1.10. Eigenvalues, symmetry

In general, a real matrix may not have a real eigenvector or eigenvalue. Thus, the matrix of 90 degree rotation in the plane,

$$\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$$

does not have a real eigenvector for obvious geometric reasons.

In this section we prove the existence of a real eigenvector (and hence, a real eigenvalue) for a real symmetric matrix.

This fact has important ramifications in surface theory, since the various notions of curvature of a surface are defined in terms of the eigenvalues of a certain symmetric matrix, or more precisely selfadjoint operator (see sections 9.1 and 10.5). Let

$$I = \begin{pmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{pmatrix}$$

be the (n, n) identity matrix. Thus

$$I = (\delta_{ij})_{\substack{i=1,\ldots,n \\ j=1,\ldots,n}}$$

where δ_{ij} is the Kronecker delta. Let B be an (n, n)-matrix.

Definition 1.10.1. A real number λ is called an eigenvalue of B if

$$\det(B - \lambda I) = 0.$$

Theorem 1.10.2. *If $\lambda \in \mathbb{R}$ is an eigenvalue of B, then there is a vector $v \in \mathbb{R}^n$, $v \neq 0$, such that*

$$Bv = \lambda v. \tag{1.10.1}$$

The proof is standard.

Definition 1.10.3. A nonzero vector satisfying (1.10.1) is called an eigenvector belonging to λ.

Recall the formula for the Euclidean inner product:

$$\langle v, w \rangle = v^1 w^1 + \cdots + v^n w^n = \sum_{i=1}^{n} v^i w^i.$$

Recall that all of our vectors are *column* vectors.

Lemma 1.10.4. *The inner product can be expressed in terms of matrix multiplication in the following fashion:*

$$\langle v, w \rangle = v^t w.$$

Recall the following property of the transpose: $(AB)^t = B^t A^t$.

Lemma 1.10.5. *A real matrix B is symmetric if and only if for all $v, w \in \mathbb{R}^n$, one has*

$$\langle Bv, w \rangle = \langle v, Bw \rangle.$$

PROOF. We have

$$\langle Bv, w\rangle = (Bv)^t w = v^t B^t w = \langle v, B^t w\rangle = \langle v, Bw\rangle,$$

proving the lemma. □

1.11. Finding an eigenvector of a symmetric matrix

THEOREM 1.11.1. *Every real symmetric matrix possesses a real eigenvector.*

We will give two proofs of this important theorem. The first proof is simpler and passes via complexification.

FIRST PROOF. where $n \geq 1$, and let B be an $n \times n$ real symmetric matrix, $B \in M_{n,n}(\mathbb{R})$. As such, it defines a linear map

$$B_{\mathbb{R}} : \mathbb{R}^n \to \mathbb{R}^n$$

sending $v \in \mathbb{R}^n$ to Bv, as usual. We now use the field extension $\mathbb{R} \subset \mathbb{C}$ and view B as a complex matrix

$$B \in M_{n,n}(\mathbb{C}).$$

Then the matrix B defines a complex linear map

$$B_{\mathbb{C}} : \mathbb{C}^n \to \mathbb{C}^n$$

sending $v \in \mathbb{C}^n$ to Bv. The characteristic polynomial of $B_{\mathbb{C}}$ is a polynomial of positive degree $n > 0$ and therefore has a root

$$\lambda \in \mathbb{C}$$

by the fundamental theorem of algebra. Let $v \in \mathbb{C}^n$ be an associated eigenvector. Let $\langle\ ,\ \rangle$ be the standard Hermitian inner product in $\mathbb{C}^n$. Recall that the Hermitian inner product is linear in one variable and *skew-linear* in the other. Thus, we have[4]

$$\langle z, w\rangle = \sum_i^n z_i \overline{w_i}$$

Then $\langle Bv, v\rangle = \langle v, Bv\rangle$ since B is real symmetric. Hence

$$\langle \lambda v, v\rangle = \langle v, \lambda v\rangle = \overline{\lambda}\langle v, v\rangle,$$

and therefore λ is real. □

The second proof is somewhat longer but has the advantage of being more geometric, as well as more concrete in the choice of the vector in question.[5]

[4]Alternatively, some texts adopt the convention $\langle z, w\rangle = \sum_i^n \overline{z_i} w_i$.
[5]We present the alternative proof here.

Second proof. Let $S \subset \mathbb{R}^n$ be the unit sphere $S = \{v \in \mathbb{R}^n \mid \|v\| = 1\}$. Given a symmetric matrix B, define a function $f : S \to \mathbb{R}$ as the restriction to S of the function, also denoted f, given by $f(v) = \langle v, Bv\rangle$. Let v_0 be a maximum of f restricted to S. Let $V_0^\perp \subset \mathbb{R}^n$ be the orthogonal complement of the line spanned by v_0. Let $w \in V_0^\perp$. Consider the curve $v_0 + tw$, $t \geq 0$ (see also a different choice of curve in Remark 1.11.2 below). Then $\left.\frac{d}{dt}\right|_{t=0} f(v_0 + tw) = 0$ since v_0 is a maximum and w is tangent to the sphere. Now

$$
\begin{aligned}
\left.\frac{d}{dt}\right|_0 f(v_0 + tw) &= \left.\frac{d}{dt}\right|_0 \langle v_0 + tw, B(v_0 + tw)\rangle \\
&= \left.\frac{d}{dt}\right|_0 (\langle v_0, Bv\rangle + t\langle v_0, Bw\rangle + t\langle w, Bv_0\rangle + t^2\langle w, Bv\rangle) \\
&= \langle v_0, Bw\rangle + \langle w, Bv_0\rangle \\
&= \langle Bw, v_0\rangle + \langle Bv_0, w\rangle \\
&= \langle B^t v_0, w\rangle + \langle Bv_0, w\rangle \\
&= \langle (B^t + B)v_0, w\rangle \\
&= 2\langle Bv_0, w\rangle \quad \text{by symmetry of } B.
\end{aligned}
$$

Thus $\langle Bv_0, w\rangle = 0$ for all $w \in V_0^\perp$. Hence Bv_0 is proportional to v_0 and so v_0 is an eigenvector of B. □

Remark 1.11.2. Our calculation used the curve $v_0 + tw$ which, while tangent to S at v_0 (see section 6.5), does not lie on S. If one prefers, one can use instead the curve $(\cos t)v_0 + (\sin t)w$ lying on S. Then

$$
\left.\frac{d}{dt}\right|_{t=0} \langle (\cos t)v_0 + (\sin t)w, B((\cos t)v_0 + (\sin t)w)\rangle = \cdots = \langle (B^t + B)v_0, w\rangle
$$

and one argues by symmetry as before.

CHAPTER 2

Self-adjoint operators, conic sections

First we reinforce the material on index notation from the previous lecture.

2.1. Trace of product in index notation

The following result is important in its own right. We reproduce it here because its proof is a good illustration of the uses of the Einstein index notation.

THEOREM 2.1.1. *Let A and B be square $n \times n$ matrices. Then $tr(AB) = tr(BA)$.*

PROOF. Let $A = (a^i{}_j)$ and $B = (b^i{}_j)$. Then

$$\mathrm{tr}(AB) = \mathrm{tr}(a^i{}_k b^k_j) = a^i{}_k b^k_i$$

by definition of trace (see Definition 1.6.1). Meanwhile,

$$\mathrm{tr}(BA) = \mathrm{tr}(b^k_i a^i{}_j) = b^k_i a^i{}_k = a^i{}_k b^k_i = \mathrm{tr}(AB),$$

proving the theorem. □

2.2. Inner product spaces and self-adjoint operators

In the previous section we worked with real matrices and showed that the symmetry of a matrix guarantees the existence of a real eigenvector. In a more general situation where a basis is not available, a similar statement holds for a special type of endomorphism of a real vector space.

DEFINITION 2.2.1. Let V be a real inner product space. An endomorphism $B : V \to V$ is *selfadjoint* if one has

$$\langle Bv, w\rangle = \langle v, Bw\rangle \quad \forall v, w \in V. \tag{2.2.1}$$

COROLLARY 2.2.2. *Every selfadjoint endomorphism of a real inner product space admits a real eigenvector.*

PROOF. The selfadjointness was the relevant property in the proof of Theorem 1.11.1. □

2.3. Orthogonal diagonalisation of symmetric matrices

THEOREM 2.3.1. *Every symmetric matrix can be orthogonally diagonalized.*

PROOF. A symmetric $n \times n$ matrix S defines a selfadjoint endomorphism f_S of the real inner product space $V = \mathbb{R}^n$ given by

$$f_S : V \to V, \quad v \mapsto Sv.$$

By Corollary 2.2.2, every selfadjoint endomorphism has a real eigenvector $v_1 \in V$, which we can assume to be a unit vector:

$$|v_1| = 1.$$

Let $\lambda_1 \in \mathbb{R}$ be its eigenvalue. Now we let $V_1 = V$ and set

$$V_2 \subset V_1$$

be the orthogonal complement of the line $\mathbb{R}v_1 \subset V_1$. Thus we have an orthogonal decomposition

$$V_1 = \mathbb{R}v_1 \oplus V_2.$$

Note that V_2 is invariant under f_S, since if $w \in V_2$ then

$$\langle f_S(w), v_1 \rangle = \langle Sw, v_1 \rangle = \langle w, Sv_1 \rangle = \langle w, \lambda_1 v_1 \rangle = \lambda_1 \langle w, v_1 \rangle = 0.$$

The restriction of f_S to V_2 is still selfadjoint by inheriting the property (2.2.1). Namely, since property (2.2.1) holds for all vectors $v, w \in V$, it still holds if these vectors are restricted to vary in a subspace $V_2 \subset V$, i.e.,

$$\langle Bv, w \rangle = \langle v, Bw \rangle \quad \forall v, w \in V_2. \tag{2.3.1}$$

Arguing inductively, we obtain an orthonormal basis consisting of eigenvectors $v_1, \ldots, v_n \in V$. Denote by $\lambda_1, \ldots, \lambda_n \in \mathbb{R}$ their eigenvalues. Let

$$P = [v_1 \ldots v_n]$$

be the orthogonal $n \times n$ matrix whose columns are the vectors v_i, so that we have

$$P^{-1} = P^t. \tag{2.3.2}$$

Consider the diagonal matrix $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$. By construction, we have

$$S = P \Lambda P^t$$

from (2.3.2), or equivalently,

$$SP = P\Lambda.$$

Indeed, to verify the relation $SP = P\Lambda$, note that both sides are equal to the square matrix $[\lambda_1 v_1 \ \lambda_2 v_2 \ \ldots \ \lambda_n v_n]$. □

2.4. Classification of conic sections: diagonalisation

A conic section[1] (or *conic* for short) in the plane is by definition a curve defined by the following equation in the (x, y)-plane:

$$ax^2 + 2bxy + cy^2 + dx + ey + f = 0, \quad a, b, c, d, e, f \in \mathbb{R}. \tag{2.4.1}$$

Here we chose the coefficient of the xy term to be $2b$ rather than b so as to simplify formulas like (2.4.2) below. Let $X = \begin{pmatrix} x \\ y \end{pmatrix}$ and let

$$S = \begin{pmatrix} a & b \\ b & c \end{pmatrix}. \tag{2.4.2}$$

Then

$$X^t S X = ax^2 + 2bxy + cy^2.$$

THEOREM 2.4.1. *Up to an orthogonal transformation, every conic section can be written in a "diagonal" form*

$$\lambda_1 x'^2 + \lambda_2 y'^2 + d'x' + e'y' + f = 0, \tag{2.4.3}$$

where the coefficients λ_1 and λ_2 are the eigenvalues of the matrix S.

PROOF. Consider the row vector

$$T = \begin{pmatrix} d & e \end{pmatrix}.$$

Then $TX = dx + ey$. Thus equation (2.4.1) becomes

$$X^t S X + TX + f = 0. \tag{2.4.4}$$

We now apply Theorem 2.3.1 to orthogonally diagonalize S to obtain $S = P\Lambda P^t$. Substituting this into (2.4.4) yields

$$X^t P \Lambda P^t X + TX + f = 0.$$

We set $X' = P^t X$. Then $X = PX'$ since P is orthogonal. Furthermore, we have

$$(X')^t = (P^t X)^t = (X^t)(P^t)^t = (X^t)P.$$

Hence we obtain

$$(X')^t \Lambda X' + TPX' + f = 0.$$

Letting $T' = TP$, we obtain

$$(X')^t \Lambda X' + T'X' + f = 0.$$

Letting x' and y' be the components of X', i.e. $X' = \begin{pmatrix} x' \\ y' \end{pmatrix}$, we obtain (2.4.3), as required. □

[1] chatach charut

To obtain more precise information about the conic, we need to specify certain nondegeneracy conditions, as discussed in Section 2.5.

2.5. Classification of conics: trichotomy, nondegeneracy

We use the diagonalisation result of the previous section so as to classify conic sections into three types (under suitable nondegeneracy conditions): ellipse, parabola, hyberbola.

THEOREM 2.5.1. *Suppose* $\det(S) \neq 0$*, i.e., S is of rank* 2*. Then, up to a translation, the conic section can be written in the form*

$$\lambda_1 (x'')^2 + \lambda_2 (y'')^2 + f'' = 0 \tag{2.5.1}$$

(note that the constant term is changed).

PROOF. If the determinant is nonzero then both eigenvalues λ_i are nonzero. The term $d'x'$ in (2.4.3) can be absorbed into the quadratic term $\lambda_1 x'^2$ by completing the square as follows:

$$\begin{aligned}\lambda_1 x'^2 + d'x' &= \lambda_1 \left(x'^2 + 2\frac{d'}{2\lambda_1} x' \right) \\ &= \lambda_1 \left(x'^2 + 2\frac{d'}{2\lambda_1} x' + \left(\frac{d'}{2\lambda_1}\right)^2 \right) - \lambda_1 \left(\frac{d'}{2\lambda_1}\right)^2 \\ &= \lambda_1 \left(x' + \frac{d'}{2\lambda_1} \right)^2 - \frac{d'^2}{4\lambda_1},\end{aligned}$$

and we set

$$x'' = x' + \frac{d'}{2\lambda_1}.$$

Similarly $e'y'$ can be absorbed into $\lambda_2 y'^2$. Geometrically this corresponds to a translation along the axes x' and y', proving the theorem. □

DEFINITION 2.5.2. A conic section is called a *hyperbola* if $\lambda_1\lambda_2 < 0$, provided the following nondegeneracy condition is satisfied: the constant f'' in equation (2.5.1) is nonzero.

REMARK 2.5.3. If the constant is zero, then instead of a hyperbola we obtain a pair of transverse lines as the solution set.

DEFINITION 2.5.4. A conic section is called an *ellipse* if $\lambda_1\lambda_2 > 0$, provided the following nondegeneracy condition is satisfied: the constant f'' in equation (2.5.1) is nonzero and has the opposite sign as compared to the sign of λ_1.

REMARK 2.5.5. If the constant is zero then the ellipse degenerates to a single point $x'' = y'' = 0$. If $f'' \neq 0$ has the same sign as λ_1 then the solution set is empty.

If the determinant of S is zero then one cannot eliminate the linear term. Therefore we continue working with the equation $\lambda_1 x'^2 + \lambda_2 y'^2 + d'x' + e'y' + f = 0$ from (2.4.3).

DEFINITION 2.5.6. The conic is a *parabola* if the following two conditions are satisfied:

(1) the matrix S is of rank 1 (this is equivalent to saying that $\lambda_1\lambda_2 = 0$ *and* one of the λ_i is nonzero);
(2) if $\lambda_1 = 0$ then $d' \neq 0$, and if $\lambda_2 = 0$ then $e' \neq 0$.

Since the determinant of the matrix S is the product of its eigenvalues, we obtain the following corollary.

COROLLARY 2.5.7. *If the conic $ax^2 + 2bxy + cy^2 + dx + ey + f = 0$ is an ellipse then $ac - b^2 > 0$. If $ac - b^2 > 0$ and the solution locus is neither empty nor a single point, then it is an ellipse.*

COROLLARY 2.5.8. *If the conic $ax^2 + 2bxy + cy^2 + dx + ey + f = 0$ is a hyperbola then $ac - b^2 < 0$. If $ac - b^2 < 0$ and the solution locus is not a pair of transverse lines, then the conic is a hyperbola.*

2.6. Quadratic surfaces

A quadratic surface in $\mathbb{R}^3$ is the locus of points satisfying the equation

$$ax^2 + 2bxy + cy^2 + 2dxz + fz^2 + 2gyz + hx + iy + jz + k = 0, \quad (2.6.1)$$

where $a, b, c, d, e, f, g, h, i, j, k \in \mathbb{R}$.

To bring this to standard form, we apply an orthogonal diagonalisation procedure similar to that employed in Section (2.5). Thus, we define matrices S, X, and D by setting

$$S = \begin{pmatrix} a & b & d \\ b & c & g \\ d & g & f \end{pmatrix}, \quad X = \begin{pmatrix} x \\ y \\ z \end{pmatrix}, \quad T = \begin{pmatrix} h & i & j \end{pmatrix},$$

so that the quadratic part of (2.6.1) becomes $X^t S X$, and the linear part becomes DX. Then equation (2.6.1) takes the form

$$X^t S X + DX + k = 0.$$

Orthogonally diagonalizing S as before, we conclude that equation (2.6.1) can be simplified to

$$ax^2 + by^2 + cz^2 + dx + ey + fz + g = 0, \quad (2.6.2)$$

with new variables x, y, z and new coefficients $a, b, c, d, e, f, g \in \mathbb{R}$.

The classification of quadratic surfaces is more involved than the case of curves, and will not be pursued here. However, we point out some important special cases.

DEFINITION 2.6.1. A quadratic surface is called an *ellipsoid* if the coefficients a, b, c in (2.6.2) are nonzero and have the same sign, and moreover the solution locus is neither a single point nor the empty set.

REMARK 2.6.2. The nondegeneracy condition can be ensured by assuming that the linear terms in (2.6.2) all vanish, and the constant term g has the opposite sign as compared to the sign of, say, the coefficient a.

Additional special cases are

- the paraboloid $z = ax^2 + by^2$,
- the hyperbolic paraboloid $z = x^2 - y^2$,
- the hyperboloid of one sheet[2] with equation $z^2 = x^2 + y^2 - 1$,
- the hyperboloid of two sheets[3] with equation $z^2 = x^2 + y^2 + 1$.

2.7. Jacobi's criterion

The type of quadratic surface one obtains depends critically on the signs of the eigenvalues of the matrix S. The signs of the eigenvalues can be determined without diagonalisation by means of Jacobi's criterion.

DEFINITION 2.7.1. Given a matrix A over a field F, let Δ_k denote the $k \times k$ upper-left block, called a *principal minor*.

DEFINITION 2.7.2. Two matrices are *equivalent* if they are congruent (rather than similar), meaning that we transform a matrix A by $B^t AB$ (rather than by conjugation $B^{-1}AB$).

Here B is of course *not* assumed to be orthogonal.

THEOREM 2.7.3 (Jacobi). *Let $A \in M_n(F)$ be a symmetric matrix, and assume $\det(\Delta_k) \neq 0$ for $k = 1, \ldots, n$. Then A is equivalent to the matrix*

$$diag\left(\frac{1}{\Delta_1}, \frac{\Delta_1}{\det\Delta_2}, \ldots, \frac{\det\Delta_{n-1}}{\det\Delta_n}\right).$$

[2] chad-yeriati

[3] du-yeriati

EXAMPLE 2.7.4. For a 2×2 symmetric matrix $\begin{pmatrix} a & b \\ b & d \end{pmatrix}$ Jacobi's criterion affirms the equivalence to $\begin{pmatrix} \frac{1}{a} & 0 \\ 0 & \frac{a}{ad-b^2} \end{pmatrix}$.

PROOF. Take the vector $b_k = \Delta_k^{-1} e_k \in F^k$ of length k, and pad it with zeros up to length n. Consider the matrix $B = (b_{ij})$ whose column vectors are $b_1, \ldots, b_n$. By Cramer's formula, the diagonal coefficients of B satisfy

$$b_{kk} = \det \begin{pmatrix} \Delta_{k-1} & 0 \\ 0 & 1/\det\Delta_k \end{pmatrix} = \det\Delta_{k-1}/\Delta_k,$$

so $\det(B) = \prod_{k=1}^n b_{kk} = 1/\det(A) \neq 0$. Compute that $B^t AB$ is lower triangular with diagonal $b_{11}, \ldots, b_{kk}$. Being symmetric, it is diagonal. □

REMARK 2.7.5. If some minor Δ_k is not invertible, then A cannot be definite.

Applying this result in the case of a real symmetric matrix, we obtain the following corollary.

COROLLARY 2.7.6. *Let A be symmetric. Then A is positive definite if and only if all* $\det(\Delta_k) > 0$.

Define minors in general (choose rows and columns $i_1, \ldots, i_t$). Permuting rows and columns, we obtain the following corollary.

COROLLARY 2.7.7. *Let A be symmetric positive definite matrix. Then all "diagonal" minors are positive definite (and in particular have positive determinants).*

An immediate application of this is determining whether or not a quadratic surface is an ellipsoid, without having to orthogonally diagonalize the matrix of coefficients.

EXAMPLE 2.7.8. Determine whether or not the quadratic surface

$$x^2 + xy + y^2 + xz + z^2 + yz + x + y + z - 2 = 0 \tag{2.7.1}$$

is an ellipsoid.

To solve the problem, we first construct the corresponding matrix

$$S = \begin{pmatrix} 1 & 1/2 & 1/2 \\ 1/2 & 1 & 1/2 \\ 1/2 & 1/2 & 1 \end{pmatrix}$$

and then calculate the principal minors $\Delta_1 = 1$, $\Delta_2 = 1 \cdot 1 - \frac{1}{2} \cdot \frac{1}{2} = 3/4$, and

$$\Delta_3 = 1 + \frac{1}{8} + \frac{1}{8} - \frac{1}{4} - \frac{1}{4} - \frac{1}{4} = \frac{1}{2}.$$

Thus all principal minors are positive and therefore the surface is an ellipsoid, provided we can show it is nondegenerate. To check nondegeneracy, notice that (2.7.1) has at least two distinct solutions: $(x, y, z) = (1, 0, 0)$ and $(x, y, z) = (0, 1, 0)$. Therefore it is a nondegenerate ellipsoid.

CHAPTER 3

Hessian, curves, and curvature

3.1. Exercise on index notation

THEOREM 3.1.1. *Every* 2×2 *matrix* $A = (a^i{}_j)$ *satisfies the identity*

$$a^i{}_k a^k{}_j + q\ \delta^i{}_j = a^k{}_k a^i{}_j, \tag{3.1.1}$$

where $q = \det(A)$.

PROOF. We will use the Cayley-Hamilton theorem which asserts that $p_A(A) = 0$ where $p_A(\lambda)$ is the characteristic polynomial of A. In the rank 2 case, we have $p_A(\lambda) = \lambda^2 - (\mathrm{tr}A)\lambda + \det(A)$, and therefore we obtain

$$A^2 - (\mathrm{tr}A)A + \det(A)I = 0,$$

which in index notation gives $a^i{}_k a^k{}_j - a^k{}_k a^i{}_j + q\ \delta^i{}_j = 0$. This is equivalent to (3.1.1). □

3.2. Hessian, minima, maxima, saddle points

Given a smooth (C^2) function $f(u^1, \ldots, u^n)$ of n variables, denote by

$$H_f = (f_{ij})_{i=1,\ldots,n;\, j=1,\ldots,n}$$

the Hessian matrix of f, *i.e.* the $n \times n$ matrix of second partial derivatives

$$f_{ij} = \frac{\partial^2 f}{\partial u^i \partial u^j}$$

of f.

THEOREM 3.2.1 (equality of mixed partials). *In terms of the anti-symmetrisation notation defined above* (1.7.2), *we have the identity*

$$f_{[ij]} = 0. \tag{3.2.1}$$

Recall that the gradient of f at a point $p = (u^1, \ldots, u^n)$ is the vector

$$\nabla f = \begin{pmatrix} \frac{\partial f}{\partial u^1} \\ \frac{\partial f}{\partial u^2} \\ \vdots \\ \frac{\partial f}{\partial u^n} \end{pmatrix}$$

DEFINITION 3.2.2. A critical point p of f is a point satisfying

$$\nabla f(p) = 0,$$

i.e. $\frac{\partial f}{\partial u^i}(p) = 0$ for all $i = 1, \ldots, n$.

EXAMPLE 3.2.3 (maxima, minima, saddle points). Let $n = 2$. Then the *sign* of the determinant

$$\det(H_f) = \frac{\partial^2 f}{\partial u^1 \partial u^1} \frac{\partial^2 f}{\partial u^2 \partial u^2} - \left(\frac{\partial^2 f}{\partial u^1 \partial u^2} \right)^2$$

at a critical point has geometric significance. Namely, if

$$\det(H_f(p)) > 0,$$

then p is a maximum or a minimum. If $\det(H_f(p)) < 0$, then p is a *saddle point.*[1]

EXAMPLE 3.2.4 (Quadratic surfaces). Quadratic surfaces are a rich source of examples.

(1) The origin is a critical point for the function whose graph is the paraboloid $z = x^2 + y^2$. In the case of the paraboloid the critical point is a minimum.
(2) Similar remarks apply to the top sheet of the hyperboloid of two sheets, namely $z = \sqrt{x^2 + y^2 + 1}$, where we also get a minimum.
(3) The origin is a critical point for the function whose graph is the hyperbolic paraboloid $z = x^2 - y^2$. In the case of the hyperbolic paraboloid the critical point is a saddle point.

In addition to the sign, the *value* of $H_f(p)$ also has geometric significance, expressed by the following theorem.

THEOREM 3.2.5. *Let* $p \in \mathbb{R}^2$ *be a critical point of* f. *Then the value of* $\det(H_f(p))$ *is precisely the Gaussian curvature at* $(p, f(p)) \in \mathbb{R}^3$ *of the surface given by the graph of* f *in* $\mathbb{R}^3$.

See Definition 9.7.1 for more details.

3.3. Parametric representation of a curve

There are two main ways of representing a curve in the plane: parametric and implicit.

A curve in the plane can represented by a pair of coordinates evolving as a function of time t, called the parameter:

$$\alpha(t) = (\alpha^1(t), \alpha^2(t)), \quad t \in [a, b],$$

[1]Ukaf

so that $x(t) = \alpha^1(t)$ and $y(t) = \alpha^2(t)$. Thus a curve can be viewed as a map

$$\alpha : [a, b] \to \mathbb{R}^2. \tag{3.3.1}$$

Let C be the image of the map (3.3.1). Then C is the geometric curve independent of parametrisation. Thus, changing the parametrisation by setting $t = t(s)$ and replacing α by a new curve $\beta(s) = \alpha(t(s))$ preserves the geometric curve.

DEFINITION 3.3.1. A parametrisation is called *regular* if $\alpha'(t) \neq 0$ for all t.

3.4. Implicit representation of a curve

A curve in the (x, y)-plane can also be represented implicitly as the solution set of an equation

$$F(x, y) = 0,$$

where F is a function always assumed sufficiently smooth. We will denote the corresponding curve C_F. Thus, a circle of radius r corresponds to the choice of the function

$$F(x, y) = x^2 + y^2 - r^2.$$

Further examples are given below.

(1) The function $F(x, y) = y - x^2$ defines a parabola.
(2) The function $F(x, y) = xy - 1$ defines a hyperbola.
(3) The function $F(x, y) = x^2 - y^2 - 1$ defines a hyperbola.

In each of these cases, it is easy to find a parametrisation (at least of a part of the curve), by solving the equation for one of the variables. Thus, in the case of the circle, we choose the positive square root to obtain $y = \sqrt{r^2 - x^2}$, giving a parametrisation of the upperhalf circle by means of the pair of formulas

$$\alpha^1(t) = t, \quad \alpha^2(t) = \sqrt{r^2 - t^2}.$$

Note this is not all of the curve C_F.

Unlike the above examples, in general it is difficult to find an explicit parametrisation. Locally one can always find one in theory under a suitable nondegeneracy condition, expressed by the implicit function theorem.

3.5. Implicit function theorem

THEOREM 3.5.1 (implicit function theorem). *Let $F(x, y)$ be a smooth function. Suppose the gradient of F does not vanish at a point $p \in C_F$, in other words*

$$\nabla F(p) \neq 0.$$

Then there exists a regular parametrisation $(\alpha^1(t), \alpha^2(t))$ of the curve C_F in a neighborhood of p.

A useful special case is the following result.

THEOREM 3.5.2 (implicit function theorem: special case). *Let $F(x, y)$ be a smooth function, and suppose that*

$$\frac{\partial F}{\partial y}(p) \neq 0.$$

Then there exists a parametrisation $y = \alpha^2(x)$, in other words $\alpha(t) = (t, \alpha^2(t))$ of the curve C_F in a neighborhood of p.

EXAMPLE 3.5.3. In the case of the circle $x^2+y^2 = r^2$, the point $(r, 0)$ on the x-axis fails to satisfy the hypothesis of Theorem 3.5.2. The curve cannot be represented by a differentiable function $y = y(x)$ in a neighborhood of this point ("vertical tangent").

3.6. Unit speed parametrisation

We review the standard calculus topic of the curvature of a curve, which is indispensable to understanding principal curvatures of a surface (*cf.* Theorem 10.2.1 and Theorem 10.5.4).

Consider a parametrized curve $\alpha(t) = (\alpha^1(t), \alpha^2(t))$ in the plane. Denote by C the underlying geometric curve, i.e., the image of α:

$$C = \left\{(x, y) \in \mathbb{R}^2 \mid (\exists t) : \quad x = \alpha^1(t), y = \alpha^2(t)\right\}.$$

DEFINITION 3.6.1. We say $\alpha = \alpha(t)$ is a unit speed curve if $\left|\frac{d\alpha}{dt}\right| = 1$, *i.e.* $\left(\frac{d\alpha^1}{dt}\right)^2 + \left(\frac{d\alpha^2}{dt}\right)^2 = 1$ for all $t \in [a, b]$.

When dealing with a unit speed curve, it is customary to denote the parameter (called arclength) by s.

EXAMPLE 3.6.2. Let $r > 0$. Then the curve

$$\alpha(s) = \left(r \cos \frac{s}{r}, r \sin \frac{s}{r}\right)$$

is a unit speed parametrisation of the circle of radius r. Indeed, we have

$$\left|\frac{d\alpha}{ds}\right| = \sqrt{\left(r\frac{1}{r}\left(-\sin \frac{s}{r}\right)\right)^2 + \left(r\frac{1}{r}\cos \frac{s}{r}\right)^2} = \sqrt{\sin^2 \frac{s}{r} + \cos^2 \frac{s}{r}} = 1.$$

3.7. Geodesic curvature

Our main interest will be in space curves, when one cannot in general assign a sign to the curvature. Therefore in the definition below we do not concern ourselves with the sign of the curvature of plane curves, either. In this section, only local properties of curvature of curves will be studied. A global result on the curvature of curves may be found in Section 4.6.

DEFINITION 3.7.1. The (geodesic) curvature function $k_\alpha(s) \geq 0$ of a unit speed curve $\alpha(s)$ is defined by setting

$$k_\alpha(s) = \left| \frac{d^2\alpha}{ds^2} \right|. \tag{3.7.1}$$

EXAMPLE 3.7.2. For the circle of radius r parametrized as above, we have

$$\frac{d^2\alpha}{ds^2} = \left(r\frac{1}{r^2}\left(-\cos\frac{s}{r}\right), r\frac{1}{r^2}\left(-\sin\frac{s}{r}\right)\right)$$

at $s = 0$, and so the curvature satisfies

$$k_\alpha = \left| \frac{1}{r}\left(-\cos\frac{s}{r}, -\sin\frac{s}{r}\right)\right| = \frac{1}{r}.$$

Note that in this case, the curvature is independent of the point, *i.e.* is a constant function of s.

In Section 4.2, we will give a formula for curvature with respect to an arbitrary parametrisation (not necessarily arclength).

In Section 4.4, the curvature will be expressed in terms of the angle formed by the tangent vector with the positive x-axis.

3.8. Osculating circle of a curve

To give a more geometric description of the curvature in terms of the osculating circle, we first recall the following fact about the second derivative.

THEOREM 3.8.1. *The second derivative of a function $f(x)$ may be computed from a triple of points $f(x), f(x+h), f(x-h)$ that are infinitely close to each other, as follows:*

$$f''(x) = \lim_{h\to 0} \frac{f(x+h) + f(x-h) - 2f(x)}{h^2}.$$

The theorem remains valid for vector-valued functions, see Definition 3.8.2. We see that the second derivative can be calculated from the value of the function at a triple of nearby points $x, x+h, x-h$.

DEFINITION 3.8.2. The *osculating circle* to the curve parametrized by α at the point $\alpha(s)$ is obtained by choosing a circle passing through the three points

$$\alpha(s),\ \alpha(s-h),\ \alpha(s+h)$$

for infinitesimal h. The standard part[2] of the resulting circle is the osculating circle; equivalently, we take limit as h tends to zero.

Note that the osculating circle and the curve are "better than tangent" (they have second order tangency). Since the second derivative is computed from the same triple of points for $\alpha(s)$ and for the osculating circle (*cf.* Remark 3.8.1), we have the following [**We55**, p. 13].

THEOREM 3.8.3. *The curvatures of the osculating circle and the curve at the point of tangency are equal.*[3]

3.9. Radius of curvature

It is helpful to recall Leibniz's and Cauchy's definition of the radius of curvature of a curve (see Cauchy [**2**]): the radius of curvature is the distance from the curve to the intersection point of two infinitely close normals to the curve. In more detail, we have the following.

DEFINITION 3.9.1. The radius of curvature of a curve C at a point p is the distance from p to the intersection point of the normals to the curve at infinitely close points p and p' of C.

The intersection point of two infinitely close normals is the center of the osculating circle at p.

[2]See Keisler [**Ke74**].

[3]For example, let $y = f(x)$. Compute the curvature of the graph of f when $f(x) = ax^2$. Let $B = (x, x^2)$. Let A be the midpoint of OB. Let C be the intersection of the perpendicular bisector of OB with the y-axis. Let $D = (0, x)$,

Triangle OAC yields

$$\sin\psi = \frac{OA}{OC} = \frac{\frac{1}{2}\sqrt{x^2+(ax^2)^2}}{r},$$

triangle OBD yields

$$\sin\psi = \frac{BD}{OB} = \frac{ax^2}{\sqrt{x^2+a^2x^4}},$$

and

$$\frac{\frac{1}{2}\sqrt{x^2+a^2x^4}}{r} = \frac{ax^2}{\sqrt{x^2+a^2x^4}},$$

so that $\frac{1}{2}(x^2+a^2x^4) = arx^2$, and $\frac{1}{2}(1+a^2x^2) = ar$, so that $r = \frac{1+a^2x^2}{2a}$. Taking the limit as $x \to 0$, we obtain $r = \frac{1}{2a}$, hence $k = \frac{1}{r} = 2a = f''(0)$. Thus the curvature of the parabola at its vertex equals the second derivative with respect to x (even though x is not the arclength parameter of the graph).

3.10. Examples of second order operators

If a curve is given implicitly as the locus (solution set) of an equation $F(x, y) = 0$, one can calculate the geodesic curvature by means of the theorems given in the next section. We start with a definition.

DEFINITION 3.10.1. The *flat Laplacian* Δ_0 is the differential operator is defined by

$$\Delta_0 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}.$$

This means that when we apply Δ to a smooth function $F = F(x, y)$, we obtain

$$\Delta_0 F = \frac{\partial^2 F}{\partial x^2} + \frac{\partial^2 F}{\partial y^2}.$$

We also introduce the traditional shorter notation

$$F_{xx} = \frac{\partial^2 F}{\partial x^2}, \quad F_{yy} = \frac{\partial^2 F}{\partial y^2}, \quad F_{xy} = \frac{\partial^2 F}{\partial x\, \partial y}.$$

Then we obtain the equivalent formula

$$\Delta_0(F) = F_{xx} + F_{yy}.$$

EXAMPLE 3.10.2. If $F(x, y) = x^2 + y^2 - r^2$, then $\frac{\partial^2 F}{\partial x^2} = 2$ and similarly $\frac{\partial^2 F}{\partial y^2} = 2$, hence $\Delta_0 F = 4$.

We will be interested in the following operator.

DEFINITION 3.10.3. The Bateman-Reiss operator D_B is defined by

$$D_B(F) = F_{xx}F_y^2 - 2F_{xy}F_xF_y + F_{yy}F_x^2, \tag{3.10.1}$$

which is a non-linear second order differential operator.

The subscript "B" stands for Bateman, as in the Bateman equation $F_{xx}F_y^2 - 2F_{xy}F_xF_y + F_{yy}F_x^2 = 0$. Alternatively, the operator can be represented by the determinant

$$D_B(F) = -\det \begin{pmatrix} 0 & F_x & F_y \\ F_x & F_{xx} & F_{xy} \\ F_y & F_{xy} & F_{yy} \end{pmatrix}$$

This was treated in detail by Goldman [**Go05**, p. 637, formula (3.1)]. The same operator occurs in the Reiss relation in algebraic geometry (see Griffiths and Harris [**GriH78**, p. 677].[4]

[4]Michel Reiss (1805-1869).

3.11. Geodesic curvature for an implicit curve

The operator defined in the previous section allows us to calculate the curvature of a curve presented in implicit form, without having to look for a parametrisation. Let $C_F \subset \mathbb{R}^2$ be a curve defined implicitly by $F(x,y)=0$.

THEOREM 3.11.1. *Let $p \in C_F$, and suppose $\nabla F(p) \neq 0$. Then the geodesic curvature k of C_F at the point p is given by*

$$k = \frac{|D_B(F)|}{|\nabla F|^3},$$

where D_B is the Bateman-Reiss operator defined in (3.10.1).

EXAMPLE 3.11.2. In the case of the circle of radius r defined by the equation $F(x,y)=0$, where $F = x^2+y^2-r^2$, we obtain

$$F_x = 2x, \quad F_y = 2y, \quad \nabla F = (2x, 2y)^t,$$

and therefore $|\nabla F| = 2\sqrt{x^2+y^2} = 2r$. Meanwhile, $F_{xx} = 2$, $F_{yy} = 2$, $F_{xy} = 0$, hence

$$D_B(F) = 2(2y)^2 + 2(2x)^2 = 8r^2,$$

and therefore curvature is k$= \frac{8r^2}{8r^3} = \frac{1}{r}$.

3.12. Curvature of graph of function

THEOREM 3.12.1. *Let x_0 be a critical point of $f(x)$, and consider the graph of f at $(x_0, f(x_0))$. Then the curvature of the graph equals*

$$k = |f''(x_0)|.$$

PROOF. We parametrize the graph by $\alpha(t) = (t, f(t))$. Then we have $\alpha''(t) = (0, f''(t))$ and $|\alpha''(t)| = |f''(t)|$, which is the expected answer at a critical point. However, the parametrisation $(t, f(t))$ is not a unit speed parametrisation of the graph.

Intuitively, the second order Taylor polynomial of f at x_0 has the same osculating circle as f, and therefore it suffices to check the result for a standard curve such as a circle or a parabola.

We check that applying the characterisation of curvature in terms of the Bateman operator $D_B(F) = F_{xx}F_y^2 - 2F_{xy}F_xF_y + F_{yy}F_x^2$ gives the same answer. Let $F(x,y) = -f(x)+y$. At the critical point x_0, we have

$$\nabla F = (-f'(x_0), 1)^t = (1,0)^t,$$

while

$$D_B(F) = -f''(x)(1)^2 - 0 + 0 == -f''(x).$$

Hence the curvature at this point satisfies

$$k = \frac{|D_B F|}{|\nabla F|} = \frac{|f''(x_0)|}{1} = |f''(x_0)|,$$

as required. □

3.13. Existence of arclength

The unit speed parametrisation is sometimes called the arclength parametrisation of the geometric curve C.

THEOREM 3.13.1. *Suppose a curve $\alpha(t)$ satisfies $\alpha'(t) \neq 0$ at every point. Then there exists a unit speed parametrition $\beta(s) = \alpha(t(s))$, defined by equation* (3.13.1) *below, of the underlying geometric curve C.*

PROOF. Recall length of graph of $f(x)$ from a to b :

$$L = \int_a^b \sqrt{1 + (f'(x))^2}\, dx \ .$$

The element of length (infinitesimal increment) ds decomposes by Pythagoras' theorem as follows:

$$ds^2 = dx^2 + dy^2.$$

More generally, for a curve $\alpha(t) = (\alpha^1(t), \alpha^2(t))$, we have the following formula for the length:

$$L = \int_a^b \sqrt{\left(\frac{d(\alpha^1)}{dt}\right)^2 + \left(\frac{d\alpha^2}{dt}\right)^2} dt = \int_a^b \left|\frac{d\alpha}{dt}\right| dt.$$

We define the new parameter $s = s(t)$ by setting

$$s(t) = \int_a^t \left|\frac{d\alpha}{d\tau}\right| d\tau, \tag{3.13.1}$$

where τ is a dummy variable (internal variable of integration). By the Fundamental Theorem of Calculus, we have $\frac{ds}{dt} = \left|\frac{d\alpha}{dt}\right|$. Let $\beta(s) = \alpha(t(s))$. Then by chain rule

$$\frac{d\beta}{ds} = \frac{d\alpha}{dt}\frac{dt}{ds} = \frac{d\alpha}{dt}\frac{1}{ds/dt} = \frac{d\alpha}{dt}\frac{1}{|d\alpha/dt|}.$$

Thus $\left|\frac{d\beta}{ds}\right| = 1$. □

EXAMPLE 3.13.2. The curve $\alpha(t) = (t^3, t^2)$ is smooth but not regular. Its graph exhibits a cusp. In this case it is impossible to find an arclength parametrisation of the curve.

EXAMPLE 3.13.3. Let $f(x) = \frac{1}{3}(2+x^2)^{3/2}$. Find (an implicit form of) arc length parametrisation of the graph of f.

REMARK 3.13.4 (Curves in $\mathbb{R}^3$). A space curve may be written in coordinates as

$$\alpha(s) = (\alpha^1(s), \alpha^2(s), \alpha^3(s)).$$

Here s is the arc length if $\left|\frac{d\alpha}{ds}\right| = 1$ *i.e.* $\sum_{i=1}^{3}\left(\frac{d\alpha^i}{ds}\right)^2 = 1$.

EXAMPLE 3.13.5. Helix $\alpha(t) = (a\cos\omega t, a\sin\omega t, bt)$.
(i) make a drawing in case $a = b = \omega = 1$.
(ii) parametrize by arc length.
(iii) compute the curvature.

CHAPTER 4

Total curvature, lattices, tori

4.1. Index notation

For a matrix A of size 3×3, the characteristic polynomial $p_A(\lambda)$ has the form $p_A(\lambda) = \lambda^3 - \mathrm{Tr}(A)\lambda^2 + s(A)\lambda - q(A)\lambda^0$. Here $q(A) = \det(A)$ and $s(A) = \lambda_1\lambda_2 + \lambda_1\lambda_3 + \lambda_2\lambda_3$, where λ_i are the eigenvalues of A. By Cayley-Hamilton theorem, we have

$$p_A(A) = 0. \tag{4.1.1}$$

Exercise 4.1.1. Express the equation (4.1.1) in index notation.

4.2. Curvature with respect to an arbitrary parameter

The formula for the curvature of a plane curve is particularly simple with respect to the arclength parameter s (see formula (3.7.1)). However, it can be expressed in terms of an arbitrary parameter t of a regular parametrisation, as well.

Theorem 4.2.1. *With respect to an arbirary parameter t, the formula for the curvature of a regular plane curve $\alpha(t) = (x(t), y(t))$ is*

$$k_\alpha(t) = \frac{|x'y'' - y'x''|}{(x'^2 + y'^2)^{3/2}} \tag{4.2.1}$$

Here $x' = dx/dt$, $y' = dy/dt$, etc. Alternatively, denoting by

$$\alpha' = d\alpha/dt$$

the tangent vector to the curve $\alpha(t) = (x(t), y(t))$, we obtain

$$\kappa = \frac{|x'y'' - y'x''|}{|\alpha'|^3}.$$

For a proof, see (4.4.2).

Remark 4.2.2. In Section 4.4, curvature will be expressed in terms of the angle θ formed by the tangent vector with the positive x-axis.

Example 4.2.3. Calculate the curvature of the graph of $y = f(x)$ at a critical point x_0 using formula (4.2.1).

4.3. Jordan curves in the plane and parameter θ

Recall that a Jordan curve in the plane is a non-selfintersecting closed curve. A Jordan curve can be represented by a continuous one-to-one map

$$\alpha : S^1 \to \mathbb{R}^2.$$

THEOREM 4.3.1 (Jordan curve theorem). *A Jordan curve separates the plane into two open regions: a bounded region and an unbounded region.*

The bounded region is called the *interior* region.

We will only deal with smooth (i.e., infinitely differentiable) regular (i.e, $\alpha' \neq 0$) maps α.

Let $\alpha(s)$ be an arclength parametrisation of a Jordan curve, denoted C, in the plane:

$$C \subset \mathbb{C}.$$

Such a parametrisation exists by Theorem 3.13.1. Let $v(s) = \alpha'(s)$ be its velocity vector at the point $\alpha(s) \in \mathbb{C}$. The vector $v(s)$ can be thought of as a point of the unit circle $S^1 \subset \mathbb{C}$ (by translating the initial point of the vector $v(s)$ to the origin). Therefore v can be thought of as a map to the unit circle called the *Gauss map*, as follows.

The Gauss map is usually defined using the normal vector (see Remark 4.6.3), but for plane curves it can be defined even more simply in terms of the tangent vector, because the two vectors differ by a 90° rotation in the plane.

DEFINITION 4.3.2. The *Gauss map* is the map

$$v : C \to S^1, \quad \alpha(s) \mapsto v(s). \tag{4.3.1}$$

We identify $\mathbb{R}^2$ with $\mathbb{C}$ so that a vector in the plane can be written as a complex number.

DEFINITION 4.3.3 (function theta). We write $v(s) = e^{i\theta(s)}$, where the angle $\theta(s)$ is measured counterclockwise, from the positive ray of the x-axis, to the vector $v(s)$.

Using the complex logarithm, we can also express $\theta(s)$ as follows:

$$\theta(s) = \frac{1}{i}\log v(s) = -i \log v(s).$$

Recall that we have the relation

$$\frac{d}{d\theta}e^{i\theta} = ie^{i\theta}.$$

4.4. Curvature expressed in terms of $\theta(s)$

Let s be an arclength parameter. The tangent vector $v(s)$ to a curve $\alpha(s)$ is expressed as $v(s) = e^{i\theta(s)}$ (see Definition 4.3.3).

THEOREM 4.4.1. *We have the following expression for the curvature k_α of the curve α:*

$$k_\alpha(s) = \left|\frac{d\theta}{ds}\right|. \tag{4.4.1}$$

PROOF. Recall that $v(s) = e^{i\theta(s)}$. By chain rule, we have

$$\frac{dv}{ds} = ie^{i\theta(s)}\frac{d\theta}{ds},$$

and therefore the curvature $k_\alpha(s)$ satisfies

$$k_\alpha(s) = |\alpha''(s)| = \left|\frac{dv}{ds}\right| = \frac{d\theta}{ds},$$

since $|i| = 1$. □

COROLLARY 4.4.2. *Let $\alpha(t) = (x(t), y(t))$ be an arbitrary parametrisation (not necessarily arclength) of a curve. We have the following expression for the curvature k_α:*

$$k_\alpha = \frac{|x'y'' - y'x''|}{|\alpha'|^3} \tag{4.4.2}$$

PROOF. With respect to an arbitrary parameter t, the components of the tangent vector $v(t) = \alpha'(t)$ are $x'(t)$ and $y'(t)$. Hence we have

$$\tan\theta = \frac{y'}{x'}. \tag{4.4.3}$$

Differentiating (4.4.3) with respect to t, we obtain

$$\frac{1}{\cos^2\theta}\frac{d\theta}{dt} = \frac{x'y'' - x''y'}{x'^2}. \tag{4.4.4}$$

Meanwhile

$$x' = \frac{dx}{dt} = \frac{dx}{ds}\frac{ds}{dt} = \cos\theta\frac{ds}{dt}$$

and

$$\frac{d\theta}{dt} = \frac{d\theta}{ds}\frac{ds}{dt}. \tag{4.4.5}$$

Furthermore, $\frac{dt}{ds} = |v(t)| = |\alpha'(t)| = |\frac{d\alpha}{dt}|$. Solving (4.4.5) for $\frac{d\theta}{ds}$, we obtain from (4.4.4) that

$$\frac{d\theta}{ds} = \frac{|x'y'' - y'x''|}{|\alpha'|^3}$$

and by (4.4.1) we obtain (4.4.2). □

4.5. Convexity

DEFINITION 4.5.1. A smooth Jordan curve C is called *strictly convex* if one of the following two equivalent conditions is satisfied:

(1) every interval joining a pair of points of C is contained in the interior region;
(2) consider the tangent line to C at a point $p \in C$; then the complement $C \setminus \{p\}$ lies entirely in one of the open halfplanes defined by the tangent line, for all $p \in C$.

THEOREM 4.5.2. *Assume that the smooth regular Jordan curve is strictly convex, and parametrized counterclockwise. Then the Gauss map $v : C \to S^1$ of* (4.3.1) *is one-to-one and onto, and $\theta(s)$ is monotone increasing.*

PROOF. First we show the "onto" part. First we consider the case of "horizontal" vectors $v(s)$. These occur at points on C with maximal and minimal imaginary part, i.e., the y-coordinate. By applying Rolle's theorem to the function $y(s)$, we obtain that these points have horizontal tangent vectors. Such points correspond to the values of θ equal to 0 and π.

Similarly, to obtain the pair of "opposite" tangent vectors

$$v = e^{i\theta} \text{ and } -v = e^{i(\theta+\pi)},$$

we consider the vector $w_0 = e^{i(\theta+\frac{\pi}{2})}$ orthogonal to v. We then seek the extrema of the the function f given by the scalar product

$$f(s) = \langle \alpha(s), w_0 \rangle$$

where $\alpha(s)$ is a parametrisation of the curve C. At an extremum s_0 of the function f, we have

$$\frac{d}{ds}(f(s))|_{s=s_0} = \left\langle \frac{d\alpha}{ds}, w_0 \right\rangle = 0.$$

Hence the tangent vector $v(s_0) = \alpha'(s_0)$ at each extremum of f is parallel to v. As v ranges over S^1, we thus obtain the points on the curve where the tangent vector is parallel to v.

To show that the Gauss map (4.3.1) is one-to-one, we proceed by contradiction (relying on the law of excluded middle). Suppose on the contrary that two distinct points $p \in C$ and $q \in C$ have the same tangent vector $v = e^{i\theta}$. Then the tangent lines T_p and T_q to C at p and q are parallel. By definition of convexity, the curve C lies on the same side of each of the tangent lines T_p and T_q. Hence the tangent lines must coincide: $T_p = T_q$. Thus, both p and q must lie on the common line $T_p = T_q$. This implies that the arc of the curve between p

and q is contained in C. This contradicts the hypothesis that the curve is strictly convex. The contradiction proves that the map is one-to-one.

Note that a counterclockwise parametrisation corresponds to $\theta(s)$ increasing, while a clockwise parametrisation corresponds to $\theta(s)$ decreasing. □

4.6. Total curvature of a convex Jordan curve

The result on the total curvature[1] of a Jordan curve is of interest in its own right. Furthermore, the result serves to motivate an analogous statement of the Gauss–Bonnet theorem for surfaces in Section 12.3.

DEFINITION 4.6.1. The total curvature $\mathrm{Tot}(C)$ of a curve C with arclenth parametrisation $\alpha(s) : [a, b] \to \mathbb{R}^2$ is the integral

$$\mathrm{Tot}(C) = \int_a^b k_\alpha(s)ds. \tag{4.6.1}$$

When a curve C is closed, i.e, $\alpha(a) = \alpha(b)$, it is more convenient to write the integral (4.6.1) using the notation of a line integral

$$\mathrm{Tot}(C) = \oint_C k_\alpha(s)ds. \tag{4.6.2}$$

THEOREM 4.6.2. *The total curvature of each strictly convex Jordan curve C with arclength parametrisation $\alpha(s)$ equals 2π, i.e.,*

$$\mathrm{Tot}(C) = \oint_C k_\alpha(s)ds = 2\pi.$$

PROOF. Let $\alpha(s)$ be a unit speed parametrisation so that $\alpha(0)$ be the lowest point of the curve, and assume the curve is parametrized counterclockwise. Let $v(s) = \alpha'(s)$. Then $v(0) = e^{i0} = 1$ and $\theta(0) = 0$. The function $\theta = \theta(s)$ is monotone increasing from 0 to 2π (see Theorem 4.5.2). Applying the change of variable formula for integration, we obtain

$$\mathrm{Tot}(C) = \oint_C k_\alpha(s)ds = \oint_C \left|\frac{dv}{ds}\right| ds = \oint_C \frac{d\theta}{ds}ds = \int_0^{2\pi} d\theta = 2\pi,$$

proving the theorem. □

REMARK 4.6.3. We can also consider the normal vector $n(s)$ to the curve. The normal vector satisfies $n(s) = e^{i(\theta+\frac{\pi}{2})} = ie^{i(\theta)}$ and $|\frac{dn}{ds}| = |\frac{dv}{ds}| = \frac{d\theta}{ds}$. Therefore we can also calculate the total curvature as follows:

$$\oint_C k_\alpha(s)ds = \oint_C \left|\frac{dn}{ds}\right| ds = \oint_C \frac{d\theta}{ds}ds = \int_0^{2\pi} d\theta = 2\pi,$$

[1] Akmumiyut kolelet

with $n(s)$ in place of $v(s)$.

Remark 4.6.4. The theorem in fact holds for an arbitrary regular Jordan curve, provided we adjust the definition of curvature to allow for negative sign (in such case $\theta(s)$ will not be a monotone function). We have dealt only with the convex case in order to simplify the topological considerations. See further in Section 4.7.

Remark 4.6.5. A similar calculation will yield the Gauss–Bonnet theorem for convex surfaces in Section 12.3.

4.7. Signed curvature $\tilde{k}_\alpha$ of Jordan curves in the plane

Let $\alpha(s)$ be an arclength parametrisation of a Jordan curve in the plane. We assume that the curve is parametrized counterclockwise. As in Section 4.4, a continuous branch of $\theta(s)$ can be chosen where $\theta(s)$ is the angle measured counterclockwise from the positive x-axis to the tangent vector $v(s) = \alpha'(s)$.

Note that if the Jordan curve is not convex, the function $\theta(s)$ will not be monotone and at certain points its derivative may take negative values: $\theta'(s) < 0$. Once we have such a continuous branch, we can define the signed curvature $\tilde{k}_\alpha$ as follows.

Definition 4.7.1. The signed curvature $\tilde{k}_\alpha(s)$ of a plane Jordan curve $\alpha(s)$ oriented counterclockwise is defined by setting

$$\tilde{k}_\alpha(s) = \frac{d\theta}{ds}.$$

We have the following generalisation of Theorem 4.6.2 on the total curvature of a convex curve.

Theorem 4.7.2. *The total signed curvature $\tilde{k}_\alpha$ of each Jordan curve C with arclength parametrisation $\alpha(s)$ equals 2π, i.e.,*

$$\widetilde{\mathrm{Tot}}(C) = \int_a^b \tilde{k}_\alpha(s)ds = 2\pi.$$

Proof. We exploit the continuous branch $\theta(s)$ as in Theorem 4.6.2, but without the absolute value signs on the derivative of $\theta(s)$. Applying the change of variable formula for integration, we obtain

$$\widetilde{\mathrm{Tot}}(C) = \oint_C \tilde{k}_\alpha(s)ds = \oint_C \frac{d\theta}{ds}ds = \int_0^{2\pi} d\theta = 2\pi,$$

proving the theorem. □

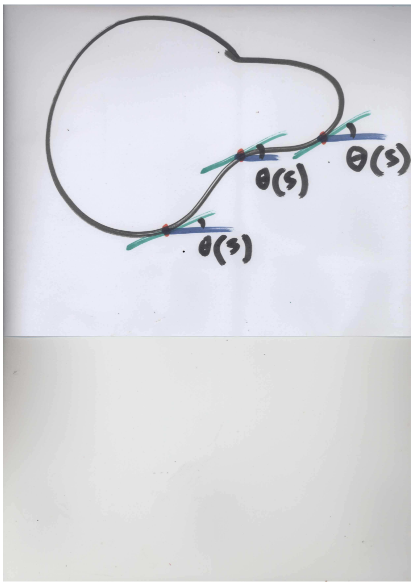

FIGURE 4.7.1. Three points on a curve with the same tangent direction, but the middle one corresponds to $\theta'(s) < 0$.

REMARK 4.7.3. For nonconvex Jordan curves, the Gauss map to the circle will not be one-to-one, but will still have an algebraic degree one. This phenomenon has a far-reaching analogue for imbedded surfaces where the algebraic degree is proportional to the Euler characteristic.

DEFINITION 4.7.4. The *algebraic degree* of a map $\theta : C \to S^1$ at a point $z \in S^1$ is defined to be the sum

$$\sum_{\theta^{-1}(z)} \operatorname{sign}\left(\frac{d\theta}{ds}\right),$$

where the sum runs over all points in the inverse image of z.

For Jordan curves (i.e., imbedded curves), the algebraic degree is always 1. This is the topological ingredient behind the theorem on the total curvature of a Jordan curve.

CHAPTER 5

Rotation index of a closed curve, lattices

5.1. Rotation index of a closed curve in the plane

Given a regular closed plane curve of length L, a continuous branch of $\theta(s)$ can be chosen even if the curve is not simple, obtaining a map

$$\theta : [0, L] \to \mathbb{R}.$$

As in Section 4.7, we can assume that $\theta(0) = 0$. Then $\theta(L)$ is necessarily an integer multiple of 2π. See Millman & Parker [**MP77**, p. 55].

Definition 5.1.1. The rotation index ι_α of a closed unit speed plane curve $\alpha(s)$ is the integer

$$\iota_\alpha = \frac{\theta(L) - \theta(0)}{2\pi}.$$

Example 5.1.2. The rotation index of a figure-8 curve is 0.

Theorem 5.1.3. *Let $\alpha(s)$ be an arclength parametrisation of a geometric curve C. Then the rotation index is related to the total signed curvature $\widetilde{\mathrm{Tot}}(C)$ as follows:*

$$\iota_\alpha = \frac{\widetilde{\mathrm{Tot}}(C)}{2\pi}. \tag{5.1.1}$$

The proof is the same as in Section 4.7. An analogous relation between the Euler characteristic of a surface and its total Gaussian curvature appears in Section 12.4.

5.2. Connected components of curves

Until now we have only considered connected curves. A curve may in general have several connected components.[1]

Definition 5.2.1. Two points p, q on a curve $C \subset \mathbb{R}^2$ are said to lie in the same connected component if there exists a continuous map $h : [0, 1] \to C$ such that $h(0) = p$ and $h(1) = q$.

[1] Rechivei k'shirut.

This defines an equivalence relation on the curve C, and decomposes it as a dijoint union

$$C = \bigsqcup_i C_i$$

Definition 5.2.2. The set of connected components of C is denoted $\pi_0(C)$. The number of connected components is denoted $|\pi_0(C)|$.

Example 5.2.3. Let $F(x, y) = (x^2+y^2-1)((x-10)^2+y^2-1)$, and let C_F be the curve defined by $F(x, y) = 0$. Then C_F is the union of a pair of disjoint circles. Therefore it has two connected components:

$$|\pi_0(C_F)| = 2.$$

The total curvature can be similarly defined for a non-connected curve, by summing the integrals over each connected components.

Definition 5.2.4. The total curvature of a curve $C = \bigsqcup_i C_i$ is

$$\mathrm{Tot}(C) = \sum_i \mathrm{Tot}(C_i).$$

We have the following generalisation of the theorem on total curvature of a curve.

Theorem 5.2.5. *If each connected component of a curve C is a strictly convex Jordan curve, then the total curvature of C is*

$$\mathrm{Tot}(C) = 2\pi\, |\pi_0(C)|.$$

Proof. We apply the previous theorem to each connected component, and sum the resulting total curvatures. □

5.3. Circle via the exponential map

We will discuss lattices and their fundamental domains in Euclidean space $\mathbb{R}^b$ in the next section. Here we give an intuitive introduction in the simplest case $b = 1$.

Every lattice (discrete[2] subgroup; see definition below in Section 5.4) in $\mathbb{R} = \mathbb{R}^1$ is of the form

$$L_\alpha = \alpha\mathbb{Z} \subset \mathbb{R},$$

for some real $\alpha > 0$. It is spanned by the vector αe_1 (or $-\alpha e_1$).

The lattice $L_\alpha \subset \mathbb{R}$ is in fact an additive subgroup. Therefore we can form the group-theoretic quotient $\mathbb{R}/L_\alpha$. This quotient is a circle (see Theorem 5.3.1). A *fundamental domain* for the quotient may be

[2] b'didah

chosen to be the interval $[0, \alpha]$. The 1-volume, *i.e.* the length, of the quotient circle is therefore precisely α.

We will give an equivalent description in terms of the complex function e^z.

THEOREM 5.3.1. *The quotient group $\mathbb{R}/L_\alpha$ is isomorphic to the circle $S^1 \subset \mathbb{C}$.*

PROOF. Consider the map $\hat{\phi} : \mathbb{R} \to \mathbb{C}$ defined by

$$\hat{\phi}(x) = e^{\frac{i2\pi x}{\alpha}}.$$

By the usual addition rule for the exponential function, this map is a homomorphism from the additive structure on $\mathbb{R}$ to the multiplicative structure in the group $\mathbb{C} \setminus \{0\}$. Namely, we have

$$\hat{\phi}(x + y) = \hat{\phi}(x)\hat{\phi}(y) \quad \forall x \forall y \in \mathbb{R}.$$

Furthermore, we have $\hat{\phi}(x + \alpha m) = \hat{\phi}(x)$ for all $m \in \mathbb{Z}$. Thus $L_\alpha = \ker \hat{\phi}$. By the group-theoretic isomorphism theorem, the map $\hat{\phi}$ descends to a map

$$\phi : \mathbb{R}/L_\alpha \to \mathbb{C},$$

which is injective. Its image is the unit circle $S^1 \subset \mathbb{C}$, which is a group under multiplication. $\square$

5.4. Lattice, fundamental domain

Let $b > 0$ be an integer.

DEFINITION 5.4.1. A *lattice* L in Euclidean space $\mathbb{R}^b$ is the integer span of a linearly independent set of b vectors.

Thus, if vectors $v_1, \ldots, v_b$ are linearly independent, then they span a lattice

$$L = \{n_1 v_1 + \ldots + n_b v_b : n_i \in \mathbb{Z}\} = \mathbb{Z}v_1 + \mathbb{Z}v_2 \cdots + \mathbb{Z}v_b$$

Note that the subgroup is isomorphic to $\mathbb{Z}^b$.

DEFINITION 5.4.2. An *orbit* of a point $x_0 \in \mathbb{R}^b$ under the action of a lattice L is the subset of $\mathbb{R}^b$ given by the collection of elements

$$\{x_0 + g \mid g \in L\}.$$

DEFINITION 5.4.3. The quotient

$$\mathbb{R}^b/L$$

(which can be understood at the group-theoretic level as in the case of the circle $\mathbb{R}/L$) is called a b-torus.

DEFINITION 5.4.4. A *fundamental domain* for the torus $\mathbb{R}^b/L$ is a closed set $F \subset \mathbb{R}^b$ satisfying the following three conditions:

- every orbit meets F in at least one point;
- every orbit meets the interior $\mathrm{Int}(F)$ of F in at most one point;
- the boundary ∂F is of zero b-dimensional volume (and can be thought of as a union of $(n-1)$-dimensional hyperplanes).

In the literature, one often replaces "n-dimensional volume" by n-dimensional "Lebesgue measure".

EXAMPLE 5.4.5. The parallelepiped spanned by a collection of basis vectors for L is such a fundamental domain.

More concretely, consider the following example.

EXAMPLE 5.4.6. The vectors e_1 and e_2 in $\mathbb{R}^2$ span the unit square which is a fundamental domain for the lattice $\mathbb{Z}^2 \subset \mathbb{R}^2 = \mathbb{C}$ of Gaussian integers.

EXAMPLE 5.4.7. Consider the vectors $v = (1,0)$ and $w = (\frac{1}{2}, \frac{\sqrt{3}}{2})$ in $\mathbb{R}^2 = \mathbb{C}$. Their span is a parallelogram giving a fundamental domain for the lattice of Eisenstein integers (see Example 5.5.2).

DEFINITION 5.4.8. The *total volume* of the b-torus $\mathbb{R}^b/L$ is by definition the b-volume of a fundamental domain.

It is shown in advanced calculus that the total volume thus defined is independent of the choice of a fundamental domain.

5.5. Lattices in the plane

Let $b = 2$. Every lattice $L \subset \mathbb{R}^2$ is of the form

$$L = \mathrm{Span}_{\mathbb{Z}}(v, w) \subset \mathbb{R}^2,$$

where $\{v, w\}$ is a linearly independent set. For example, let α and β be nonzero reals. Set

$$L_{\alpha,\beta} = \mathrm{Span}_{\mathbb{Z}}(\alpha e_1, \beta e_2) \subset \mathbb{R}^2.$$

This lattice admits an orthogonal basis, namely $\{\alpha e_1, \beta e_2\}$.

EXAMPLE 5.5.1 (Gaussian integers). For the standard lattice $\mathbb{Z}^b \subset \mathbb{R}^b$, the torus $\mathbb{T}^b = \mathbb{R}^b/\mathbb{Z}^b$ satisfies $\mathrm{vol}(\mathbb{T}^b) = 1$ as it has the unit cube as a fundamental domain.

In dimension 2, the resulting lattice in $\mathbb{C} = \mathbb{R}^2$ is called the *Gaussian integers*. It contains 4 elements of least length. These are the fourth roots of unity.

EXAMPLE 5.5.2 (Eisenstein integers). Consider the lattice $L_E \subset \mathbb{R}^2 = \mathbb{C}$ spanned by $1 \in \mathbb{C}$ and the sixth root of unity $e^{\frac{2\pi i}{6}} \in \mathbb{C}$:

$$L_E = \mathrm{Span}_{\mathbb{Z}}(e^{i\pi/3}, 1) = \mathbb{Z}\, e^{i\pi/3} + \mathbb{Z}\, 1. \tag{5.5.1}$$

The resulting lattice is called the *Eisenstein integers*. The torus $\mathbb{T}^2 = \mathbb{R}^2/L_E$ satisfies $\mathrm{area}(\mathbb{T}^2) = \frac{\sqrt{3}}{2}$. The Eisenstein lattice contains 6 elements of least length, namely all the sixth roots of unity.

5.6. Successive minima of a lattice

Let B be Euclidean space, and let $\| \ \|$ be the Euclidean norm. Let $L \subset (B, \| \ \|)$ be a lattice, which is by definition of maximal rank $\mathrm{rank}(L) = \dim(B)$.

EXAMPLE 5.6.1. The first successive minimum, $\lambda_1(L, \| \ \|)$ is the least length of a nonzero vector in L.

We can express the definition symbolically by means of the formula

$$\lambda_1(L, \| \ \|) = \min \left\{ \|v_1\| \ \middle|\ v_1 \in L \setminus \{0\} \right\}.$$

We illustrate the geometric meaning of λ_1 in terms of the circle of Theorem 5.3.1.

THEOREM 5.6.2. *Consider a lattice* $L \subset \mathbb{R}$. *Then the circle* $\mathbb{R}/L$ *satisfies*

$$\mathrm{length}(\mathbb{R}/L) = \lambda_1(L).$$

PROOF. This follows by choosing the fundamental domain $F = [0, \alpha]$ where $\alpha = \lambda_1(L)$, so that $L = \alpha\mathbb{Z}$, cf. Example 5.4 above. □

DEFINITION 5.6.3. For $k = 2$, define the *second successive minimum* of the lattice L with $rank(L) \geq 2$ as follows. Given a pair of vectors $S = \{v, w\}$ in L, define the "length"[3] $|S|$ of S by setting

$$|S| = \max(\|v\|, \|w\|).$$

Then the second successive minimum, $\lambda_2(L, \| \ \|)$ is the least "length" of a pair of non-proportional vectors in L:

$$\lambda_2(L) = \inf_S |S|,$$

where S runs over all linearly independent (*i.e.* non-proportional) pairs of vectors $\{v, w\} \subset L$.

EXAMPLE 5.6.4. For Gaussian and Eisenstein integers, $\lambda_1 = \lambda_2 = 1$.

EXAMPLE 5.6.5. We have $\lambda_1(L_{\alpha,\beta}) = \min(|\alpha|, |\beta|)$ and $\lambda_2(L_{\alpha,\beta}) = \max(|\alpha|, |\beta|)$.

[3]Quotation marks: merka'ot.

5.7. Gram matrix

The volume of the torus $\mathbb{R}^b/L$ (see Definition 5.4.3) is also called the covolume of the lattice L. It is by definition the volume of a fundamental domain for L, *e.g.* a parallelepiped spanned by a $\mathbb{Z}$-basis for L.

DEFINITION 5.7.1. Given a finite set $S = \{v_i\}_{i=1,\ldots,n}$ in $\mathbb{R}^b$, we define its *Gram matrix* as the matrix of inner products

$$Gram(S) = (\langle v_i, v_j \rangle)_{i=1,\ldots,n;\, j=1,\ldots,n}. \tag{5.7.1}$$

THEOREM 5.7.2. *Let $b = n$. The volume of the torus $\mathbb{R}^b/L$ is the square root of the determinant of the Gram matrix of a basis for the lattice L.*

Thus, the parallelepiped P spanned by the vectors $\{v_i\}$ satisfies

$$\mathrm{vol}(P) = \sqrt{\det(Gram(S))}. \tag{5.7.2}$$

PROOF. Let A be the square matrix whose columns are the column vectors $v_1, v_2 \ldots, v_n$ in $\mathbb{R}^n$. It is shown in linear algebra that

$$\mathrm{vol}(P) = |\det(A)|.$$

Let $B = A^t A$, and let $B = (b_{ij})$. Then

$$b_{ij} = v_i^t\, v_j = \langle v_i, v_j \rangle$$

Hence $B = Gram(S)$. Thus

$$\det(Gram(S)) = \det(A^t\, A) = \det(A)^2 = \mathrm{vol}(P)^2$$

proving the theorem. □

5.8. Sphere and torus as topological surfaces

The topology of surfaces will be discussed in more detail in Chapter 16.17. For now, we will recall that a compact surface can be either orientable or non-orientable. An orientable surface is characterized topologically by its *genus*, i.e. number of "handles".

Recall that the unit sphere in $\mathbb{R}^3$ can be represented implicitly by the equation

$$x^2 + y^2 + z^2 = 1.$$

Parametric representations of surfaces are discussed in Section 7.2.

EXAMPLE 5.8.1. The sphere has genus 0 (no handles).

THEOREM 5.8.2. *The 2-torus is characterized topologically in one of the following four equivalent ways:*

(1) *the Cartesian product of a pair of circles: $S^1 \times S^1$;*

(2) *the surface of revolution in* $\mathbb{R}^3$ *obtained by starting with the following circle in the* (x, z)*-plane:* $(x-2)^2 + z^2 = 1$ *(for example), and rotating it around the* z*-axis;*
(3) *a quotient* $\mathbb{R}^2/L$ *of the plane by a lattice* L*;*
(4) *a compact* 2*-dimensional manifold of genus* 1.

The equivalence between items (2) and (3) can be seen by marking a pair of generators of L by different color, and using the same colors to indicate the corresponding circles on the imbedded torus of revolution, as follows:

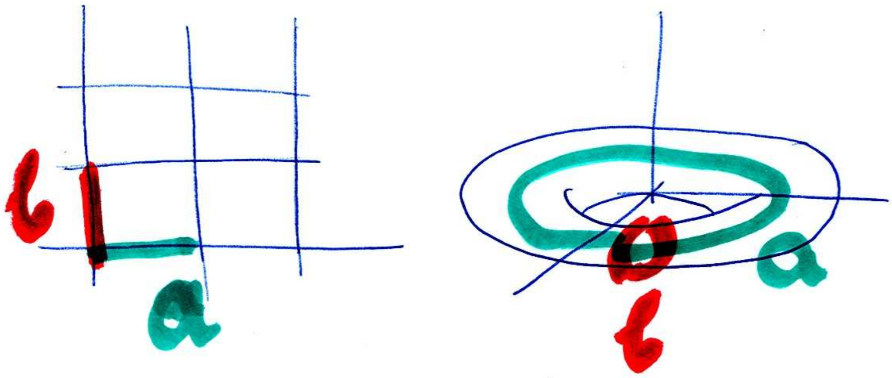

FIGURE 5.8.1. Torus viewed by means of its lattice (left) and by means of a Euclidean imbedding (right)

Note by comparison that a circle can be represented either by its fundamental domain which is $[0, 2\pi]$ (with endpoints identified), or as the unit circle imbedded in the plane.

CHAPTER 6

Hermite constant, Clairaut's relation

6.1. Hermite constant

The Hermite constant γ_b is defined in one of the following two equivalent ways:

(1) γ_b is the *square* of the maximal first successive minimum λ_1, among all lattices of unit covolume;
(2) γ_b is defined by the formula

$$\sqrt{\gamma_b} = \sup \left\{ \frac{\lambda_1(L)}{\mathrm{vol}(\mathbb{R}^b/L)^{\frac{1}{b}}} \,\middle|\, L \subseteq (\mathbb{R}^b, \| \,\|) \right\}, \tag{6.1.1}$$

where the supremum is extended over all lattices L in $\mathbb{R}^b$ with a Euclidean norm $\| \,\|$.

Remark 6.1.1 (Dense packings). A lattice realizing the supremum may be thought of as the one realizing the densest packing in $\mathbb{R}^b$ when we place the balls of radius $\frac{1}{2}\lambda_1(L)$ at the points of L.

6.2. Standard fundamental domain

We will discuss the case $b = 2$ in detail. An important role is played in this dimension by the standard fundamental domain.

Definition 6.2.1. The *standard fundamental domain*, denoted D, is the set

$$D = \left\{ z \in \mathbb{C} \,\middle|\, |z| \geq 1,\ |\mathrm{Re}(z)| \leq \frac{1}{2},\ \mathrm{Im}(z) > 0 \right\} \tag{6.2.1}$$

cf. [**Ser73**, p. 78].

The domain D a fundamental domain for the action of $PSL(2, \mathbb{Z})$ in the upperhalf plane of $\mathbb{C}$.

Lemma 6.2.2. *Multiplying a lattice $L \subset \mathbb{C}$ by nonzero complex numbers does not change the value of the quotient*

$$\frac{\lambda_1(L)^2}{\mathrm{area}(\mathbb{C}/L)}.$$

PROOF. We write such a complex number as $re^{i\theta}$. Note that multiplication by $re^{i\theta}$ can be thought of as a composition of a scaling by the real factor r, and rotation by angle θ. The rotation is an isometry (congruence) that preserves all lengths, and in particular the length $\lambda_1(L)$ and the area of the quotient torus.

Meanwhile, multiplication by r results in a cancellation

$$\frac{\lambda_1(rL)^2}{\mathrm{area}(\mathbb{C}/rL)} = \frac{(r\lambda_1(L))^2}{r^2\,\mathrm{area}(\mathbb{C}/L)} = \frac{r^2\lambda_1(L)^2}{r^2\,\mathrm{area}(\mathbb{C}/L)} = \frac{\lambda_1(L)^2}{\mathrm{area}(\mathbb{C}/L)},$$

proving the lemma. □

6.3. Conformal parameter τ

Two lattices in $\mathbb{C}$ are said to be *similar* if one is obtained from the other by multiplication by a nonzero complex number.

THEOREM 6.3.1. *Every lattice in $\mathbb{C}$ is similar to a lattice spanned by $\{\tau, 1\}$ where τ is in the standard fundamental domain D of* (6.2.1). *The value $\tau = e^{i\pi/3}$ corresponds to the Eisenstein integers* (5.5.1).

PROOF. Let $L \subset \mathbb{C}$ be a lattice. Choose a "shortest" vector $z \in L$, *i.e.* we have $|z| = \lambda_1(L)$. By Lemma 6.2.2, we may replace the lattice L by the lattice $z^{-1}L$.

Thus, we may assume without loss of generality that the complex number $+1 \in \mathbb{C}$ is a shortest element in the lattice L. Thus we have $\lambda_1(L) = 1$. Now complete the element $+1$ to a $\mathbb{Z}$-basis

$$\{\bar{\tau}, +1\}$$

for L. Here we may assume, by replacing $\bar{\tau}$ by $-\bar{\tau}$ if necessary, that $\mathrm{Im}(\bar{\tau}) > 0$.

Now consider the real part $\mathrm{Re}(\bar{\tau})$. We adjust the basis by adding a suitable integer k to $\bar{\tau}$:

$$\tau = \bar{\tau} - k \quad \text{where} \quad k = \left[\mathrm{Re}(\bar{\tau}) + \tfrac{1}{2}\right] \tag{6.3.1}$$

(the brackets denote the integer part), so it satisfies the condition

$$-\frac{1}{2} \leq \mathrm{Re}(\tau) \leq \frac{1}{2}.$$

Since $\tau \in L$, we have $|\tau| \geq \lambda_1(L) = 1$. Therefore the element τ lies in the standard fundamental domain (6.2.1). □

EXAMPLE 6.3.2. For the "rectangular" lattice $L_{\alpha,\beta} = \mathrm{Span}_{\mathbb{Z}}(\alpha, \beta i)$, we obtain

$$\tau(L_{\alpha,\beta}) = \begin{cases} \frac{|\beta|}{|\alpha|} i & \text{if } |\beta| > |\alpha| \\ \frac{|\alpha|}{|\beta|} i & \text{if } |\alpha| > |\beta|. \end{cases}$$

COROLLARY 6.3.3. *Let* $b = 2$. *Then we have the following value for the Hermite constant:* $\gamma_2 = \frac{2}{\sqrt{3}} = 1.1547....$ *The corresponding optimal lattice is homothetic to the* $\mathbb{Z}$*-span of cube roots of unity in* $\mathbb{C}$ *(i.e. the Eisenstein integers).*

PROOF. Choose τ as in (6.3.1) above. The pair

$$\{\tau, +1\}$$

is a basis for the lattice. The imaginary part satisfies $\mathrm{Im}(\tau) \geq \frac{\sqrt{3}}{2}$, with equality possible precisely for

$$\tau = e^{i\frac{\pi}{3}} \text{ or } \tau = e^{i\frac{2\pi}{3}}.$$

Moreover, if $\tau = r\exp(i\theta)$, then

$$\sin\theta = \frac{\mathrm{Im}(\tau)}{|\tau|} \geq \frac{\sqrt{3}}{2}.$$

The proof is concluded by calculating the area of the parallelogram in $\mathbb{C}$ spanned by τ and $+1$;

$$\frac{\lambda_1(L)^2}{\mathrm{area}(\mathbb{C}/L)} = \frac{1}{|\tau|\sin\theta} \leq \frac{2}{\sqrt{3}},$$

proving the theorem. □

DEFINITION 6.3.4. A $\tau \in D$ is said to be the *conformal parameter* of a flat torus T^2 if T^2 is similar to a torus $\mathbb{C}/L$ where $L = \mathbb{Z}\tau + \mathbb{Z}1$.

EXAMPLE 6.3.5. In dimensions $b \geq 3$, the Hermite constants are harder to compute, but explicit values (as well as the associated critical lattices) are known for small dimensions, *e.g.* $\gamma_3 = 2^{\frac{1}{3}} = 1.2599...$, while $\gamma_4 = \sqrt{2} = 1.4142....$

6.4. Spherical coordinates

Spherical coordinates will be useful in understanding surfaces of revolution (see Section 7.5).

DEFINITION 6.4.1. Spherical coordinates (ρ, θ, φ) in $\mathbb{R}^3$ are defined by the following formulas. We have

$$\rho = \sqrt{x^2 + y^2 + z^2} = \sqrt{r^2 + z^2}$$

is the distance to the origin, while φ is the angle with the z-axis, so that $\cos\varphi = z/\rho$. Here θ is the angle inherited from polar coordinates in the x, y plane, so that $\tan\theta = \frac{y}{x}$.

REMARK 6.4.2. The interval of definition for the variable φ is $\varphi \in [0, \pi]$ since $\varphi = \arccos\frac{z}{\rho}$ and the range of the arccos function is $[0, \pi]$. Meanwhile $\theta \in [0, 2\pi]$ as usual.

The unit sphere $S^2 \subset \mathbb{R}^3$ is defined by

$$S^2 = \{\rho = 1\}.$$

A *latitude*[1] on the unit sphere is a circle satisfying the equation

$$\varphi = \text{constant}.$$

A latitude is parallel to the equator. The equator $\varphi = \pi/2$ is the only latitude that's a great circle. A latitude can be parametrized by setting $\theta(t) = t, \varphi(t) = \text{constant}$. Note that on the unit sphere $\{\rho = 1\}$, we have

$$r = \sin\varphi, \tag{6.4.1}$$

where r is the distance to the z-axis.

6.5. Great circles parametrized and implicit

In Section 8.4, we will encounter the geodesic equation. This is a system of nonlinear second order differential equations.

We seek to provide a geometric intuition for this equation. We will establish a connection between solutions of this system, on the one hand, and spherical great circles, on the other. Such a connection is established via the intermediary of Clairaut's relation for a variable point on a great circle:

$$r(t)\cos\gamma(t) = \text{const}$$

where r is the distance to the (vertical) axis of revolution and γ is the angle with the latitudinal circle (*cf.* Theorem 6.6.1). In this section, we will verify Clairaut's relation synthetically for great circles, and also show that the latter satisfy a first order differential equation. In Section 8.4 we will complete the connection by deriving Clairaut's relation from the geodesic equation.

Let $S^2 \subset \mathbb{R}^3$ be the unit sphere defined by the equation

$$x^2 + y^2 + z^2 = 1.$$

A plane P through the origin is given by an equation

$$ax + by + cz = 0, \tag{6.5.1}$$

[1] kav rochav

where the vector

$$\begin{pmatrix} a \\ b \\ c \end{pmatrix} \neq \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

is nonzero.

DEFINITION 6.5.1. A great circle G of S^2 is given by an intersection

$$G = S^2 \cap P.$$

EXAMPLE 6.5.2 (A parametrisation of the equator of S^2).

$$\alpha(t) = (\cos t)e_1 + (\sin t)e_2.$$

In general, every great circle can be parametrized by

$$\alpha(t) = (\cos t)v + (\sin t)w$$

where v and w are orthonormal.

EXAMPLE 6.5.3 (an implicit (non-parametric) representation of a great circle). Recall that we have

$$x = r\cos\theta = \rho\sin\varphi\cos\theta; \; y = \rho\sin\varphi\sin\theta; \; z = \rho\cos\varphi\,. \tag{6.5.2}$$

If the circle lies in the plane $ax+by+cz=0$ where a, b, c are fixed, the great circle in coordinates (θ, φ) is defined implicitly by the equation

$$a\sin\varphi\cos\theta + b\sin\varphi\sin\theta + c\cos\theta = 0,$$

as in (6.5.1).

LEMMA 6.5.4. *Scalar product of vector valued functions satisfies Leibniz's rule:*

$$\langle f, g\rangle' = \langle f', g\rangle + \langle f, g'\rangle. \tag{6.5.3}$$

PROOF. See [**Leib**]. In more detail, let (f_1, f_2) be components of f, and let (g_1, g_2) be components of g. Then

$$\langle f, g\rangle' = (f_1g_1 + f_2g_2)' = f_1g_1' + f_1'g_1 + f_2g_2' + f_2'g_2 = \langle f', g\rangle + \langle f, g'\rangle,$$

completing the proof. □

LEMMA 6.5.5. *Let $\alpha : \mathbb{R} \to S^2$ be a parametrized curve on the sphere $S^2 \subset \mathbb{R}^3$. Then the tangent vector $\frac{d\alpha}{dt}$ is perpendicular to the position vector $\alpha(t)$:*

$$\left\langle \alpha(t), \frac{d\alpha}{dt} \right\rangle = 0.$$

PROOF. We have $\langle \alpha(t), \alpha(t) \rangle = 1$ by definition of S^2. We apply the operator $\frac{d}{dt}$ to obtain

$$\frac{d}{dt}\langle \alpha(t), \alpha(t) \rangle = 0.$$

Next, we apply Leibniz's rule (6.5.3) to obtain

$$\Big\langle \alpha(t), \frac{d\alpha}{dt} \Big\rangle + \Big\langle \frac{d\alpha}{dt}, \alpha(t) \Big\rangle = 2\Big\langle \alpha(t), \frac{d\alpha}{dt} \Big\rangle = \frac{d}{dt}(1) = 0,$$

completing the proof. □

6.6. Clairaut's relation

THEOREM 6.6.1 (Clairaut's relation). *Let $\alpha(t)$ be a regular parametrisation of a great circle G on $S^2 \subset \mathbb{R}^3$. Let $r(t)$ denote the distance from $\alpha(t)$ to the z-axis, and let $\gamma(t)$ denote the angle between the tangent vector $\dot\alpha(t)$ to the curve and the vector tangent to the latitude at the point $\alpha(t)$. Then*

$$r(t) \cos \gamma(t) = \text{const}.$$

Here the constant has value $\text{const} = r_{\min}$, *where $r_{\min}$ is the least distance from a point of G to the z-axis.*

DEFINITION 6.6.2. A *spherical triangle* is the following collection of data: the vertices are points of S^2, the sides are arcs of great circles, while the angles are defined to be the angles between tangent vectors to the sides. Here we assume that

(1) all sides have length $< \pi$,
(2) the three vertices do not lie on a common great circle.

REMARK 6.6.3 (Spherical sine law).

$$\frac{\sin a}{\sin \alpha} = \frac{\sin b}{\sin \beta} = \frac{\sin c}{\sin \gamma}$$

EXAMPLE 6.6.4. Let $\gamma = \frac{\pi}{2}$, then $\sin a = \sin c \sin \gamma$. Note that for small values of a, c we recapture the Euclidean formula

$$a = c \sin \alpha$$

as the limiting case.

DEFINITION 6.6.5. A longitude (meridian)[2] is (half) a great circle passing through the North Pole $e_3 = (0, 0, 1)$.

[2]kav orech

LEMMA 6.6.6. *Assume G is not a longitude (meridian) and let $p \in G$ be the point with the maximal z-coordinate among the points of G. Then the great circle G has the following equivalent properties:*

(1) *it is perpendicular at p to the longitude (meridian) passing through p, and*
(2) *G is tangent at p to the latitudinal circle.*

PROOF. Let $\alpha(t)$ be a parametrisation of G with $\alpha(0) = p$. Since the function $\langle \alpha(t), e_3 \rangle$ achieves its maximum at $t = 0$, we have

$$\frac{d}{dt}\bigg|_0 \langle \alpha(t), e_3 \rangle = 0.$$

By Leibniz's rule,

$$\left\langle \frac{d\alpha}{dt}\bigg|_0, e_3 \right\rangle = -\left\langle \alpha(t), \frac{de_3}{dt} \right\rangle = 0,$$

since e_3 is constant. Let $\dot{\alpha}(t) = \dfrac{d\alpha}{dt}$. Thus $\langle \dot{\alpha}(0), e_3 \rangle = 0$. Also $\langle \dot{\alpha}(0), p \rangle = 0$ by Lemma 6.5.5. Since the tangent vector to the longitude (meridian) through p lies in the plane spanned by p and e_3, the lemma follows. □

6.7. Proof of Clairaut's relation

PROOF OF CLAIRAUT'S RELATION. Note that we have an angle of $\frac{\pi}{2} - \gamma$ between $\dot{\alpha}(t)$ and the vector tangent to the longitude (meridian) at $\alpha(t)$ (this is equivalent to the vanishing of the metric coefficient g_{12} for a surface of revolution).

Let $p \in S^2$ be the point of $\alpha(t)$ with maximal z-coordinate. Note that if $\varphi(t)$ is the spherical coordinate φ of $\alpha(t)$, then $r(t) = \sin \varphi(t)$. Consider the spherical triangle with vertices $\alpha(t)$, p, and the north pole e_3.

By Lemma 6.6.6, the angle at p is $\dfrac{\pi}{2}$. Hence by the law of sines,

$$\sin c \sin\left(\frac{\pi}{2} - \gamma\right) = \sin b,$$

where

- $c = \varphi(t) =$ arc of longitude (meridian) joining $\alpha(t)$ to e_3;
- $b =$ arc of longitude (meridian) joining p to e_3.

Note that $r = \sin c$ by (6.4.1). Since b is independent of t, the theorem is proved. □

Note that the parametrisation in Clairaut's formula need not be arclength. Assume G is not a longitude (meridian), so we can parametrize it by the value of the spherical coordinate θ. Note that $\rho \equiv 1$.

By the implicit function theorem, we can think of G as defined by a function

$$\varphi = \varphi(\theta).$$

THEOREM 6.7.1. *The great circle G satisfies the differential equation*

$$\frac{1}{r^2} + \frac{1}{r^4}\left(\frac{d\varphi}{d\theta}\right)^2 = \frac{1}{\mathrm{const}^2}, \tag{6.7.1}$$

where $r = \sin\varphi$ *and* $\mathrm{const} = \sin\varphi_{\min}$ *from Theorem 6.6.1.*

PROOF. An 'element of length', ds, along C decomposes into a longitudinal (along a longitude (meridian), north-south) displacement $d\varphi$, and a latitudinal (east-west) displacement $rd\theta$. These are related by

$$ds\sin\gamma = d\varphi$$
$$ds\cos\gamma = rd\theta$$

so that $ds^2 = (d\varphi)^2 + (rd\theta)^2$. Hence

$$rd\theta\tan\gamma = ds\sin\gamma = d\varphi,$$

or $\tan\gamma = \frac{d\varphi}{rd\theta}$. Hence $\cos^2\gamma = \frac{1}{1+\left(\frac{d\varphi}{rd\theta}\right)^2}$. Therefore by Clairaut's relation (Theorem 6.6.1), we obtain

$$\left(\frac{\mathrm{const}}{r}\right)^2\left(1+\left(\frac{d\varphi}{rd\theta}\right)^2\right) = 1. \tag{6.7.2}$$

Equivalently,

$$1+\left(\frac{d\varphi}{rd\theta}\right)^2 = \left(\frac{r}{\mathrm{const}}\right)^2$$

or

$$\left(\frac{d\varphi}{rd\theta}\right)^2 = \frac{r^2}{\mathrm{const}^2} - 1$$

or

$$\frac{d\varphi}{rd\theta} = \sqrt{\frac{r^2}{\mathrm{const}^2} - 1}$$

or

$$\frac{1}{r}\frac{d\varphi}{d\theta} = \sqrt{\frac{r^2}{\mathrm{const}^2} - 1}$$

or

$$\frac{1}{\sin\varphi}\frac{d\varphi}{d\theta} = \sqrt{\frac{\sin^2\varphi}{\mathrm{const}^2} - 1} \tag{6.7.3}$$

Equivalently,

$$\frac{1}{r^2} + \frac{1}{r^4}\left(\frac{d\varphi}{d\theta}\right)^2 = \frac{1}{\text{const}^2} \qquad \text{where} \quad r = \sin\varphi.$$

This equation is solved explicitly in terms of integrals in Example 8.4.3 below. □

Remark 6.7.2. At a point where φ is not extremal as a function of θ, theorem on the uniqueness of solution applies and gives a unique geodesic through the point.

However, at a point of maximal φ, the hypothesis of the uniqueness theorem does not apply. Namely, the square root expression on the right hand side of (6.7.3) does not satisfy the Lipschitz condition as the expression under the square root sign vanishes. In fact, uniqueness fails at this point, as a latitude (which is not a geodesic) satisfies the differential equation, as well. Here we have $r = \text{const}$, and at an extremal value of φ one can no longer solve the equation by separation of variables (as this would involve division by the radical expression which vanishes at the extremal value of φ). At this point, there is a degeneracy and general results about uniqueness of solution cannot be applied.

CHAPTER 7

Local geometry of surfaces; first fundamental form

7.1. Local geometry of surfaces

The differential geometry of surfaces in Euclidean 3-space starts with the observation that they inherit a metric structure from the ambient space (i.e. the Euclidean space). We would like to understand which geometric properties of this structure are intrinsic, in a sense to be clarified.

Remark 7.1.1. Following [**Ar74**, Appendix 1, p. 301], note that a piece of paper may be placed flat on a table, or it may be rolled into a cylinder, or it may be rolled into a cone. However, it cannot be transformed into the surface of a sphere, that is, without tearing or stretching. Understanding this phenomenon quantitatively is our goal, *cf.* Figure 11.6.1.

7.2. Regular surface; Jacobian

Our starting point is the first fundamental form, obtained by restricting the 3-dimensional inner product. What does the first fundamental form measure? A helpful observation to keep in mind is that it allows one to measure the length of curves on the surface.

Consider a surface $M \subset \mathbb{R}^3$ parametrized by a map $\underline{x}(u^1, u^2)$ or

$$\underline{x} : \mathbb{R}^2 \to \mathbb{R}^3. \tag{7.2.1}$$

We will always assume that $\underline{x}$ is differentiable. From now on we will frequently omit the underline in $\underline{x}$, and write simply "x", to lighten the notation. Denote by J_x its Jacobian matrix, *i.e.* the 3×2 matrix

$$J_x = \left(\frac{dx^i}{du^j} \right), \tag{7.2.2}$$

where x^i are the three components of the vector valued function $\underline{x}$.

Definition 7.2.1. The parametrisation $\underline{x}$ of the surface M is called *regular* if one of the following equivalent conditions is satisfied:

(1) the vectors $\frac{\partial x}{\partial u^1}$ and $\frac{\partial x}{\partial u^2}$ are linearly independent;
(2) the Jacobian matrix (7.2.2) is of rank 2.

EXAMPLE 7.2.2. Let $\beta > 0$ be fixed, and consider the function $f(x,y) = \sqrt{\beta^2 - x^2 - y^2}$. The graph of f can be parametrized as follows:

$$\underline{x} = (u^1, u^2, f(u^1, u^2)).$$

This provides a parametrisation of the (open) northern hemisphere. Then $\frac{\partial \underline{x}}{\partial u^1} = (1, 0, f_x)^t$ while $\frac{\partial \underline{x}}{\partial u^2} = (0, 1, f_y)^t$. Note that we have

$$f_x = \frac{-x}{f} = \frac{-x}{\sqrt{\beta^2 - x^2 - y^2}}, \quad f_y = \frac{-y}{f}. \tag{7.2.3}$$

The southern hemisphere can be similarly parametrized by

$$\underline{x}' = (u^1, u^2, -f(u^1, u^2)).$$

The formulas for the $\frac{\partial \underline{x}}{\partial u^i}$ are then $\frac{\partial \underline{x}}{\partial u^1} = (1, 0, -f_x)^t = (1, 0, \frac{x}{\sqrt{\beta^2 - x^2 - y^2}})$, while $\frac{\partial \underline{x}}{\partial u^2} = (0, 1, -f_y)^t = (0, 1, \frac{y}{\sqrt{\beta^2 - x^2 - y^2}})$.

7.3. First fundamental form of a surface

Let $\langle\, , \rangle$ denote the inner product in $\mathbb{R}^3$. For $i = 1, 2$ and $j = 1, 2$, define functions $g_{ij} = g_{ij}(u^1, u^2)$ called "metric coefficients" by

$$g_{ij}(u^1, u^2) = \left\langle \frac{\partial x}{\partial u^i}, \frac{\partial x}{\partial u^j} \right\rangle. \tag{7.3.1}$$

REMARK 7.3.1. We have $g_{ij} = g_{ji}$ as the inner product is symmetric.

EXAMPLE 7.3.2 (Metric coefficients for the graph of a function). The graph of a function f satisfies

$$\left\langle \frac{\partial \underline{x}}{\partial u^1}, \frac{\partial \underline{x}}{\partial u^1} \right\rangle = 1 + f_x^2,$$

$$\left\langle \frac{\partial \underline{x}}{\partial u^1}, \frac{\partial \underline{x}}{\partial u^2} \right\rangle = f_x f_y,$$

$$\left\langle \frac{\partial \underline{x}}{\partial u^2}, \frac{\partial \underline{x}}{\partial u^2} \right\rangle = 1 + f_y^2.$$

Therefore in this case we have

$$(g_{ij}) = \begin{pmatrix} 1 + f_x^2 & f_x f_y \\ f_x f_y & 1 + f_y^2 \end{pmatrix}$$

In the case of the hemisphere, the partial derivatives are as in formula (7.2.3).

EXAMPLE 7.3.3 (Metric coefficients in spherical coordinates). Consider the parametrisation of the unit sphere as a function of spherical coordinates, so that $\underline{x} = \underline{x}(\theta, \varphi)$. We have $x = \sin\varphi\cos\theta$, while $y = \sin\varphi\sin\theta$, and $z = \cos\varphi$. Setting $u^1 = \theta$ and $u^2 = \varphi$, we obtain

$$\frac{\partial \underline{x}}{\partial u^1} = (-\sin\varphi\sin\theta, \sin\varphi\cos\theta, 0)$$

and

$$\frac{\partial \underline{x}}{\partial u^2} = (\cos\varphi\cos\theta, \cos\varphi\sin\theta, -\sin\varphi),$$

Thus in this case we have

$$(g_{ij}) = \begin{pmatrix} \sin^2\varphi & 0 \\ 0 & 1 \end{pmatrix}$$

so that $\det(g_{ij}) = \sin^2\varphi$ and $\sqrt{\det(g_{ij})} = \sin\varphi$.

DEFINITION 7.3.4. The vectors $x_i = \dfrac{\partial x}{\partial u^i}$, $i = 1, 2$, are called the tangent vectors to the surface M.

THEOREM 7.3.5. *The matrix (g_{ij}) is the Gram matrix of the pair of tangent vectors:*

$$(g_{ij}) = J_x^T \, J_x,$$

cf. formula (5.7.2).

The proof is immediate.

DEFINITION 7.3.6. The plane spanned by the vectors $x_1(u^1, u^2)$ and $x_2(u^1, u^2)$ is called the *tangent plane* to the surface M at the point $p = x(u^1, u^2)$, and denoted

$$T_pM.$$

DEFINITION 7.3.7. The *first fundamental form* I_p of the surface M at p is the bilinear form on T_p defined by setting

$$I_p : T_pM \times T_pM \to \mathbb{R},$$

defined by the restriction of the ambient Euclidean inner product:

$$I_p(v, w) = \langle v, w \rangle_{\mathbb{R}^3},$$

for all $v, w \in T_pM$. With respect to the basis (frame) $\{x_1, x_2\}$, it is given by the matrix (g_{ij}), where

$$g_{ij} = \langle x_i, x_j \rangle.$$

The coefficients g_{ij} are sometimes called 'metric coefficients.'

Like curves, surfaces can be represented either implicitly or parametrically.

7.4. Plane, cylinder

The example of the sphere was discussed in the previous section. We now consider two more examples.

EXAMPLE 7.4.1. The x, y-plane in $\mathbb{R}^3$ is defined implicitly by $z = 0$. Let $\underline{x}(u^1, u^2) = (u^1, u^2, 0) \in \mathbb{R}^3$. This is a parametrisation of the xy-plane in $\mathbb{R}^3$. Then $x_1 = (1, 0, 0)^t$ and $x_2 = (0, 1, 0)^t$. Then

$$g_{11} = \langle x_1, x_1 \rangle = \left\langle \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \right\rangle = 1,$$

etcetera. Thus we have $(g_{ij}) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = I.$

EXAMPLE 7.4.2. Let $x(u^1, u^2) = (\cos u^1, \sin u^1, u^2)$. This formula provides a parametrisation of the cylinder. We have

$$x_1 = (-\sin u^1, \cos u^1, 0)^t$$

and $x_2 = (0, 0, 1)^t$, while

$$g_{11} = \left\langle \begin{pmatrix} -\sin u^1 \\ \cos u^1 \\ 0 \end{pmatrix}, \begin{pmatrix} -\sin u^1 \\ \cos u^1 \\ 0 \end{pmatrix} \right\rangle = \sin^2 u^1 + \cos^2 u^1 = 1,$$

etc. Thus $(g_{ij}) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = I.$

REMARK 7.4.3. The first fundamental form does **not** contain all the information (even up to Euclidean congruences) about the surface. Indeed, the plane and the cylinder have the same first fundamental form, but are geometrically distinct imbedded surfaces.

7.5. Surfaces of revolution

EXAMPLE 7.5.1 (Surfaces of revolution). Here it is customary to use the notation $u^1 = \theta$ and $u^2 = \phi$. The starting point is a curve C in the xz-plane, parametrized by a pair of functions

$$x = f(\phi), \quad z = g(\phi).$$

The surface of revolution (around the z-axis) defined by C is parametrized as follows:

$$\underline{x}(\theta, \phi) = (f(\phi) \cos \theta, f(\phi) \sin \theta, g(\phi)). \tag{7.5.1}$$

If we start with the vertical line $f(\phi) = 1$, $g(\phi) = \phi$, the resulting surface of revolution is the cylinder. For $f(\phi) = \sin \phi$ and $g(\phi) = \cos \phi$, we obtain the sphere S^2 in spherical coordinates.

THEOREM 7.5.2. *If ϕ is the arclength parameter of the curve C, then the first fundamental form of the corresponding surface of revolution* (7.5.1) *is given by*

$$(g_{ij}) = \begin{pmatrix} f^2 & 0 \\ 0 & 1 \end{pmatrix}.$$

PROOF. We have

$$x_1 = \frac{\partial x}{\partial \theta} = (-f\sin\theta, f\cos\theta, 0),$$
$$x_2 = \frac{\partial x}{\partial \phi} = \left(\frac{df}{d\phi}\cos\theta, \frac{df}{d\phi}\sin\theta, \frac{dg}{d\phi}\right)$$

therefore

$$g_{11} = \left| \begin{matrix} -f\sin\theta \\ f\cos\theta \\ 0 \end{matrix} \right|^2 = f^2\sin^2\theta + f^2\cos^2\theta = f^2$$

and

$$g_{22} = \left| \begin{pmatrix} \frac{df}{d\phi}\cos\theta \\ \frac{df}{d\phi}\sin\theta \\ \frac{dg}{d\phi} \end{pmatrix} \right|^2 = \left(\frac{df}{d\phi}\right)^2 (\cos^2\theta + \sin^2\theta) + \left(\frac{dg}{d\phi}\right)^2 = \left(\frac{df}{d\phi}\right)^2 + \left(\frac{dg}{d\phi}\right)^2$$

and

$$g_{12} = \left\langle \begin{pmatrix} -f\sin\theta \\ f\cos\theta \\ 0 \end{pmatrix}, \begin{pmatrix} \frac{df}{d\phi}\cos\theta \\ \frac{df}{d\phi}\sin\theta \\ \frac{dg}{d\phi} \end{pmatrix} \right\rangle = -f\frac{df}{d\phi}\sin\theta\cos\theta + f\frac{df}{d\phi}\cos\theta\sin\theta = 0.$$

Thus

$$(g_{ij}) = \begin{pmatrix} f^2 & 0 \\ 0 & \left(\frac{df}{d\phi}\right)^2 + \left(\frac{dg}{d\phi}\right)^2 \end{pmatrix}.$$

In the case of an arclength parametrisation, we obtain $g_{22} = 1$, proving the theorem. □

EXAMPLE 7.5.3. Consider the curve $(\sin\phi, \cos\phi)$ in the x, z-plane. The resulting surface of revolution is the sphere, where the ϕ parameter coincides with the angle φ of spherical coordinates. Thus, for the sphere S^2 we obtain

$$(g_{ij}) = \begin{pmatrix} \sin^2\phi & 0 \\ 0 & 1 \end{pmatrix}.$$

7.6. Pseudosphere

EXAMPLE 7.6.1. The pseudosphere (so called because its Gaussian curvature equals -1) is the surface of revolution generated by a curve called the *tractrix*, parametrized by (f, g). Here $f(\phi) = e^{\phi}$ and

$$g(\phi) = \int_0^{\phi} \sqrt{1 - e^{2\psi}} d\psi,$$

where $-\infty < \phi \leq 0$. The tractrix generates a surface of revolution with $g_{11} = e^{2\phi}$, while

$$\begin{aligned} g_{22} &= (e^{\phi})^2 + (\sqrt{1 - e^{2\phi}})^2 \\ &= e^{2\phi} + 1 - e^{2\phi} = 1. \end{aligned}$$

Thus $(g_{ij}) = \begin{pmatrix} e^{2\phi} & 0 \\ 0 & 1 \end{pmatrix}$.

EXERCISE 7.6.2. Compute the coefficients g_{ij} for the standard parametrisation of the graph of $z = f(x, y)$ by $(u^1, u^2, f(u^1, u^2))$.

7.7. Einstein summation convention

In the next section we will use again the Einstein summation convention. Let's review some exercises exploiting this.

EXERCISE 7.7.1. A matrix A is called *idempotent* if $A^2 = A$. Write the idempotency condition in indices with Einstein summation convention (without Σ's).

EXERCISE 7.7.2. Matrices A and B are *similar* if there exists an invertible matrix P such that $AP - PB = 0$. Write the similarity condition in indices, as the vanishing of the (i, j)th coefficient of the difference $AP - PB$.

7.8. Measuring length of curves on surfaces

Let us now explain how to measure the length of curves on a surface in terms of the metric coefficients of the surface. Let

$$\alpha : [a, b] \to \mathbb{R}^2, \quad \alpha(t) = (\alpha^1(t), \alpha^2(t))$$

be a plane curve.

THEOREM 7.8.1. *Let $\beta(t) = \underline{x} \circ \alpha(t)$ be a curve on the surface $\underline{x}$. Then the length L of β is given by the formula*

$$L = \int_a^b \sqrt{g_{ij} \frac{d\alpha^i}{dt} \frac{d\alpha^j}{dt}} dt.$$

PROOF. The length L of β is calculated as follows using chain rule:

$$\begin{aligned} L &= \int_a^b \left| \frac{d\beta}{dt} \right| dt = \int_a^b \left| \sum_{i=1}^2 \frac{\partial x}{\partial u^i} \frac{d\alpha^i}{dt} \right| dt \\ &= \int_a^b \left\langle x_i \frac{d\alpha^i}{dt}, x_j \frac{d\alpha^j}{dt} \right\rangle^{1/2} dt \\ &= \int_a^b \left(\langle x_i, x_j \rangle \frac{d\alpha^i}{dt} \frac{d\alpha^j}{dt} \right)^{1/2} dt = \\ &= \int_a^b \sqrt{g_{ij}(\alpha(t)) \frac{d\alpha^i}{dt} \frac{d\alpha^j}{dt}} dt. \end{aligned}$$

Thus

$$L = \int_a^b \sqrt{g_{ij}(\alpha(t)) \frac{d\alpha^i}{dt} \frac{d\alpha^j}{dt}} dt.$$

□

EXAMPLE 7.8.2. If $g_{ij} = \delta_{ij}$ is the identity matrix, then the length of the curve $\beta = \underline{x} \circ \alpha$ on the surface equals the length of the curve α of $\mathbb{R}^2$. Indeed, If $g_{ij} = \delta_{ij}$, then $L = \displaystyle\int_a^b \sqrt{\left(\frac{d\alpha^1}{dt} \right)^2 + \left(\frac{d\alpha^2}{dt} \right)^2} dt =$ the length of $\alpha(t)$ in $\mathbb{R}^2$. The parametrisation in this case is an isometry.

To simplify notation, let $x = x(t) = \alpha^1(t)$ and $y = y(t) = \alpha^2(t)$. Then the infinitesimal element of arclength is

$$ds = \sqrt{dx^2 + dy^2},$$

and the length of the curve is

$$\int ds = \int \sqrt{dx^2 + dy^2}.$$

7.9. The symbols Γ_{ij}^k of a surface

The symbols Γ_{ij}^k, roughly speaking,[1] account for how the surface twists in space. They are, however, coordinate dependent and have no intrinsic geometric meaning. We will see that the symbols Γ_{ij}^k also control the behavior of geodesics on the surface. Here geodesics can be thought of as curves that are to the surface what straight lines are to a plane, or what great circles are to a sphere, *cf.* Definition 8.3.2.

Let $x : \mathbb{R}^2 \to \mathbb{R}^3$ be a regular parametrized surface, and let $x_i = \dfrac{\partial x}{\partial u^i}$.

[1]In the first approximation: kiruv rishon.

Definition 7.9.1. The normal vector $n(u^1, u^2)$ to a regular surface at the point $x(u^1, u^2) \in \mathbb{R}^3$ is defined in terms of the vector product, *cf.* Definition 1.9.3, as follows:

$$n = \frac{x_1 \times x_2}{|x_1 \times x_2|},$$

so that $\langle n, x_i \rangle = 0 \quad \forall i$.

The vectors x_1, x_2, n form a basis (frame) for $\mathbb{R}^3$ (since x is regular by hypothesis). Recall that the g_{ij} were defined in terms of first partial derivatives of x. Meanwhile, the symbols Γ^k_{ij} are defined in terms of the second partial derivatives (since they are not even tensors, one does not stagger the indices). Namely, they are the coefficients of the decomposition of the second partial derivative vector x_{ij}, defined by

$$x_{ij} = \frac{\partial^2 x}{\partial u^i \partial u^j}$$

with respect to the basis, or frame,[2] (x_1, x_2, n) :

Definition 7.9.2. The symbols Γ^k_{ij} are uniquely determined by the formula

$$x_{ij} = \Gamma^1_{ij} x_1 + \Gamma^2_{ij} x_2 + L_{ij} n$$

(the coefficients L_{ij} will be discussed in Section 10.1).

7.10. Basic properties of the symbols Γ^k_{ij}

Proposition 7.10.1. *We have the following formula for the symbols:*

$$\Gamma^k_{ij} = \langle x_{ij}, x_\ell \rangle g^{\ell k},$$

where (g^{ij}) *is the inverse matrix of* g_{ij}.

Proof. We have

$$\langle x_{ij}, x_\ell \rangle = \langle \Gamma^k_{ij} x_k + L_{ij} n, x_\ell \rangle = \langle \Gamma^k_{ij} x_k, x_\ell \rangle + \langle L_{ij} n, x_\ell \rangle = \Gamma^k_{ij} g_{k\ell}.$$

We now multiply by $g^{\ell m}$ and sum:

$$\langle x_{ij}, x_\ell \rangle g^{\ell m} = \Gamma^k_{ij} g_{k\ell} g^{\ell m} = \Gamma^k_{ij} \delta^m_k = \Gamma^m_{ij}.$$

This is equivalent to the desired formula. □

Remark 7.10.2. We have the following relation: $\Gamma^k_{ij} = \Gamma^k_{ji}$, or $\Gamma^k_{[ij]} = 0 \quad \forall ijk$.

---

[2] Maarechet yichus

EXAMPLE 7.10.3 (The plane). Calculation of the symbols Γ_{ij}^k for the plane $x(u^1, u^2) = (u^1, u^2, 0)$. We have $x_1 = (1, 0, 0)$, $x_2 = (0, 1, 0)$. Thus we have $x_{ij} = 0 \quad \forall i, j$. Hence $\Gamma_{ij}^k \equiv 0 \quad \forall i, j, k$.

EXAMPLE 7.10.4 (The cylinder). Calculation of the symbols for the cylinder $x(u^1, u^2) = (\cos u^1, \sin u^1, u^2)$. The normal vector is $n = (\cos u^1, \sin u^1, 0)$, while $x_1 = (-\sin u^1, \cos u^1, 0)$, $x_2 = (0, 0, 1)$. Thus we have $x_{22} = 0$, $x_{21} = 0$ and so $\Gamma_{22}^k = 0$ and $\Gamma_{12}^k = \Gamma_{21}^k = 0 \quad \forall k$.

Meanwhile, $x_{11} = (-\cos u^1, -\sin u^2, 0)$. This vector is proportional to n:

$$x_{11} = 0x_1 + 0x_2 + (-1)n.$$

Hence $\Gamma_{11}^k = 0 \quad \forall k$.

DEFINITION 7.10.5. We will use the following notation for the partial derivative of g_{ij}:

$$g_{ij;k} = \frac{\partial}{\partial u^k}(g_{ij}).$$

LEMMA 7.10.6. *In terms of the symmetrisation notation introduced in Section 1.7, we have*

$$g_{ij;k} = 2g_{m\{i}\Gamma_{j\}k}^m.$$

PROOF. Indeed,

$$g_{ij;k} = \partial u^k \langle x_i, x_j \rangle = \langle x_{ik}, x_j \rangle + \langle x_i, x_{jk} \rangle = \langle \Gamma_{ik}^m x_m, x_j \rangle + \langle \Gamma_{jk}^m x_m, x_i \rangle.$$

By definition of the metric coefficients, we have

$$g_{ij;k} = \Gamma_{ik}^m g_{mj} + \Gamma_{jk}^m g_{mi} = g_{mj}\Gamma_{ik}^m + g_{mi}\Gamma_{jk}^m = 2g_{m\{j}\Gamma_{i\}k}^m$$

or $2g_{m\{i}\Gamma_{j\}k}^m$. □

An important role in the theory is played by the intrinsic nature of the coefficients Γ. Namely, we will prove that they are determined by the metric coefficients alone, and are therefore independent of the ambient (extrinsic) geometry of the surface, i.e., the way it "sits" in 3-space.

7.11. Intrinsic nature of the symbols Γ_{ij}^k

THEOREM 7.11.1. *The symbols Γ_{ij}^k can be expressed in terms of the first fundamental form and its derivatives as follows :*

$$\Gamma_{ij}^k = \frac{1}{2}(g_{i\ell;j} - g_{ij;\ell} + g_{j\ell;i})g^{\ell k},$$

where g^{ij} is the inverse matrix of g_{ij}.

Γ^1_{ij}	$j=1$	$j=2$
$i=1$	0	$\frac{1}{f}\frac{df}{d\phi}$
$i=2$	$\frac{1}{f}\frac{df}{d\phi}$	0

TABLE 7.12.1. Symbols Γ^k_{ij} of a surface of revolution (7.12.1)

PROOF. Applying Lemma 7.10.6 three times, we obtain

$$\begin{aligned} g_{i\ell;j} - g_{ij;\ell} + g_{j\ell;i} &= 2g_{m\{i}\Gamma^m_{\ell\}j} - 2g_{m\{i}\Gamma^m_{j\}\ell} + 2g_{m\{j}\Gamma^m_{\ell\}i} \\ &= g_{mi}\Gamma^m_{\ell j} + g_{m\ell}\Gamma^m_{ij} - g_{mi}\Gamma^m_{j\ell} - g_{mj}\Gamma^m_{i\ell} + g_{mj}\Gamma^m_{\ell i} + g_{m\ell}\Gamma^m_{ji} \\ &= 2g_{m\ell}\Gamma^m_{ji}. \end{aligned}$$

Thus $\frac{1}{2}(g_{i\ell;j}-g_{ij;\ell}+g_{j\ell;i})g^{\ell k} = \Gamma^m_{ij}g_{m\ell}g^{\ell k} = \Gamma^m_{ij}\delta^k_m = \Gamma^k_{ij}$, as required. □

7.12. Symbols Γ^k_{ij} for a surface of revolution

Recall that a surface of revolution is obtained by starting with a curve $f(\phi), g(\phi)$ in the (x,z) plane, and rotating it around the z-axis, obtaining the parametrisation $x(\theta,\phi) = (f(\phi)\cos\theta, f(\phi)\sin\theta, g(\phi))$. Thus we adopt the notation

$$u^1 = \theta, \quad u^2 = \phi.$$

LEMMA 7.12.1. *For a surface of revolution we have* $\Gamma^1_{11} = \Gamma^1_{22} = 0$, *while* $\Gamma^1_{12} = \frac{f\frac{\mathrm{d}f}{\mathrm{d}\phi}}{f^2}$, *cf. Table 7.12.1.*

PROOF. For the surface of revolution

$$x(\theta,\phi) = (f(\phi)\cos\theta, f(\phi)\sin\theta, g(\phi)), \tag{7.12.1}$$

the metric coefficients are given by the matrix

$$(g_{ij}) = \begin{pmatrix} f^2 & 0 \\ 0 & \left(\frac{df}{d\phi}\right)^2 + \left(\frac{dg}{d\phi}\right)^2 \end{pmatrix}.$$

Since the off-diagonal coefficients $g_{12} = 0$ vanish, the coefficients of the inverse matrix satisfy

$$g^{ii} = \frac{1}{g_{ii}}. \tag{7.12.2}$$

We have $\dfrac{\partial}{\partial\theta}(g_{ii}) = 0$ since g_{ii} depend only on ϕ. Thus the terms

$$g_{ii;1} = 0 \tag{7.12.3}$$

vanish. Let us now compute the symbols Γ^1_{ij} for $k = 1$. Using formulas (7.12.2) and (7.12.3), we obtain

$$\begin{aligned}\Gamma^1_{11} &= \frac{1}{2g_{11}}(g_{11;1} - g_{11;1} + g_{11;1}) \text{ by formula (7.12.2)}\\ &= 0 \text{ by formula (7.12.3).}\end{aligned}$$

Similarly,

$$\begin{aligned}\Gamma^1_{12} &= \frac{1}{2g_{11}}(g_{11;2} - g_{12;1} + g_{12;1}) = \frac{g_{11;2}}{2g_{11}}\\ &= \frac{\frac{\mathrm{d}}{\mathrm{d}\phi}(f^2)}{2f^2} = \frac{f\frac{\mathrm{d}f}{\mathrm{d}\phi}}{f^2},\end{aligned}$$

while $\Gamma^1_{22} = \frac{1}{2g_{11}}(g_{12;2} - g_{22;1} + g_{12;2}) = \frac{g_{12;2}}{g_{11}} = \frac{\frac{\mathrm{d}}{\mathrm{d}\phi}(0)}{g_{11}} = 0.$ □

CHAPTER 8

Conformally equivalent metrics, geodesic equation

8.1. Metrics conformal to the standard flat metric

A particularly important class of metrics are those conformal to the flat metric, in the following sense. We will use the symbol

$$\lambda$$

for the conformal factor of the metric, as below.

LEMMA 8.1.1. *Consider a metric whose coefficients are of the form*

$$g_{ij} = \lambda(u^1, u^2)\delta_{ij}.$$

Then we have $\Gamma^1_{11} = \frac{\lambda_1}{2\lambda}$, $\Gamma^1_{22} = \frac{-\lambda_1}{2\lambda}$, *and* $\Gamma^1_{12} = \frac{\lambda_2}{2\lambda}$

The values of the coefficients are listed in Table 8.1.1.

PROOF. By hypothesis, we have $g_{11} = g_{22} = \lambda(u^1, u^2)$ while $g_{12} = 0$. We have

$$\Gamma^1_{ij} = \frac{1}{2\lambda}(g_{i1;j} - g_{ij;1} + g_{j1;i}),$$

and the lemma follows by examining the cases. $\square$

Γ^1_{ij}	$j = 1$	$j = 2$
$i = 1$	$\frac{\lambda_1}{2\lambda}$	$\frac{\lambda_2}{2\lambda}$
$i = 2$	$\frac{\lambda_2}{2\lambda}$	$\frac{-\lambda_1}{2\lambda}$

Γ^2_{ij}	$j = 1$	$j = 2$
$i = 1$	$-\frac{\lambda_2}{2\lambda}$	$\frac{\lambda_1}{2\lambda}$
$i = 2$	$\frac{\lambda_1}{2\lambda}$	$\frac{\lambda_2}{2\lambda}$

TABLE 8.1.1. Symbols Γ^k_{ij} of a conformal metric $\lambda\delta_{ij}$

8.2. Geodesics on a surface

What is a geodesic on a surface?

A geodesic on a surface can be thought of as the path of an ant crawling along the surface of an apple, according to *Gravitation*, a Physics textbook [**MiTW73**, p. 3]. Imagine that we peel off a narrow strip of the apple's skin along the ant's trajectory, and then lay it out flat on a table. What we obtain is a straight line, revealing the ant's ability to travel along the shortest path.

On the other hand, a geodesic is defined by a certain nonlinear second order ordinary differential equation, *cf.* (8.3.1). To make the geodesic equation more concrete, we will examine the case of the surfaces of revolution. Here the geodesic equation transforms into a conservation law (conservation of angular momentum)[1] called Clairaut's relation. The latter lends itself to a synthetic verification for spherical great circles, as in Theorem 6.6.1.

We will derive the geodesic equation using the calculus of variations. I once heard R. Bott point out a surprising aspect of M. Morse's foundational work in this area. Namely, Morse systematically used the length functional on the space of curves. The simple idea of using the energy functional instead of the length functional was not exploited until later. The use of energy simplifies calculations considerably, as we will see in Section 10.3.

8.3. Geodesic equation

Consider a plane curve $\underset{s}{\mathbb{R}} \underset{\alpha}{\to} \underset{(u^1,u^2)}{\mathbb{R}^2}$ where $\alpha = (\alpha^1(s), \alpha^2(s))$. If $x : \mathbb{R}^2 \to \mathbb{R}^3$ is a parametrisation of a surface M, the composition

$$\underset{s}{\mathbb{R}} \underset{\alpha}{\to} \underset{(u^1,u^2)}{\mathbb{R}^2} \underset{x}{\to} \mathbb{R}^3$$

yields a curve

$$\beta = x \circ \alpha$$

on M.

Proposition 8.3.1. *Every regular curve $\beta(t)$ on M satisfies the identity*

$$\beta'' = \left(\alpha^{i'}\alpha^{j'}\Gamma^k_{ij} + \alpha^{k''}\right) x_k + \left(L_{ij}\alpha^{j'}\alpha^{i'}\right) n$$

Proof. Write $\beta = x \circ \alpha$, then $\beta' = x_i \alpha^{i'}$. We have

$$\beta'' = \frac{d}{ds}(x_i \circ \alpha)\alpha^{i'} + x_i \alpha^{i''} = x_{ij}\alpha^{j'}\alpha^{i'} + x_k \alpha^{k''}.$$

[1] tena' zaviti

Meanwhile $x_{ij} = \Gamma_{ij}^k x_k + L_{ij} n$, proving the proposition. □

DEFINITION 8.3.2. A curve $\beta = x \circ \alpha$ is a geodesic on the surface x if the one of the following two equivalent conditions is satisfied:

(a) we have for each $k = 1, 2$,

$$(\alpha^k)'' + \Gamma_{ij}^k (\alpha^i)' (\alpha^j)' = 0 \quad \text{where} \quad ' = \frac{d}{ds}\,, \tag{8.3.1}$$

meaning that

$$(\forall k) \quad \frac{d^2\alpha^k}{ds^2} + \Gamma_{ij}^k \frac{d\alpha^i}{ds}\frac{d\alpha^j}{ds} = 0;$$

(b) the vector β'' is perpendicular to the surface and one has

$$\beta'' = L_{ij} {\alpha^i}' {\alpha^j}' n. \tag{8.3.2}$$

REMARK 8.3.3. The equations (8.3.1) will be derived using the calculus of variations, in Section 10.3. Furthermore, by Lemma 8.3.5, such a curve β must have constant speed.

PROOF OF EQUIVALENCE. If β is a geodesic, then applying Proposition 8.3.1, we obtain

$$\beta'' = L_{ij} {\alpha^i}' {\alpha^j}' n,$$

showing that the vector β'' is perpendicular to (every tangent vector of) the surface. On the other hand, if β'' is proportional to the normal vector n, then the tangential component of β'' must vanish, proving (8.3.2). □

EXAMPLE 8.3.4. In the plane, all coefficients $\Gamma_{ij}^k = 0$ vanish. Then the equation becomes $(\alpha^k)'' = 0$, *i.e.* $\alpha(s)$ is a linear function of s. In other words, the graph is a straight line. Thus, all geodesics in the plane are straight lines.

LEMMA 8.3.5. *A curve β satisfying the geodesic equation* (8.3.1) *is necessarily constant speed.*

PROOF. From formula (8.3.2), we have

$$\frac{d}{ds}\left(\|\beta'\|^2\right) = 2\langle \beta'', \beta' \rangle = L_{ij} {\alpha^i}' {\alpha^j}' {\alpha^k}' \langle n, x_k \rangle = 0,$$

proving the lemma. □

8.4. Geodesics on a surface of revolution

In the case of a surface of revolution, it is convenient to use the notation $u^1 = \theta, u^2 = \phi$ for the coordinates. For the purposes of this section, we will replace $f(\phi)$ by $r(\phi)$. This function gives the distance from a point on the surface, to the z-axis.

LEMMA 8.4.1. *The angle γ between a curve β and the latitude satisfies* $\cos\gamma = r\theta'$.

PROOF. The tangent vector to the latitude is x_1 with $|x_1| = r = f$. We have

$$\cos\gamma = \left\langle \frac{x_1}{|x_1|}, \frac{d\beta}{ds} \right\rangle \quad \text{where} \quad x_1 = \frac{\partial x}{\partial \theta}(\theta, \phi).$$

Thus

$$\begin{aligned} \cos\gamma &= \frac{1}{|x_1|}\langle x_1, \underbrace{x_1\theta' + x_2\phi'}_{\text{chain rule}}\rangle = \frac{\theta'}{|x_1|}\underbrace{\langle x_1, x_1\rangle}_{|x_1|^2} + \frac{\phi'}{|x_2|}\underbrace{\langle x_1, x_2\rangle}_{g_{12}=0} \\ &= \theta'|x_1| = r\theta', \end{aligned}$$

proving the lemma. □

THEOREM 8.4.2. *Let $x = r(\phi)\cos\theta$, $y = r(\phi)\sin\theta$, and $z = g(\phi)$ be the components of a surface of revolution. Then the geodesic equation for $k = 1$ is equivalent to Clairaut's relation* $r\cos\gamma = \text{const}$ *(cf. Theorem 6.6.1).*

PROOF. The symbols for $k = 1$ are given by Lemma 7.12.1. We will use the shorthand notation $\theta(s), \phi(s)$ respectively for $\alpha^1(s), \alpha^2(s)$. Let $' = \frac{d}{ds}$. The equation of geodesic $\beta = x \circ \alpha$ for $k = 1$ becomes

$$\begin{aligned} 0 &= \theta'' + 2\Gamma_{12}^1\theta'\phi' \\ &= \theta'' + \frac{2r\frac{dr}{d\phi}}{r^2}\theta'\phi' \\ &= r^2\theta'' + 2r\frac{dr}{d\phi}\theta'\phi' \\ &= (r^2\theta')'. \end{aligned}$$

Integrating, we obtain

$$r^2\theta' = \text{const.} \tag{8.4.1}$$

Since the curve β is constant speed by Lemma 8.3.5, we can assume the paramenter s is arclength.

By Lemma 8.4.1 we have $r\cos\gamma = r^2\theta' = \text{const}$ from (8.4.1), proving Clairaut's relation. □

The rest of the material in this section is optional.

In the case of a surface of revolution, the geodesic equation can be integrated by means of primitives :

$$1 = \left\langle \frac{dx}{ds}, \frac{dx}{ds} \right\rangle = \langle x_1\theta^{'} + x_2\phi^{'}, x_1\theta^{'} + x_2\phi^{'} \rangle$$

$$= g_{11}(\theta^{'})^2 + g_{22}(\phi^{'})^2 = f^2(\theta^{'})^2 + \left(\underbrace{\left(\frac{df}{d\phi}\right)^2 + \left(\frac{dg}{d\phi}\right)^2}_{g_{22}} \right) (\phi^{'})^2.$$

Thus, $1 = f^2 \left(\frac{d\theta}{ds}\right)^2 + g_{22}\left(\frac{d\phi}{ds}\right)^2$. Now multiply by $\left(\frac{ds}{d\theta}\right)^2$:

$$\left(\frac{ds}{d\theta}\right)^2 = f^2 + g_{22}\left(\frac{d\phi}{ds}\frac{ds}{d\theta}\right)^2 = f^2 + g_{22}\left(\frac{d\phi}{d\theta}\right)^2.$$

Now from Clairaut's relation we have

$$\frac{ds}{d\theta} = \frac{f^2}{c},$$

hence we obtain the formula $\frac{f^4}{c^2} = f^2 + g_{22}\left(\frac{d\phi}{d\theta}\right)^2$ or

$$\frac{1}{c^2} = \frac{1}{f^2} + \frac{g_{22}}{f^4}\left(\frac{d\phi}{d\theta}\right)^2$$

which is the equation (6.7.1) for a geodesic on the sphere. Furthermore, we have

$$\frac{d\phi}{d\theta} = \sqrt{\frac{\frac{f^4}{c^2} - f^2}{g_{22}}} = \frac{f}{c}\sqrt{\frac{f^2 - c^2}{\left(\frac{df}{d\phi}\right)^2 + \left(\frac{dg}{d\phi}\right)^2}}.$$

Thus

$$\theta = c\int \frac{1}{f}\sqrt{\frac{\left(\frac{df}{d\phi}\right)^2 + \left(\frac{dg}{d\phi}\right)^2}{f^2 - c^2}}d\phi + \text{const}.$$

EXAMPLE 8.4.3. On S^2, we have $f(\phi) = \sin\phi, g(\phi) = \cos\phi$, thus we obtain

$$\theta = c\int \frac{d\phi}{\sin\phi\sqrt{\sin^2\phi - c^2}} + \text{const},$$

which is the equation of a great circle in spherical coordinates.

Following some preliminaries on areas, directional derivatives, and Hessians, we will deal with a central object in the differential geometry of surfaces in Euclidean space, namely the Weingarten map, in Section 9.4.

8.5. Polar, cylindrical, spherical coordinates; integration

In this section, we review material from calculus on polar, cylindrical, and spherical coordinates.

The polar coordinates (r, θ) in the plane arise naturally in complex analysis (of one complex variable).

Definition 8.5.1. Polar coordinates (koordinatot koteviot) (r, θ) satisfy $r^2 = x^2 + y^2$ and $x = r\cos\theta$, $y = r\sin\theta$.

It is shown in elementary calculus that the area of a region D in the plane in polar coordinates is calculated using the area element

$$dA = r\,dr\,d\theta.$$

Thus, an integral is of the form

$$\int_D dA = \iint r dr d\theta.$$

Cylindrical coordinates in Euclidean 3-space are studied in Vector Calculus.

Definition 8.5.2. Cylindrical coordinates (koordinatot gliliot)

$$(r, \theta, z)$$

are a natural extension of the polar coordinates (r, θ) in the plane.

The volume of an open region D is calculated with respect to cylindrical coordinates using the volume element

$$dV = r\,dr\,d\theta\,dz.$$

Namely, an integral is of the form

$$\int_D dV = \iiint r dr\,d\theta\,dz.$$

Example 8.5.3. Find the volume of a right circular cone with height h and base a circle of radius b.

Spherical coordinates[2]

$$(\rho, \theta, \varphi)$$

in Euclidean 3-space are studied in Vector Calculus.

[2]koordinatot kaduriot

DEFINITION 8.5.4. Spherical coordinates (ρ, θ, φ) are defined as follows. The coordinate ρ is the distance from the point to the origin, satisfying

$$\rho^2 = x^2 + y^2 + z^2,$$

or $\rho^2 = r^2+z^2$, where $r^2 = x^2+y^2$. If we project the point orthogonally to the (x, y)-plane, the polar coordinates of its image, (r, θ), satisfy $x = r\cos\theta$ and $y = r\sin\theta$. The last coordiate φ of the spherical coordinates is the angle between the position vector of the point and the third basis vector e_3 in 3-space (pointing upward along the z-axis). Thus

$$z = \rho\cos\varphi$$

while

$$r = \rho\sin\varphi\,.$$

Here we have the bounds $0 \leq \rho$, $0 \leq \theta \leq 2\pi$, and $0 \leq \varphi \leq \pi$ (note the different upper bounds for θ and φ). Recall that the area of a spherical region D is calculated using a volume element of the form

$$dV = \rho^2 \sin\varphi \;d\rho\, d\theta\, d\,\varphi,$$

so that the volume of a region D is

$$\int_D dV = \iiint_D \rho^2 \sin\varphi \;d\rho\, d\theta\, d\,\varphi\,.$$

EXAMPLE 8.5.5. Calculate the volume of the spherical shell between spheres of radius $\rho_0 > 0$ and $\rho_1 \geq \rho_0$.

The area of a spherical region on a sphere of radius $\rho = \rho_1$ is calculated using the area element

$$dA = \rho_1^2 \sin\varphi \;d\theta\, d\,\varphi\,.$$

Thus the area of a spherical region D on a sphere of radius ρ_1 is given by the integral

$$\int_D dA = \iint \rho_1^2 \sin\varphi \;d\theta\, d\,\varphi = \rho_1^2 \iint \sin\varphi \;d\theta\, d\,\varphi$$

EXAMPLE 8.5.6. Calculate the area of the spherical region on a sphere of radius ρ_1 contained in the first octant, (so that all three Cartesian coordinates are positive).

8.6. Measuring area on surfaces

DEFINITION 8.6.1 (Computation of area). The *area* of the surface $x : U \to \mathbb{R}^3$ is computed by integrating the expression

$$\sqrt{\det(g_{ij})}du^1du^2 \tag{8.6.1}$$

over the domain U of the map x. Thus $\text{area}(x) = \int_U \sqrt{\det(g_{ij})}du^1du^2$.

The presence of the square root in the formula is explained in infinitesimal calculus in terms of the Gram matrix, *cf.* (5.7.1).

EXAMPLE 8.6.2. Consider the parametrisation given by spherical coordinates on the unit sphere. Then the integrand is

$$\sin\varphi \; d\theta \; d\varphi,$$

and we recover the formula familiar from calculus for the area of a region D on the unit sphere:

$$\text{area}(D) = \iint_D \sin\varphi \; d\theta \; d\varphi \,.$$

CHAPTER 9

Directional derivative and Weingarten map

9.1. Directional derivative

We will represent a vector v in the $\mathbb{R}^n$ plane as the velocity vector $v = \frac{d\alpha}{dt}$ of a curve $\alpha(t)$, at $t = 0$. Typically we will be interested in the case $n = 3$ (or 2).

Definition 9.1.1. Given a function f of n variables, its directional derivative[1] $\nabla_v f$ at a point p, in the direction of a vector v is defined by setting

$$\nabla_v f = \left. \frac{d\left(f \circ \alpha(t)\right)}{dt} \right|_{t=0}.$$

Lemma 9.1.2. *The definition of directional derivative is independent of the choice of the curve $\alpha(t)$ representing the vector v.*

Proof. The lemma is proved in Elementary calculus [**Ke74**]. □

Let $p = \underline{x}(u^1, u^2)$ be a point of a surface in $\mathbb{R}^3$.

Definition 9.1.3. The tangent plane to the surface $\underline{x} = \underline{x}(u^1, u^2)$ at the point p is denoted T_p and is defined "naively" to be the plane passing through p and spanned by vectors x_1 and x_2, or alternatively as the plane perpendicular to the normal vector n at p.

Thus we have an orthogonal decomposition

$$\mathbb{R}^3 = T_p \overset{\perp}{\oplus} \mathbb{R}n.$$

Example 9.1.4. Suppose $\underline{x}(u^1, u^2)$ is a parametrisation of the unit sphere $S^2 \subset \mathbb{R}^3$. At a point $(a, b, c) \in S^2$, the normal vector is the position vector itself: $n = \begin{pmatrix} a \\ b \\ c \end{pmatrix}$. In other words, $n(u^1, u^2) = \underline{x}(u^1, u^2)$.

[1] nigzeret kivunit

9.2. Extending vector field along surface to open set in $\mathbb{R}^3$

Now let us return to the set-up

$$\underset{t}{\mathbb{R}} \underset{\alpha}{\rightarrow} \underset{(u^1,u^2)}{\mathbb{R}^2} \underset{\underline{x}}{\rightarrow} \mathbb{R}^3.$$

We consider the curve $\beta = x \circ \alpha$. Let $v \in T_p$ be a tangent vector at a point $p \in M$, defined by $v = \frac{d\beta}{dt}\big|_{t=0}$, where $\beta(0) = p$. By chain rule, $v = \frac{d\alpha^i}{dt} x_i$. The normal vector $n \circ \alpha(t)$ along $\beta(t)$ varies from point to point on the surface, but is not defined in an open neighnorhood in $\mathbb{R}^3$. We would like to extned it to a vector field in an open neighborhood of $p \in \mathbb{R}^3$. The curve β represents the class of curves with initial vector v.

LEMMA 9.2.1. *the gradient ∇F of a function $F = F(x, y, z)$ is perpendicular to the level surface $F(x, y, z) = 0$ of F.*

This was shown in elementary calculus.

THEOREM 9.2.2. *Consider a regular parametrisation $x(u^1, u^2)$ as before, as well as its normal vector $n = n(u^1, u^2)$. One can extend n to a vector field $N(x, y, z)$ defined in an open neighborhood of $p \in \mathbb{R}^3$, so that we have*

$$n(u^1, u^2) = N\left(\underline{x}(u^1, u^2)\right). \tag{9.2.1}$$

PROOF. We apply a version of the implicit function theorem for surfaces to represent the surface implicitly by an equation $F(x, y, z) = 0$, where F is defined in an open neighborhood of $p \in M$, and $\nabla F \neq 0$ at p. By Lemma 9.2.1, the normalisation

$$\frac{1}{|\nabla F|} \nabla F$$

of the gradient ∇F of F gives the required extension. □

PROPOSITION 9.2.3. *Let $p \in M$, and $v \in T_pM$ where $v = \beta'(0)$. Let N be the vector field extending $n(u^1, u^2)$. Then the directional derivative $\nabla_v N$ satisfies*

$$\nabla_v N = \left.\frac{d\left(n \circ \alpha(t)\right)}{dt}\right|_{t=0}.$$

PROOF. By (9.2.1), the function $n(t)$ satisfies the relation

$$n(\alpha(t)) = N(\beta(t)),$$

where $\beta = x \circ \alpha$. The gradient $\nabla_v N$ can be calculated using this particular curve β. By Lemma 9.1.2, the gradient is independent of the choice of the curve. We therefore obtain

$$\nabla_v N = \frac{d\,(N \circ \beta(t))}{dt} = \frac{d\,(n \circ \alpha(t))}{dt},$$

proving the proposition. □

9.3. Hessian of a function at a critical point

Consider a function $f(x, y)$ of two variables in the neighborhood of a critical point, where $\nabla f = 0$, and consider its graph.

REMARK 9.3.1. The tangent plane of the graph of f at a critical point of f is a horizontal plane.

The Hessian (matrix of second derivatives) of the function at the critical point captures the main features of the behavior of the function in a neighborhood of the critical point. Thus, we have the following typical result concerning the surface given by the graph of the function in $\mathbb{R}^3$.

THEOREM 9.3.2. *At a critical point of f, assume that the eigenvalues of the Hessian are nonzero. If the eigenvalues of the Hessian have opposite sign, then the graph of f is a saddle point. If the eigenvalues have the same sign, the graph is a local minimum or maximum.*

REMARK 9.3.3. If one thinks of the Hessian as a linear transformation (an endomorphism) of the horizontal plane, given by the matrix of second derivatives, then the Hessian of a function at a critical point becomes a special case of the *Weingarten* map, defined in the next section.

9.4. From Hessian to Weingarten map

Now consider the more general framework of a parametrized regular surface M in $\mathbb{R}^3$.

REMARK 9.4.1. Instead of working with a matrix of second derivatives, we will give a definition of an endomorphism of the tangent plane in a coordinate-free fashion.

We extend n to a vector field $N(x, y, z)$ defined in an open neighborhood of $p \in \mathbb{R}^3$, so that we have

$$n(u^1, u^2) = N\left(x(u^1, u^2)\right).$$

DEFINITION 9.4.2. Let $p \in M$. Denote by T_pM its tangent plane at p. The **Weingarten map** (also known as the shape operator)

$$W_p : T_p \to T_p$$

is the endomorphism of the tangent plane given by directional derivative of the extension N of n:

$$W(v) = \nabla_v N = \left.\frac{d}{dt}\right|_{t=0} n \circ \alpha(t), \tag{9.4.1}$$

where $\beta = x \circ \alpha$ is chosen so that $\beta(0) = p$ while $\beta'(0) = v$.

LEMMA 9.4.3. *The map W is well defined.*

PROOF. We have to show that the right hand side of formula (9.4.1), which is a priori a vector in $\mathbb{R}^3$, indeed produces a vector in the tangent plane. By Leibniz's rule,

$$\langle W(v), n\rangle = \langle \nabla_v N, n\rangle = \frac{1}{2}\frac{d}{dt}\langle n \circ \alpha(t), n \circ \alpha(t)\rangle = 0,$$

and therefore $W(v)$ indeed lies in T_p, proving the lemma. □

The connection with the Hessian is given in the following theorem.

THEOREM 9.4.4. *The operator W is a selfadjoint endomorphism of the tangent plane T_p, and satisfies*

$$\langle W(x_i), x_j\rangle = -\left\langle n, \frac{\partial^2 x}{\partial u^i \partial u^j}\right\rangle.$$

PROOF. As in the definition of the Weingarten map, $\nabla_v N$ is the derivative of N along a curve with initial vector v. We can choose the curve

$$\gamma(t) = x(t, a).$$

Then we have $\gamma'(t) = x_1$. Therefore

$$\nabla_{x_1} N = \frac{\partial n}{\partial u^1}$$

by definition of the directional derivative. Thus the coordinate lines $x(u^1, a)$ and $x(b, u^2)$ of the chart (u^1, u^2) allow us to write $\nabla_{x_j} N = \frac{\partial}{\partial u^j} n$. Therefore we have

$$\begin{aligned}\langle W(x_i), x_j\rangle &= \left\langle \frac{\partial n}{\partial u^i}, x_j\right\rangle \\ &= \frac{\partial}{\partial u^i}\langle n, x_j\rangle - \langle n, \frac{\partial}{\partial u^i} x_j\rangle \\ &= -\left\langle n, \frac{\partial^2 x}{\partial u^i \partial u^j}\right\rangle,\end{aligned}$$

which is an expression symmetric in i and j. This proves the selfadjointness of W by verifying it for a set of basis vectors. □

THEOREM 9.4.5. *The eigenvalues of the Weingarten map are real.*

This is immediate from the fact that the endomorphism W_p of T_pM is selfadjoint.

DEFINITION 9.4.6. The principal curvatures, denoted k_1 and k_2, are the eigenvalues of the Weingarten map W.

This will be discussed in more detail in the next chapter, around Definition 10.2.9.

EXAMPLE 9.4.7. Consider the plane $x(u^1, u^2) = (u^1, u^2, 0)$. We have $x_1 = e_1$, $x_2 = e_2$, while the normal vector field n

$$n = \frac{x_1 \times x_2}{(x_1 \times x_2)} = e_1 \times e_2 = e_3$$

is constant. It can therefore be extended to a constant vector field N defined in an open neighborhood in $\mathbb{R}^3$. Thus $\nabla_v N \equiv 0$ and $W(v) \equiv 0$, and the Weingarten map is identically zero.

In the next section we will present nonzero examples of the Weingarten map.

9.5. Weingarten map of sphere and cylinder

LEMMA 9.5.1. *The Weingarten map of the sphere of radius $r > 0$ at every point p of the sphere is the scalar map $T_p \to T_p$ given by*

$$\frac{1}{r} Id = \frac{1}{r} Id_{T_p},$$

where Id is the identity map of T_p.

PROOF. On the sphere, we have $n \circ \alpha(t) = \frac{1}{r}\beta(t)$. Hence

$$\begin{aligned} W(v) &= \nabla_v N(\beta(t)) \\ &= \left.\frac{d}{dt}\right|_{t=0} n \circ \alpha(t) \\ &= \frac{1}{r} \left.\frac{d}{dt}\right|_{t=0} \beta(t) \\ &= \frac{1}{r} v. \end{aligned}$$

Thus $W(v) = \frac{1}{r} v$ for all v. In other words, $W = \frac{1}{r}$ Id. □

Note that the Weingarten map has rank 2 in this case.

EXAMPLE 9.5.2. For the cylinder, we have the parametrisation

$$\underline{x}(u^1, u^2) = (\cos u^1, \sin u^1, u^2),$$

and $n = (\cos u^1, \sin u^1, 0)$. As before, this can be extended to a vector field N defined in an open neighborhood in $\mathbb{R}^3$, with the usual relation $\nabla_{x_1} N = \frac{\partial}{\partial u^1} n$. Hence we have

$$\nabla_{x_1} N = \frac{\partial}{\partial u^1} n = (-\sin u^1, \cos u^1, 0). \tag{9.5.1}$$

Similarly,

$$\nabla_{x_2} N = \frac{\partial}{\partial u^2} n = (0, 0, 0). \tag{9.5.2}$$

Now let $v = v^i x_i$ be an arbitrary tangent vector. Therefore

$$\nabla_v N = \nabla_{v^i x_i} N = v_1 \nabla_{x_1} N + v_2 \nabla_{x_2} N$$

by linearity. Hence (9.5.1) and (9.5.2) yield

$$\nabla_v N = v^1 \nabla_{x_1} N = v^1 \begin{pmatrix} -\sin u^1 \\ \cos u^1 \\ 0 \end{pmatrix},$$

and therefore,

$$W(v) = v^1 \begin{pmatrix} -\sin u^1 \\ \cos u^1 \\ 0 \end{pmatrix} = v^1 x_1,$$

where $v = v^i x_i$. Thus, $W(x_1) = x_1$, while $W(x_2) = 0$.

Note that the Weingarten map has rank 1 in this case.

9.6. Coefficients $L^i_{\ j}$ of Weingarten map

Since the two vectors (x_1, x_2) from a basis of the tangent plane T_p, we may introduce the following definition.

DEFINITION 9.6.1. The coefficients $L^i_{\ j}$ of the Weingarten map are defined by

$$W(x_j) = L^i_{\ j} x_i.$$

EXAMPLE 9.6.2. For the plane, we have $(L^i_{\ j}) \equiv 0$, for the sphere,

$$(L^i_{\ j}) = \begin{pmatrix} \frac{1}{r} & 0 \\ 0 & \frac{1}{r} \end{pmatrix} = \frac{1}{r} \delta^i_{\ j}.$$

For the cylinder, we have $L^1_{\ 1} = 1$. The remaining coefficients vanish, so that

$$(L^i_{\ j}) = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}.$$

9.7. Gaussian curvature

We continue with the terminology and notation of the previous section.

DEFINITION 9.7.1. The **Gaussian curvature** function $K = K(u^1, u^2)$ of the surface $\underline{x}$ is the determinant of the Weingarten map :

$$K = \det(L^i{}_j) = L^1{}_1 L^2{}_2 - L^1{}_2 L^2{}_1 = 2L^1{}_{[1} L^2{}_{2]}$$

(the skew-symmetrisation notation was defined in 1.7).

EXAMPLE 9.7.2. Cylinder, plane $K = 0$, *cf.* Figure 11.6.1. For a sphere of radius r, we have $K = \det \begin{pmatrix} \frac{1}{r} & 0 \\ 0 & \frac{1}{r} \end{pmatrix} = \frac{1}{r^2}$.

REMARK 9.7.3 (Sign of Gaussian curvature). Of particular geometric significance is the sign of the Gaussian curvature. The geometric meaning of negative Gaussian curvature is a saddle point. The geometric meaning of positive Gaussian curvature is a point of convexity, such as local minimum or local maximum of the graph of a function of two variables.

CHAPTER 10

Second fundamental form, theorema egregium

10.1. Second fundamental form

DEFINITION 10.1.1. The second fundamental form $\mathbf{II}_p$ is the bilinear form on the tangent plane T_p defined for $u, v \in T_p$ by

$$\mathbf{II}_p(u,v) = -\langle \nabla_u n, v\rangle.$$

It may be helpful to keep in mind the observation that the second fundamental form measures the curvature of geodesics on $\underline{x}$ viewed as curves of $\mathbb{R}^3$ (*cf.* 3.7.1) as illustrated by Theorem 10.2.1 below.

DEFINITION 10.1.2. The coefficients L_{ij} of the second fundamental form are defined to be

$$L_{ij} = \mathbf{II}(x_i, x_j) = -\left\langle \frac{\partial n}{\partial u^i}, x_j \right\rangle.$$

LEMMA 10.1.3. *The coefficients L_{ij} of the second fundamental form are symmetric in i and j, more precisely $L_{ij} = +\langle x_{ij}, n\rangle$.*

PROOF. We have $\langle n, x_i\rangle = 0$. Hence $\frac{\partial}{\partial u^j}\langle n, x_i\rangle = 0$, *i.e.*

$$\left\langle \frac{\partial}{\partial u^j} n, x_i \right\rangle + \langle n, x_{ij}\rangle = 0$$

or

$$\langle n, x_{ij}\rangle = -\langle \nabla_{x_j} n, x_i\rangle = +\mathbf{II}(x_j, x_i)$$

and the proof is concluded by the equality of mixed partials. □

LEMMA 10.1.4. *The second fundamental form allows us to identify the normal component of the second partials of the map x, namely: $x_{ij} = \Gamma_{ij}^k x_k + L_{ij} n$.*

PROOF. Let $x_{ij} = \Gamma_{ij}^k x_k + cn$ and form the inner product with n to obtain $L_{ij} = \langle x_{ij}, n\rangle = 0 + c\langle n, n\rangle = c$. □

10.2. Geodesics and second fundamental form

THEOREM 10.2.1. *Let $\beta(s)$ be a unit speed geodesic on the surface in $\mathbb{R}^3$, so that $\beta'(s) \in T_pM$ where $p = \beta(s)$. Then*

$$|\mathbf{II}(\beta', \beta')| = k_\beta(s),$$

where k_β is the curvature of β as a curve in $\mathbb{R}^3$.

PROOF. We apply formula (8.3.2). Since $|n| = 1$, we have

$$k_\beta \stackrel{\text{def}}{=} |\beta''| = |L_{ij}\alpha^{i'}\alpha^{j'}|.$$

Also

$$\mathbf{II}(\beta', \beta') = \mathbf{II}(x_i\alpha^{i'}, x_j\alpha^{j'}) = \alpha^{i'}\alpha^{j'}\mathbf{II}(x_i, x_j) = \alpha^{i'}\alpha^{j'}L_{ij},$$

completing the proof. □

PROPOSITION 10.2.2. *We have the following relation between the Weingarten map and the second fundamental form:*

$$L_{ij} = -L^k{}_j g_{ki}.$$

PROOF. By definition,

$$L_{ij} = \langle x_{ij}, n\rangle = -\left\langle \frac{\partial}{\partial u^j} n, x_i \right\rangle = -\langle L^k{}_j x_k, x_i\rangle = -L^k{}_j g_{ki}.$$

□

10.2.1. Normal and geodesic curvatures. The material in this subsection is optional.

Let $x(u^1, u^2)$ be a surface in $\mathbb{R}^3$. Let $\alpha(s) = (\alpha^1(s), \alpha^2(s))$ a curve in $\mathbb{R}^2$, and consider the curve $\beta = x \circ \alpha$ on the surface x. Consider also the normal vector to the surface, denoted n. The tangent unit vector $\beta' = \frac{d\beta}{ds}$ is perpendicular to n, so $\beta', n, \beta' \times n$ are three unit vector spanning $\mathbb{R}^3$. As β' is a unit vector, it is perpendicular to β'', and therefore β'' is a linear combination of n and $\beta' \times n$. Thus

$$\beta'' = k_n n + k_g(\beta' \times n),$$

where $k_n, k_g \in \mathbb{R}$ are called the *normal* and the *geodesic curvature* (resp.). Note that

$$k_n = \beta'' \cdot n, \;\; k_g = \beta'' \cdot (\beta' \times n).$$

If k is the curvature of β, then we have that $k^2 = ||\beta''||^2 = k_n^2 + k_g^2$.

EXERCISE 10.2.3. *Let $\beta(s)$ be a unit speed curve on a sphere of radius r. Then the normal curvature of β is $\pm 1/r$.*

REMARK 10.2.4. *Let π be a plane passing through the center of the sphere S in Exercise 10.2.3, and Let $C = \pi \cap S$. Then the curvature k of C is $1/r$, and thus $k_g = 0$. On the other hand, if π does not pass through the center of S (and $\pi \cap S \neq \emptyset$) then the geodesic curvature of the intersection $\neq 0$.*

PROPOSITION 10.2.5. *If a unit speed curve β on a surface is geodesic then its geodesic curvature is $k_g = 0$.*

PROOF. If β is a geodesic, then β'' is parallel to the normal vector n, so it is perpendicular to $n \times \beta''$, and therefore $k_g = \beta'' \cdot (n \times \beta'') = 0$. □

EXERCISE 10.2.6. *The inverse direction of Proposition 10.2.5 is also correct. Prove it.*

EXAMPLE 10.2.7. *Any line $\gamma(t) = at + b$ on a surface is a geodesic, as $\gamma'' = 0$ and therefore $k_g = 0$.*

EXAMPLE 10.2.8. *Take a cylinder x of radius 1 and intersect it with a plane π parallel to the xy plane. Let $C = x \cap \pi$ - it is a circle of radius 1. Thus $k = 1$. Show that $k_n = 1$ and thus C is a geodesic.*

DEFINITION 10.2.9. The principal curvatures, denoted k_1 and k_2, are the eigenvalues of the Weingarten map W.

Assume $k_1 > k_2$.

THEOREM 10.2.10. *The minimal and maximal values of the absolute value of the normal curvature $|k_n|$ at a point p of all curves on a surface passing through p are $|k_2|$ and $|k_1|$.*

PROOF. This is proven using the fact that $k_g = 0$ for geodesic curves and from Theorem 3.9.4?[1] □

10.3. Calculus of variations and the geodesic equation

Calculus of variations is known as tachsiv variatsiot.

Let $\alpha(s) = (\alpha^1(s), \alpha^2(s))$, and consider the curve $\beta = x \circ \alpha$ on the surface :

$$\underset{s}{[a, b]} \xrightarrow{\alpha} \mathbb{R}^2 \xrightarrow{x} \mathbb{R}^3.$$

Consider the energy functional

$$\mathcal{E}(\beta) = \int_a^b \|\beta'\|^2 ds.$$

[1]See which theorem this is.

Here $' = \dfrac{d}{ds}$. We have $\dfrac{d\beta}{ds} = x_i(\alpha^i)'$ by chain rule. Thus

$$\mathcal{E}(\beta) = \int_a^b \langle \beta', \beta' \rangle ds = \int_a^b \left\langle x_i \alpha^{i\prime}, x_j \alpha^{j\prime} \right\rangle ds = \int_a^b g_{ij} \alpha^{i\prime} \alpha^{j\prime} ds. \tag{10.3.1}$$

Consider a variation $\alpha(s) \to \alpha(s) + t\delta(s)$, t small, such that $\delta(a) = \delta(b) = 0$. If $\beta = x \circ \alpha$ is a critical point of $\mathcal{E}$, then for any perturbation δ vanishing at the endpoints, the following derivative vanishes:

$$\begin{aligned} 0 &= \frac{d}{dt}\bigg|_{t=0} \mathcal{E}(x \circ (\alpha + t\delta)) \\ &= \frac{d}{dt}\bigg|_{t=0} \left\{ \int_a^b g_{ij} (\alpha^i + t\delta^i)' (\alpha^j + t\delta^j)' ds \right\} \text{(from equation (10.3.1))} \\ &= \underbrace{\int_a^b \frac{\partial\,(g_{ij} \circ (\alpha + t\delta))}{\partial t} \alpha^{i'} \alpha^{j'} ds}_{A} + \underbrace{\int_a^b g_{ij} \big(\alpha^{i'} \delta^{j'} + \alpha^{j'} \delta^{i'} \big) ds}_{B}, \end{aligned}$$

so that we have

$$A + B = 0. \tag{10.3.2}$$

We will need to compute both the t-derivative and the s-derivative of the first fundamental form. The formula is given in the lemma below.

LEMMA 10.3.1. *The partial derivatives of $g_{ij} = g_{ij} \circ (\alpha(s) + t\delta(s))$ along $\beta = x \circ \alpha$ are given by the following formulas:*

$$\frac{\partial}{\partial t}(g_{ij} \circ (\alpha + t\delta)) = (\langle x_{ik}, x_j \rangle + \langle x_i, x_{jk} \rangle)\delta^k,$$

and

$$\frac{\partial g_{ik}}{\partial s} = (\langle x_{im}, x_k \rangle + \langle x_i, x_{km} \rangle)(\alpha^m)'.$$

PROOF. We have

$$\begin{aligned} \frac{\partial}{\partial t}(g_{ij} \circ (\alpha + t\delta)) &= \frac{\partial}{\partial t} \langle x_i \circ (\alpha + t\delta)), x_j \circ (\alpha + t\delta) \rangle \\ &= \left\langle \frac{\partial}{\partial t}(x_i \circ (\alpha + t\delta)), x_j \right\rangle + \left\langle x_i, \frac{\partial}{\partial t}(x_j \circ (\alpha + t\delta)) \right\rangle \\ &= \langle x_{ik}\delta^k, x_j \rangle + \langle x_i, x_{jk}\delta^k \rangle = (\langle x_{ik}, x_j \rangle + \langle x_i, x_{jk} \rangle)\delta^k. \end{aligned}$$

Furthermore, $g_{ik}' = \langle x_i, x_k \rangle' = \langle x_{im}(\alpha^m)', x_k \rangle + \langle x_i, x_{km}(\alpha^m)' \rangle$. □

LEMMA 10.3.2. *Let $f \in C^0[a,b]$. Suppose that for all $g \in C^0[a,b]$ we have $\int_a^b f(x)g(x)dx = 0$. Then $f(x) \equiv 0$. This conclusion remains true if we use only test functions $g(x)$ such that $g(a) = g(b) = 0$.*

PROOF. We try the test function $g(x) = f(x)$. Then $\int_a^b (f(x))^2 ds = 0$. Since $(f(x))^2 \geq 0$ and f is continuous, it follows that $f(x) \equiv 0$. If we want $g(x)$ to be 0 at the endpoints, it suffices to choose $g(x) = (x-a)(b-x)f(x)$. □

THEOREM 10.3.3. *Suppose $\beta = x \circ \alpha$ is a critical point of the energy functional (endpoints fixed). Then β satisfies the differential equation*

$$(\forall k) \quad (\alpha^k)'' + \Gamma_{ij}^k (\alpha^i)'(\alpha^j)' = 0.$$

PROOF. We use Lemma 10.3.1 to evaluate the term A from equation (10.3.2) as follows:

$$\begin{aligned} A &= \int_a^b (\langle x_{ik}, x_j\rangle + \langle x_i, x_{jk}\rangle)(\alpha^i)'(\alpha^j)'\delta^k ds \\ &= 2\int_a^b \langle x_{ik}, x_j\rangle \alpha^{i'} \alpha^{j'} \delta^k ds, \end{aligned}$$

since summation is over both i and j. Similarly,

$$\begin{aligned} B &= 2\int_a^b g_{ij} \alpha^{i'} \delta^{j'} ds \\ &= -2\int_a^b \left(g_{ij}\alpha^{i'}\right)' \delta^j ds \end{aligned}$$

by integration by parts, where the boundary term vanishes since $\delta(a) = \delta(b) = 0$. Hence

$$B = -2\int_a^b \left(g_{ik}\alpha^{i'}\right)' \delta^k ds$$

by changing an index of summation. Thus

$$\frac{1}{2}\frac{d}{dt}\bigg|_{t=0} (\mathcal{E}) = \int_a^b \left\{ \langle x_{ik}, x_j\rangle \alpha^{i'}\alpha^{j'} - \left(g_{ik}\alpha^{i'}\right)' \right\} \delta^k ds.$$

Since this is true for any variation (δ^k), by Lemma 10.3.2 we obtain the Euler-Lagrange equation

$$(\forall k) \quad \langle x_{ik}, x_j\rangle \alpha^{i'}\alpha^{j'} - \left(g_{ik}\alpha^{i'}\right)' \equiv 0,$$

or

$$\langle x_{ik}, x_j\rangle \alpha^{i'}\alpha^{j'} - g_{ik}{}'\alpha^{i'} - g_{ik}\alpha^{i''} = 0. \tag{10.3.3}$$

Using the formula from Lemma 10.3.1 for the s-derivative of the first fundamental form, we can rewrite the formula (10.3.3) as follows:

$$\begin{aligned} 0 &= \langle x_{ik}, x_j\rangle \alpha^{i'}\alpha^{j'} - \langle x_{im}, x_k\rangle \alpha^{m'}\alpha^{i'} - \langle x_i, x_{km}\rangle \alpha^{i'}\alpha^{m'} - g_{ik}\alpha^{i''} \\ &= -\langle \Gamma^n_{im} x_n, x_k\rangle \alpha^{m'}\alpha^{i'} - g_{ik}\alpha^{i''} \\ &= -\Gamma^n_{im} g_{nk}\alpha^{m'}\alpha^{i'} - g_{ik}\alpha^{i''}, \end{aligned}$$

where the cancellation of the first and the third term in the first line results from replacing index i by j and m by i in the third term. This is true $\forall k$. Now multiply by g^{jk} :

$$\begin{aligned} g^{jk}\Gamma^n_{im} g_{nk}\alpha^{m'}\alpha^{i'} + g^{jk} g_{ik}\alpha^{i'} &= \delta^j_n \Gamma^n_{im}\alpha^{m'}\alpha^{i'} + \delta^j_i \alpha^{i''} \\ &= \Gamma^j_{im}\alpha^{m'}\alpha^{i'} + \alpha^{j''} \\ &= 0, \end{aligned}$$

which is the desired geodesic equation. □

10.4. Three formulas for Gaussian curvature

Remark 10.4.1. Throughout this Chapter, we will make use of the Einstein summation convention, *i.e.* suppress the summation symbol $\sum$ when the summation index occurs simultaneously as a subscript and a superscript in a formula, *cf.* (11.4.1).

Theorem 10.4.2. *We have the following three equivalent formulas for the Gaussian curvature:*

(a) $K = \det(L^i{}_j) = 2L^1{}_{[1}L^2{}_{2]}$;

(b) $K = \frac{\det(L_{ij})}{\det(g_{ij})}$;

(c) $K = -\frac{2}{g_{11}} L_{1[1}L^2{}_{2]}$.

Proof. The first formula is our definition of K. To prove the second formula, we use the formula $L_{ij} = -L^k{}_j g_{ki}$ of Proposition 10.2.2. By the multiplicativity of determinant with respect to matrix multiplication,

$$\det(L^i{}_j) = (-1)^2 \frac{\det(L_{ij})}{\det(g_{ij})}.$$

Let us now prove formula (c). The proof is a calculation. Note that by definition, $2L_{1[1}L^2{}_{2]} = L_{11}L^2{}_2 - L_{12}L^2{}_1$. Hence

$$\begin{aligned} -\frac{2}{g_{11}}\left(L_{1[1}L^2{}_{2]}\right) &= \frac{1}{g_{11}}\left(\left(L^1{}_1g_{11} - L^2{}_1g_{21}\right)L^2{}_2 - \left(L^1{}_2g_{11} - L^2{}_2g_{21}\right)L^2{}_1\right) \\ &= \frac{1}{g_{11}}\left(g_{11}L^1{}_1L^2{}_2 - g_{21}L^2{}_1L^2{}_2 - g_{11}L^1{}_2L^2{}_1 + g_{21}L^2{}_2L^2{}_1\right) \\ &= \det(L^i{}_j) = K. \end{aligned}$$

Note that we can either calculate using the formula $L_{ij} = -L^k{}_jg_{ki}$, or the formula $L_{ij} = -L^k{}_ig_{kj}$. Only the former one leads to the appropriate cancellations as above. □

10.5. Principal curvatures

Definition 10.5.1. The principal curvatures, denoted k_1 and k_2, are the eigenvalues of the Weingarten map W.

Remark 10.5.2. The curvatures k_1 and k_2 are real. This follows from the selfadjointness of W (Theorem 9.4.4) and Corollary 2.2.2.

Theorem 10.5.3. *The Gaussian curvature equals the product of the principal curvatures.*

Proof. The determinant of a 2 by 2 matrix equals the product of its eigenvalues: $K = \det(L^i{}_j) = k_1k_2$. □

Theorem 10.5.4. *Let v be a unit eigenvector belonging to a principal curvature k. Let $\beta(s)$ be a geodesic satisfying $\beta'(0) = v$. Then the curvature of β as a space curve is the absolute value of k :*

$$k_\beta(0) = |k|.$$

Proof. Let $v = v^ix_i$ be a unit eigenvector of W represented by a curve $\beta(t) = x \circ \alpha(t)$, so that $v^i = \alpha^{i\prime}$. The eigenvector property $W(v) = kv$ translates into

$$L^i{}_jv^j = kv^i. \tag{10.5.1}$$

As before we may write $\beta = x \circ \alpha$. Meanwhile, by Theorem 10.2.1, we have

$$\begin{aligned} \pm k_\beta &= \mathbf{II}(\beta', \beta') \\ &= L_{ij}\alpha^{i'}\alpha^{j'} \\ &= L_{ij}v^iv^j \\ &= -L^m{}_jg_{mi}v^iv^j \\ &= \left(-L^m{}_jv^j\right)g_{mi}v^i. \end{aligned}$$

Therefore from equation (10.5.1) we obtain

$$\begin{aligned} \pm k_\beta &= -k\ v^m g_{mi} v^i \\ &= -k\ \|v\|^2 \\ &= -k, \end{aligned}$$

proving the theorem. □

COROLLARY 10.5.5. *Gaussian curvature at a point p of a regular surface in $\mathbb{R}^3$ is the product of curvatures of two perpendicular geodesics passing through p, whose tangent vectors are eigenvectors of the Weingarten map at the point.*

CHAPTER 11

Minimal surfaces, Theorema egregium

11.1. Minimal surfaces

DEFINITION 11.1.1. The Mean curvature H is half the trace of the Weingarten map: $H = \frac{1}{2}(k_1 + k_2) = \frac{1}{2}L^i{}_i$, *cf.* Table 11.6.1.

A surface $x(u^1, u^2)$ is called minimal if $H = 0$ at every point, *i.e.* $k_1 + k_2 = 0$. Geometrically, such a surface is represented by a soap film[1] see Figure 11.1.1.

DEFINITION 11.1.2. A parametrisation $\underline{x}(u^1, u^2)$ is called *isothermal* if there is a function $f = f(u^1, u^2)$ such that $g_{ij} = f^2\delta_{ij}$, *i.e.* $\langle x_1, x_1\rangle = \langle x_2, x_2\rangle$ and $\langle x_1, x_2\rangle = 0$.

[1]krum sabon, as opposed to bu'at sabon. Dip a wire (tayil) into soapy water.

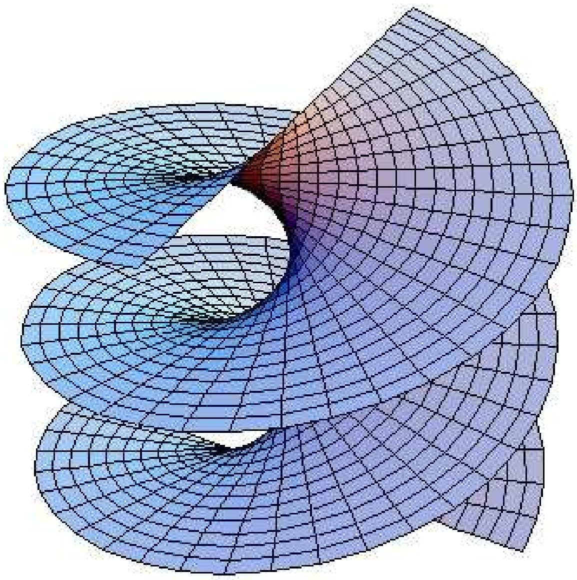

FIGURE 11.1.1. Helicoid: a minimal surface

PROPOSITION 11.1.3. *Assume $\underline{x}$ is isothermal. Then it satisfies the partial differential equation*

$$\underline{x}_{11} + \underline{x}_{22} = -2f^2 H \underline{n}.$$

PROOF. Note

$$L_{ij} = -L^m{}_j g_{mi} = -L^m{}_j f^2 \delta_{mi} = -f^2 L^i{}_j.$$

Thus

$$L^i{}_j = -\frac{L_{ij}}{f^2}$$

so that the mean curvature H satisfies

$$H = \frac{1}{2} L^i{}_i = -\frac{L_{11} + L_{22}}{2f^2}. \tag{11.1.1}$$

Since $\langle x_1, x_2 \rangle = 0$, we have $\dfrac{\partial}{\partial u^2} \langle x_1, x_2 \rangle = 0$. Therefore

$$\langle x_{12}, x_2 \rangle + \langle x_1, x_{22} \rangle = 0. \tag{11.1.2}$$

So $-\langle x_{12}, x_2 \rangle = \langle x_1, x_{22} \rangle$. By Definition 11.1.2, we have

$$\langle x_1, x_1 \rangle - \langle x_2, x_2 \rangle = 0.$$

From formula (11.1.2) we obtain

$$\begin{aligned} 0 &= \frac{\partial}{\partial u^1} \langle x_1, x_1 \rangle - \frac{\partial}{\partial u^1} \langle x_2, x_2 \rangle \\ &= 2\langle x_{11}, x_1 \rangle - 2\langle x_{21}, x_2 \rangle \\ &= 2\langle x_{11}, x_1 \rangle + 2\langle x_{22}, x_1 \rangle \\ &= 2\langle x_{11} + x_{22}, x_1 \rangle. \end{aligned}$$

Inspecting the u^2-derivatives, we similarly obtain $\langle x_{11} + x_{22}, x_2 \rangle = 0$. Since x_1, x_2 and n form an orthogonal basis, $x_{11} + x_{22}$ is proportional to n. Write $x_{11} + x_{22} = cn$. Applying (11.1.1), we obtain

$$\begin{aligned} c &= \langle x_{11} + x_{22}, n \rangle \\ &= \langle x_{11}, n \rangle + \langle x_{22}, n \rangle \\ &= L_{11} + L_{22} \\ &= -2f^2 H, \end{aligned}$$

as required. □

DEFINITION 11.1.4. Let $F : \mathbb{R}^2 \to \mathbb{R}$, $(u^1, u^2) \mapsto F(u^1, u^2)$. The *Laplacian*, denoted $\Delta(F)$, is the second-order differential operator on functions, defined by

$$\Delta F = \frac{\partial^2 F}{\partial (u^1)^2} + \frac{\partial^2 F}{\partial (u^2)^2}.$$

Functions in the kernel of the Laplacian are called *harmonic*.

DEFINITION 11.1.5. We say that F is *harmonic* if $\Delta F = 0$.

Harmonic functions are important in the study of heat flow, or heat transfer (ma'avar chom).

COROLLARY 11.1.6. *Let* $\underline{x}(u^1, u^2) = (x(u^1, u^2), y(u^1, u^2), z(u^1, u^2))$ *in* $\mathbb{R}^3$ *be a parametrized surface and assume that* $\underline{x}$ *is isothermal. Then* $\underline{x}$ *is minimal if and only if the coordinate functions* x, y, z *are harmonic.*

PROOF. From Proposition 11.1.3 we have

$$\|\Delta(\underline{x})\| = \sqrt{(\Delta x)^2 + (\Delta y)^2 + (\Delta z)^2} = 2f^2|H|,$$

proving the corollary. □

EXAMPLE 11.1.7. The catenoid is parametrized as follows:

$$x(\theta, \phi) = (a\cosh\phi\cos\theta, a\cosh\phi\sin\theta, a\phi).$$

Here $f(\phi) = a\cosh\phi$, while $g(\phi) = a\phi$. Then $g_{11} = a^2\cosh^2\phi = f^2, g_{22} = (a\sinh\phi)^2 + a^2 = a^2\cosh^2\phi = f^2, g_{12} = 0$. Calculate:

$$\begin{aligned} x_{11} + x_{22} &= (-a\cosh\phi\cos\theta, -a\cosh\phi\sin\theta, 0) \\ &\quad + (a\cosh\phi\cos\theta, a\cosh\phi\sin\theta, 0) \\ &= (0, 0, 0). \end{aligned}$$

Thus the catenoid is a minimal surface. In fact, it is the only surface of revolution which is minimal [**We55**, p. 179].

11.2. Introduction to theorema egregium; intrinsic *vs* extrinsic

Some of the material in this chapter has already been covered in earlier chapters. We include such material here for the benefit of the reader primarily interested in understanding Gaussian curvature.

The present Chapter is an introduction to Gaussian curvature. Understanding the intrinsic nature of Gaussian curvature, *i.e.* the *theorema egregium* of Gauss, clarifies the geometric classification of surfaces. A beautiful historical account and an analysis of Gauss's proof of the *theorema egregium* may be found in [**Do79**].[2]

We present a compact formula for Gaussian curvature and its proof, along the lines of the argument in M. do Carmo's book [**Ca76**]. The

[2]Our formula (11.6.2) for Gaussian curvature is similar to the traditional formula for the Riemann curvature tensor in terms of the Levi-Civita connection (note the antisymmetrisation in both formulas, and the corresponding two summands), without the burden of the connection formalism.

appeal of an old-fashioned, computational, coordinate notation proof is that it obviates the need for higher order objects such as connections, tensors, exponental map, *etc*, and is, hopefully, directly accessible to a student not yet familiar with the subject, *cf.* Remark 11.7.4.

We would like to distinguish two types of properties of a surface in Euclidean space: intrinsic and extrinsic. Understanding the extrinsic/intrinsic dichotomy is equivalent to understanding the *theorema egregium* of Gauss. The *theorema egregium* is the key insight lying at the foundation of differential geometry as conceived by B. Riemann in his essay [**Ri1854**] presented before the Royal Scientific Society of Göttingen a century and a half ago.

The *theorema egregium* asserts that an infinitesimal invariant of a surface in Euclidean space, called Gaussian curvature, is an "intrinsic" invariant of the surface Σ. In other words, Gaussian curvature of Σ is independent of its isometric imbedding in Euclidean space. This theorem paves the way for an intrinsic definition of curvature in modern Riemannian geometry.

We will prove that Gaussian curvature K is an *intrinsic* invariant of a surface in Euclidean space, in the following precise sense: K can be expressed in terms of the coefficients of the *first* fundamental form (see Section) and their derivatives alone. *A priori* the possibility of thus expressing K is not obvious, as the naive definition of K involves the coefficients of the *second* fundamental form (alternatively, of the Weingarten map).

11.3. Riemann's formula

The intrinsic nature of Gaussian curvature paves the way for a transition from classical differential geometry, to a more abstract approach of modern differential geometry. The distinction can be described roughly as follows. Classically, one studies surfaces in Euclidean space. Here the first fundamental form (g_{ij}) of the surface is the restriction of the Euclidean inner product. Meanwhile, abstractly, a surface comes equipped with a set of coefficients, which we deliberately denote by the same letters, (g_{ij}) in each coordinate patch, or equivalently, its element of length. One then proceeds to study its geometry without any reference to a Euclidean imbedding, *cf.* (16.4.2).

Such an approach was pioneered in higher dimensions in Riemann's essay. The essay contains a single formula [**Ri1854**, p. 292] (*cf.* [**Sp79**, p. 147]), namely the formula for the element of length of a surface of constant (Gaussian) curvature $K \equiv \alpha$:

$$\frac{1}{1+\frac{\alpha}{4}\sum x^2}\sqrt{\sum dx^2} \tag{11.3.1}$$

(today, of course, we would incorporate a summation index as part of the notation), *cf.* formulas (16.4.3), (16.17.1), and (17.3.1).

11.4. Preliminaries to the *theorema egregium*

The material in this section has mostly been covered already in earlier chapters.

DEFINITION 11.4.1. Given a surface Σ in 3-space which is the graph of a function of two variables, consider a *critical point* $p \in \Sigma$. The *Gaussian curvature* of the surface at the critical point p is the determinant of the Hessian of the function, *i.e.* the determinant of the two-by-two matrix of its second derivatives.

The implicit function theorem allows us to view any point of a regular surface, as such a critical point, after a suitable rotation. We have thus given the simplest possible definition of Gaussian curvature at any point of a regular surface.

REMARK 11.4.2. The appeal if this definition is that it allows one immediately to grasp the basic distinction between negative *versus* positive curvature, in terms of the dichotomy "saddle point *versus* cup/cap".

It will be more convenient to use an alternative definition in terms of the Weingarten map, which is readily shown to be equivalent to Definition 11.4.1.

We denote the coordinates in $\mathbb{R}^2$ by (u^1, u^2). In $\mathbb{R}^3$, let $\langle\, , \rangle$ be the Euclidean inner product. Let $x = x(u^1, u^2) : \mathbb{R}^2 \to \mathbb{R}^3$ be a regular parametrized surface. Here "regular" means that the vector valued function x has differential dx of rank 2 at every point. Consider partial derivatives $x_i = \frac{\partial}{\partial u^i}(x)$, where $i = 1, 2$. Thus, vectors x_1 and x_2 form a basis for the tangent plane at every point. Similarly, let

$$x_{ij} = \frac{\partial^2 x}{\partial u^i \partial u^j} \in \mathbb{R}^3.$$

Let $n = n(u^1, u^2)$ be a unit normal to the surface at the point $x(u^1, u^2)$, so that $\langle n, x_i\rangle = 0$.

DEFINITION 11.4.3 (symbols g_{ij}, L_{ij}, Γ^k_{ij}). The first fundamental form (g_{ij}) is given in coordinates by $g_{ij} = \langle x_i, x_j\rangle$. The second fundamental form (L_{ij}) is given in coordinates by $L_{ij} = \langle n, x_{ij}\rangle$. The symbols Γ_{ij}^k are uniquely defined by the decomposition

$$\begin{aligned} x_{ij} &= \Gamma_{ij}^k x_k + L_{ij} n \\ &= \Gamma_{ij}^1 x_1 + \Gamma_{ij}^2 x_2 + L_{ij} n, \end{aligned} \tag{11.4.1}$$

where the repeated (upper and lower) index k implies summation, in accordance with the Einstein summation convention, *cf.* Remark 10.4.1.

DEFINITION 11.4.4 (symbols ${L^i}_j$). The Weingarten map $({L^i}_j)$ is an endomorphism of the tangent plane, namely $\mathbb{R}x_1 \oplus \mathbb{R}x_2$. It is uniquely defined by the decomposition

$$\begin{aligned} n_j &= {L^i}_j x_i \\ &= {L^1}_j x_1 + {L^2}_j x_2, \end{aligned}$$

where $n_j = \frac{\partial}{\partial u^j}(n)$.

DEFINITION 11.4.5. We will denote by square brackets [ ] the anti-symmetrisation over the pair of indices found in between the brackets, *e.g.*

$$a_{[ij]} = \tfrac{1}{2}(a_{ij} - a_{ji}).$$

Note that $g_{[ij]} = 0$, $L_{[ij]} = 0$, and $\Gamma_{[ij]}^k = 0$. We will use the notation $\Gamma_{ij;\ell}^k$ for the ℓ-th partial derivative of the symbol Γ_{ij}^k.

11.5. An identity involving the Γ_{ij}^k and the L_{ij}

The following technical result will imply the *theorema egregium* as an easy consequence.

PROPOSITION 11.5.1. *We have the following relation*

$$\Gamma_{i[j;k]}^q + \Gamma_{i[j}^m \, \Gamma_{k]m}^q = -L_{i[j} {L^q}_{k]}$$

for each set of indices i, j, k, q (with, as usual, an implied summation over the index m).

PROOF. Let us calculate the tangential component, with respect to the basis $\{x_1, x_2, n\}$, of the third partial derivative

$$x_{ijk} = \frac{\partial^3 x}{\partial u^i \partial u^j \partial u^k}.$$

Recall that $n_k = {L^p}_k x_p$ and $x_{jk} = \Gamma_{jk}^\ell x_\ell + L_{ik} n$. Thus, we have

$$\begin{aligned} (x_{ij})_k &= \left(\Gamma_{ij}^m x_m + L_{ij} n\right)_k \\ &= \Gamma_{ij;k}^m x_m + \Gamma_{ij}^m x_{mk} + L_{ij} n_k + L_{ij;k} n \\ &= \Gamma_{ij;k}^m x_m + \Gamma_{ij}^m \left(\Gamma_{mk}^p x_p + L_{mk} n\right) + L_{ij} \left({L^p}_k x_p\right) + L_{ij;k} n. \end{aligned}$$

Grouping together the tangential terms, we obtain

$$\begin{aligned}(x_{ij})_k &= \Gamma^m_{ij;k} x_m + \Gamma^m_{ij}\left(\Gamma^p_{mk} x_p\right) + L_{ij}\left(L^p{}_k x_p\right) + (\ldots)n \\ &= \left(\Gamma^q_{ij;k} + \Gamma^m_{ij}\Gamma^q_{mk} + L_{ij}L^q{}_k\right) x_q + (\ldots)n \\ &= \left(\Gamma^q_{ij;k} + \Gamma^m_{ij}\Gamma^q_{km} + L_{ij}L^q{}_k\right) x_q + (\ldots)n,\end{aligned}$$

since the symbols Γ^q_{km} are symmetric in the two subscripts. Now the symmetry in j, k (equality of mixed partials) implies the following identity: $x_{i[jk]} = 0$. Therefore

$$\begin{aligned}0 &= x_{i[jk]} \\ &= (x_{i[j})_{k]} \\ &= \left(\Gamma^q_{i[j;k]} + \Gamma^m_{i[j}\Gamma^q_{k]m} + L_{i[j}L^q{}_{k]}\right) x_q + (\ldots)n,\end{aligned}$$

and therefore $\Gamma^q_{i[j;k]} + \Gamma^m_{i[j}\Gamma^q_{k]m} + L_{i[j}L^q{}_{k]} = 0$ for each $q = 1, 2$. □

11.6. The *theorema egregium* of Gauss

Recall from Section 10.4 that we have $K = -\frac{2}{g_{11}} L_{1[1}L^2{}_{2]}$, where by definition

$$K = \det(L^i{}_j) = 2L^1{}_{[1}L^2{}_{2]}. \tag{11.6.1}$$

THEOREM 11.6.1 (Theorema egregium). *The Gaussian curvature function $K = K(u^1, u^2)$ can be expressed in terms of the coefficients of the first fundamental form alone (and their first and second derivatives) as follows:*

$$K = \frac{2}{g_{11}}\left(\Gamma^2_{1[1;2]} + \Gamma^j_{1[1}\Gamma^2_{2]j}\right), \tag{11.6.2}$$

where the symbols Γ^k_{ij} can be expressed in terms of the derivatives of g_{ij} be the formula $\Gamma^k_{ij} = \frac{1}{2}(g_{i\ell;j} - g_{ij;\ell} + g_{j\ell;i})g^{\ell k}$, where (g^{ij}) is the inverse matrix of (g_{ij}).

PROOF OF *theorema egregium*. We present a streamlined version of do Carmo's proof [**Ca76**, p. 233]. The proof is in 3 steps.

(1) We express the third partial derivative x_{ijk} in terms of the Γ's (intrinsic information) and the L's (extrinsic information).
(2) The equality of mixed partials yields an identification of a suitable expression in terms of the Γ's, with a certain combination of the L's.
(3) The combination of the L's is expressed in terms of Gaussian curvature.

	Weingarten map $(L^i{}_j)$	Gaussian curvature K	mean curvature H
plane	$\begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$	0	0
cylinder	$\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$	0	$\frac{1}{2}$
invariance of curvature		yes	no

TABLE 11.6.1. Plane and cylinder have the same intrinsic geometry (K), but different extrinsic geometries (H)

The first two steps were carried out in Proposition 11.5.1. We choose the valus $i = j = 1$ and $k = q = 2$ for the indices. Applying Theorem 10.4.2(c), we obtain

$$\begin{aligned} \Gamma^2_{1[1;2]} + \Gamma^m_{1[1}\Gamma^2_{2]m} &= -L_{1[1}L^2{}_{2]} \\ &= g_{1i}L^i{}_{[1}L^2{}_{2]} \\ &= g_{11}L^1{}_{[1}L^2{}_{2]} \end{aligned}$$

since the term $L^2{}_{[1}L^2{}_{2]} = 0$ vanishes. This yields the desired formula for K and complete the proof of the *theorema egregium*. □

REMARK 11.6.2. Unlike Gaussian curvature K, the mean curvature $H = \frac{1}{2}L^i{}_i$ cannot be expressed in terms of the g_{ij} and their derivatives. Indeed, the plane and the cylinder have parametrisations with identical g_{ij}, but with different mean curvature, *cf.* Table 11.6.1. To summarize, Gaussian curvature is an intrinsic invariant, while mean curvature an extrinsic invariant, of the surface.

11.7. The Laplacian formula for Gaussian curvature

A (u^1, u^2) chart in which the metric becomes conformal (see Definition 16.7.1) to the standard flat metric, is referred to as *isothermal coordinates*. The existence of the latter is proved in [**Bes87**].

DEFINITION 11.7.1. The Laplace-Beltrami operator for a metric $\lambda\delta_{ij}$ in isothermal coordinates is

$$\Delta_{LB} = \frac{1}{\lambda}\left(\frac{\partial^2}{\partial(u^1)^2} + \frac{\partial^2}{\partial(u^2)^2}\right).$$

Γ^1_{ij}	$j=1$	$j=2$
$i=1$	μ_1	μ_2
$i=2$	μ_2	$-\mu_1$

Γ^2_{ij}	$j=1$	$j=2$
$i=1$	$-\mu_2$	μ_1
$i=2$	μ_1	μ_2

TABLE 11.7.1. Symbols Γ^k_{ij} of a metric $e^{2\mu(u^1,u^2)}\delta_{ij}$

The notation *means* that when we apply the operator Δ_{LB} to a function $f = f(u^1, u^2)$, we obtain

$$\Delta_{LB}(f) = \frac{1}{\lambda}\left(\frac{\partial^2 f}{\partial(u^1)^2} + \frac{\partial^2 f}{\partial(u^2)^2}\right).$$

In more readable form for $f = f(x, y)$, we have

$$\Delta_{LB}(f) = \frac{1}{\lambda}\left(\frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2}\right).$$

THEOREM 11.7.2. *Given a metric in isothermal coordinates with metric coefficients $g_{ij} = \lambda(u^1, u^2)\delta_{ij}$, its Gaussian curvature is minus half the Laplace-Beltrami operator applied to the* log *of the conformal factor λ:*

$$K = -\frac{1}{2}\Delta_{LB}\log\lambda. \tag{11.7.1}$$

PROOF. Let $\lambda = e^{2\mu}$. We have from Table 11.7.1:

$$2\Gamma^2_{1[1;2]} = \Gamma^2_{11;2} - \Gamma^2_{12;1} = -\mu_{22} - \mu_{11}.$$

It remains to verify that the $\Gamma\Gamma$ term in the expression (11.6.2) for the Gaussian curvature vanishes:

$$\begin{aligned} 2\Gamma^j_{1[1}\Gamma^2_{2]j} &= 2\Gamma^1_{1[1}\Gamma^2_{2]1} + 2\Gamma^2_{1[1}\Gamma^2_{2]2} \\ &= \Gamma^1_{11}\Gamma^2_{21} - \Gamma^1_{12}\Gamma^2_{11} + \Gamma^2_{11}\Gamma^2_{22} - \Gamma^2_{12}\Gamma^2_{12} \\ &= \mu_1\mu_1 - \mu_2(-\mu_2) + (-\mu_2)\mu_2 - \mu_1\mu_1 \\ &= 0. \end{aligned}$$

Then from formula (11.6.2) we have

$$K = \frac{2}{\lambda}\Gamma^2_{1[1;2]} = -\frac{1}{\lambda}(\mu_{11} + \mu_{22}) = -\Delta_{LB}\mu.$$

Meanwhile, $\Delta_{LB}\log\lambda = \Delta_{LB}(2\mu) = 2\Delta_{LB}(\mu)$, proving the result. □

Setting $\lambda = f^2$, we restate the theorem as follows.

THEOREM 11.7.3. *Given a metric in isothermal coordinates with metric coefficients* $g_{ij} = f^2(u^1, u^2)\delta_{ij}$*, its Gaussian curvature is minus the Laplace-Beltrami operator of the* log *of the conformal factor* f*:*

$$K = -\Delta_{LB} \log f. \tag{11.7.2}$$

Either one of the formulas (11.6.2), (11.7.1), or (11.7.2) can serve as the intrinsic definition of Gaussian curvature, replacing the extrinsic definition (11.6.1), *cf.* Remark 11.3.

REMARK 11.7.4. For a reader familiar with elements of Riemannian geometry, it is worth mentioning that the Jacobi equation

$$y'' + Ky = 0$$

of a Jacobi field y on M (expressing an infinitesimal variation by geodesics) sheds light on the nature of curvature in a way that no mere formula for K could.

Thus, in positive curvature, geodesics converge, while in negative curvature, they diverge.

However, to prove the Jacobi equation, one would need to have already an intrinsically well-defined quantity on the left hand side, $y'' + Ky$, of the Jacobi equation. In particular, one would need an already intrinsic notion of curvature K. Thus, a proof of the *theorema egregium* necessarily precedes the deeper insights provided by the Jacobi equation.

Similarly, the Gaussian curvature at $p \in M$ is the first significant term in the asymptotic expansion of the length of a "small" circle of center p. This fact, too, sheds much light on the nature of Gaussian curvature. However, to define what one means by a "small" circle, requires introducing higher order notions such as the exponential map, which are usually understood at a later stage than the notion of Gaussian curvature, *cf.* [**Ca76, Car92, Ch93, GaHL04**].

CHAPTER 12

Gauss–Bonnet theorem

12.1. Binet–Cauchy identity

THEOREM 12.1.1 (Binet–Cauchy identity). *The 3-dimensional case of the Binet–Cauchy identity is the identity*

$$(a \cdot c)(b \cdot d) = (a \cdot d)(b \cdot c) + (a \times b) \cdot (c \times d),$$

where a, b, c, and d are vectors in $\mathbb{R}^3$.

The formula can also be written as a formula giving the dot product of two wedge products, namely

$$(a \times b) \cdot (c \times d) = (a \cdot c)(b \cdot d) - (a \cdot d)(b \cdot c),$$

THEOREM 12.1.2 (Special case of Binet–Cauchy). *In the special case of unit vectors $a = c$ and $b = d$, we obtain*

$$|a \times b|^2 = |a|^2|b|^2 - |a \cdot b|^2.$$

COROLLARY 12.1.3. *Let a, b be the two tangent vectors x_1, x_2. Then*

$$|x_1 \times x_2|^2 = g_{11}g_{22} - g_{12}^2 = \det(g_{ij}).$$

When both vectors are unit vectors, we obtain the usual relation

$$1 = \cos^2(\phi) + \sin^2(\phi)$$

where ϕ is the angle between the vectors.

12.2. Area elements of the surface and of the sphere

Consider an orientable surfaces Σ imbedded in 3-space. By the Binet–Cauchy identity 12.1.1, we have

$$\sqrt{\det(g_{ij})} = \left|x_1 \times x_2\right|,$$

where $x_i = \frac{\partial x}{\partial u^i}$.

DEFINITION 12.2.1. The area element dA_Σ of the surface Σ is

$$dA_\Sigma = \sqrt{\det(g_{ij})}du^1du^2 = |x_1 \times x_2|du^1du^2 \qquad (12.2.1)$$

where the g_{ij} are the metric coefficients of the surface with respect to the parametrisation $x(u^1, u^2)$.

This notion of area is discussed in more detail in Section 16.6.

We have used the subscript Σ so as to specify which surface we are dealing with.

An orientable surface by definition admits unit normal vector $N = N_p$, at each point $p \in \Sigma$. The vector field N is a globally defined field on the surface. Note that the image vector N can be thought of as an element of the unit sphere:

$$N_p \in S^2.$$

Each point $p \in \Sigma$ lies in a neighborhood parametrized by $x(u^1, u^2)$. At a point $p = x(u^1, u^2)$, we have a normal vector

$$N_p = N_{x(u^1,u^2)}.$$

The unit normal vector $n(u^1, u^2)$, obtained by normalizing the vector product $x_1 \times x_2$, coincides with the globally selected normal N:

$$n(u^1, u^2) = N_{x(u^1,u^2)}.$$

THEOREM 12.2.2. *If Gaussian curvature is nonzero at a point* $p = x(u_0^1, u_0^2) \in \Sigma$ *of the surface, then the map* $n(u^1, u^2)$ *from* Σ *to* S^2 *produces a regular parametrisation of a neighborhood of a given point* $n(u_0^1, u_0^2) \in S^2$ *on the sphere.*

PROOF. The parametrisation is given by the map to the sphere whose Jacobian is the Weingarten map. Regularity follows from the fact that Gaussian curvature is the determinant of the Weingarten map, hence nonzero by hypothesis. Hence the vectors n_1 and n_2 are linearly independent. □

Note that we are no longer thinking of n is a normal to the original surface Σ, but rather as a parametrisation of an open neighborhood on the unit sphere. Now consider the area element of the unit sphere.

THEOREM 12.2.3. *Consider a parametrisation* $n(u^1, u^2)$ *of a neighborhood on the sphere* S^2 *as in Theorem 12.2.2. Then the area element* dA_{S^2} *can be expressed as*

$$dA_{S^2} = \sqrt{\det(\tilde{g}_{\alpha\beta})}du^1du^2 = |n_1 \times n_2|du^1du^2,$$

where $\tilde{g}_{\alpha\beta}$ *are the metric coefficients of the parametrisation* $n(u^1, u^2)$ *of the sphere, and* $n_i = \frac{\partial n(u^1,u^2)}{\partial u^i}$.

PROOF. This is the usual formula for surfaces, applied to the chosen parametrisation $n(u^1, u^2)$ as defined in Theorem 12.2.2, in place of the traditional $\underline{x}(u^1, u^2)$. □

REMARK 12.2.4. We have used the subscript to distinguish the area element dA_{S^2} of the sphere S^2 from the area element dA_Σ of the surface Σ as in (12.2.1), where the g_{ij} are the metric coefficients of the surface with respect to the parametrisation $x(u^1, u^2)$. Note that we need to distinguish the two area elements

$$dA_{S^2} \quad \text{and} \quad dA_\Sigma$$

because both will occur in the proof of the Gauss–Bonnet theorem.

12.3. Proof of Gauss-Bonnet theorem

The Gauss–Bonnet theorem for surfaces is to a certain extent analogous to the theorem on the total curvature of a plane curve (Theorem 4.6.2). In both cases, an integral of curvature turns out to have topological significance.

In the notation of the previous section, we have the metric coefficients (g_{ij}) of the parametrisation $x(u^1, u^2)$ of the surface Σ, as well as the metric coefficients $(\tilde{g}_{\alpha,\beta})$ of the sphere relative to the parametrisation $n(u^1, u^2)$ stemming from the normal of Σ.

LEMMA 12.3.1. *We have the identity*

$$\det(\tilde{g}_{\alpha\beta}) = \big(K(u^1, u^2)\big)^2 \det(g_{ij})$$

where $K = K(u^1, u^2)$ *is the Gaussian curvature of the surface* Σ.

PROOF. Let $L = (L^i{}_j)$ be the matrix of W with respect to the basis (x_1, x_2). By definition of curvature we have $K = \det(L)$. Recall that the coefficients $L^i{}_j$ of the Weingarten map are defined by

$$n_\alpha = L^i{}_\alpha x_i = x_i L^i{}_\alpha. \tag{12.3.1}$$

Consider the 3×2-matrices $A = [x_1 \quad x_2]$ and $B = [n_1 \quad n_2]$. Then (12.3.1) implies by definition that

$$B = AL.$$

Therefore the Gram matrices satisfy

$$\mathrm{Gram}(n_1, n_2) = B^t B = (AL)^t AL = L^t A^t AL = L^t \,\mathrm{Gram}(x_1, x_2)\, L.$$

Applying the determinant, we complete the proof of the theorem. □

Note that by the Cauchy-Binet formula, the desired identity is equivalent to the formula

$$|n_{u^1} \times n_{u^2}| = |\det(L^i{}_j)|\, |x_{u^1} \times x_{u^2}|, \tag{12.3.2}$$

immediate from the observation that a linear map multiplies the area of parallelograms by its determinant. Namely, the Weingarten map sends each vector x_i to n_i.

THEOREM 12.3.2 (Special case of Gauss-Bonnet). *Let Σ be a convex closed surface in $\mathbb{R}^3$, i.e., an imbedded closed surface of positive Gaussian curvature (a typical example is an ellipsoid). Then the curvature integral satisfies*

$$\int_\Sigma K(p)\, dA_\Sigma = 4\pi = 2\chi(S^2),$$

where K is the Gaussian curvature function defined at every point p of the surface.

PROOF. The convexity of the surface guarantees that the map n is one-to-one (compare with the proof of Theorem 4.5.2 on closed curves).

We examine the integrand $K\, dA_\Sigma$ in a coordinate chart (u^1, u^2) where it can be written as $K(u^1, u^2)\, dA_\Sigma$. By Lemma 12.3.1, we have

$$K(u^1, u^2)\, dA_\Sigma = K(u^1, u^2)\sqrt{\det(g_{ij})}du^1du^2 = \sqrt{\det(\tilde{g}_{\alpha\beta})}du^1du^2 = dA_{S^2}.$$

Thus, the expression $K\, dA_\Sigma$ coincides with the area element dA_{S^2} of the unit sphere S^2 in every coordinate chart. Hence we can write

$$\int_\Sigma K\, dA_\Sigma = \int_{S^2} dA_{S^2} = 4\pi,$$

proving the theorem. □

12.4. Euler characteristic

The Euler characteristic of a closed imbedded surface in Euclidean 3-space can be defined via the total Gaussian curvature.

DEFINITION 12.4.1. The Euler characteristic $\chi(\Sigma)$ of a surface Σ is defined by the relation

$$2\pi\chi(\Sigma) = \int_\Sigma K(p) dA_\Sigma. \qquad (12.4.1)$$

The relation (12.4.1) is similar to the line intergal expression for the rotation index in formula (5.1.1). To show that this definition of the Euler characteristic agrees with the usual one, it is necessary to use the notion of algebraic degree for maps between surfaces, similar to the algebraic degree of a self-map of a circle.

CHAPTER 13

Duality in algebra, calculus, and geometry

13.1. Duality in linear algebra

The theorema egregium of Gauss marks the transition from classical differential geometry of curves and surfaces imbedded in 3-space, to modern differential geometry of surfaces (and manifolds) studied intrinsically. To formulate the intrinsic viewpoint, one needs the notion of duality of vector and covector.

Let V be a real vector space. We will assume all vector spaces to be finite dimensional unless stated otherwise. Euclidean space $\mathbb{R}^n$ is an example of a real vector space of dimension n.

Example 13.1.1. The tangent plane T_pM of a regular surface M at a point $p \in M$ is a real vector space of dimension 2.

Definition 13.1.2. A *linear form*, also called *1-form*, ϕ on V is a linear functional

$$\phi : V \to \mathbb{R}.$$

Example 13.1.3. In the usual plane of vectors

$$v = v^1 e_1 + v^2 e_2$$

represented by arrows, we denote by dx the 1-form which extracts the abscissa of the vector, and by dy the 1-form which extracts the ordinate of the vector:

$$dx(v) = v^1,$$

and

$$dy(v) = v^2.$$

Thus, for a vector $v = 3e_1 + 4e_2$ with components $(3, 4)$ we obtain

$$dx(v) = 3,\ dy(v) = 4.$$

Definition 13.1.4. We similarly define the corresponding quadratic forms dx^2 and dy^2, by squaring the value of the 1-form on on v. Such quadratic forms are called *rank-1* quadratic forms.

Thus,

$$dx^2(v) = (v^1)^2,\ dy^2(v) = (v^2)^2;$$

in the particular example $v = 3e_1 + 4e_2$ we obtain $dx^2(v) = 9,\ dy^2(v) = 16$.

DEFINITION 13.1.5. The *dual* space of V, denoted V^*, is the space of all linear forms on V:

$$V^* = \{\phi \mid \phi \text{ is a 1-form on } V\}.$$

Evaluating ϕ at an element $x \in V$ produces a scalar $\phi(x) \in \mathbb{R}$.

DEFINITION 13.1.6. The *evaluation map* is the natural pairing between V and V^*, namely a linear map denoted

$$\langle\ ,\ \rangle : V \times V^* \to \mathbb{R},$$

defined by evaluating y at x, i.e., setting $\langle x, y\rangle = y(x)$, for all $x \in V$ and $y \in V^*$.

REMARK 13.1.7. Note we are using the same notation for the pairing as for the scalar product in Euclidean space. The notation is quite widespread.

DEFINITION 13.1.8. If V admits a basis of vectors (x_i), then the dual space V^* admits a unique basis, called the *dual basis* (y_j), satisfying

$$\langle x_i, y_j\rangle = \delta_{ij}, \tag{13.1.1}$$

for all $i, j = 1, \ldots, n$, where δ_{ij} is the Kronecker delta function.

EXAMPLE 13.1.9. Let $V = \mathbb{R}^2$. We have the usual basis e_1, e_2 for V. The 1-forms dx, dy form a basis for the dual space V^*.

13.2. Duality in calculus; derivations

Let E be a Euclidean space of dimension n, and let $p \in E$ be a fixed point.

DEFINITION 13.2.1. Let

$$\mathbb{D}_p = \{f : f \in C^\infty\}$$

be the space of real C^∞-functions f defined in a neighborhood of $p \in E$.

Note that $\mathbb{D}_p$ is infinite-dimensional as it includes all polynomials.

THEOREM 13.2.2. *A partial derivative $\frac{\partial}{\partial u^i}$ at p a 1-form*

$$\frac{\partial}{\partial u^i} : \mathbb{D}_p \to \mathbb{R}$$

on the space $\mathbb{D}_p$, satisfying Leibniz's rule

$$\left.\frac{\partial(fg)}{\partial u^i}\right|_p = \left.\frac{\partial f}{\partial u^i}\right|_p g(p) + f(p) \left.\frac{\partial g}{\partial u^i}\right|_p . \tag{13.2.1}$$

The formula can be written briefly as

$$\tfrac{\partial}{\partial u^i}(fg) = \tfrac{\partial}{\partial u^i}(f)g + f\tfrac{\partial}{\partial u^i}(g),$$

at the point p. This was proved in calculus. Formula (13.2.1) motivates the following more general definition of a derivation.

DEFINITION 13.2.3. A *derivation* X at p is an $\mathbb{R}$-linear 1-form

$$X : \mathbb{D}_p \to \mathbb{R}$$

on the space $\mathbb{D}_p$ satisfying Leibniz's rule:

$$X(fg) = X(f)g(p) + f(p)X(g) \tag{13.2.2}$$

for all $f, g \in \mathbb{D}_p$.

REMARK 13.2.4. Linearity of a derivation is required only with regard to scalars in $\mathbb{R}$, not functions.

The following two results are familiar from advanced calculus.

PROPOSITION 13.2.5. *Let E be an n-dimensional Euclidean space, and $p \in E$. Then the space of all derivations at p is a vector space of dimension n.*

PROOF. We will prove the result in the case of a single variable u at the point $p = 0$ (the general case is similar).

Let X be a derivation. Then

$$X(1) = X(1 \cdot 1) = 2X(1)$$

by Leibniz's rule. Therefore $X(1) = 0$, and similarly for any constant by linearity of X.

Now consider the monic polynomial $u = u^1$ of degree 1, i.e., the linear function

$$u \in \mathbb{D}_{p=0}.$$

We evaluate the derivation X at u and set $c = X(u)$.

By the Taylor remainder formula, any function $f \in \mathbb{D}_{p=0}$ can be written as

$$f(u) = a + bu + g(u)u$$

where g is smooth and $g(0) = 0$. Now we have by linearity

$$\begin{aligned} X(f) &= X(a + bu + g(u)u) \\ &= bX(u) + X(g)u(0) + g(0) \cdot 1 \\ &= bc + 0 + 0 \\ &= c\frac{\partial}{\partial u}(f). \end{aligned}$$

Thus X coincides with $c\frac{\partial}{\partial u}$ for all $f \in \mathbb{D}_p$. Hence the tangent space is 1-dimensional, proving the theorem in this case. □

DEFINITION 13.2.6. The space of derivations at p is called the *tangent space* $T_p = T_pE$ at p.

PROPOSITION 13.2.7. *Let* $(u^1, \ldots, u^n)$ *define coordinates for* E. *Then a basis for the tangent space* T_p *is given by the* n *partial derivatives*

$$\left(\frac{\partial}{\partial u^i}\right), \quad i = 1, \ldots, n.$$

DEFINITION 13.2.8. The space dual to the tangent space T_p is called the *cotangent space*, and denoted T_p^*.

Thus an element of a tangent space is a vector, while an element of a cotangent space is called a 1-form, or a *covector*.

DEFINITION 13.2.9. The basis dual to the basis $\left(\frac{\partial}{\partial u^i}\right)$ is denoted

$$(du^j), \quad j = 1, \ldots, n.$$

Thus each du^j is by definition a linear form, or cotangent vector (covector for short). We are therefore working with dual bases $\left(\frac{\partial}{\partial u^i}\right)$ for vectors, and (du^j) for covectors. The evaluation map as in (13.1.1) gives

$$\left\langle \frac{\partial}{\partial u^i}, du^j \right\rangle = du^j \left(\frac{\partial}{\partial u^i}\right) = \delta_i^j, \tag{13.2.3}$$

where δ_i^j is the Kronecker delta.

13.3. Constructing bilinear forms out of 1-forms

Recall that the polarisation formula (see definition 1.5.4) allows one to reconstruct a symmetric bilinear form $B = B(v, w)$, from the quadratic form $Q(v) = B(v, v)$, at least if the characteristic is not 2:

$$B(v, w) = \frac{1}{4}(Q(v + w) - Q(v - w)). \tag{13.3.1}$$

Similarly, one can construct bilinear forms out of the 1-forms du^i, as follows. Consider a quadratic form defined by a linear combination of the rank-1 quadratic forms $(du^i)^2$, as in Definition 13.1.4. Polarizing the quadratic form, one obtains a bilinear form on the tangent space T_p.

EXAMPLE 13.3.1. Let $\underline{v} = v^1 e_1 + v^2 e_2$ be an arbitrary vector in the plane. Let dx and dy be the standard covectors, extracting, respectively, the first and second coordinates of v. Consider the quadratic form Q given by

$$Q = E dx^2 + F dy^2.$$

Here $Q(\underline{v})$ is calculated as

$$Q(\underline{v}) = E(dx(\underline{v}))^2 + F(dy(\underline{v}))^2 = E(v^1)^2 + F(v^2)^2.$$

Polarisation then produces the bilinear form $B = B(\underline{v}, \underline{v}')$, where $\underline{v}$ and $\underline{v}'$ are arbitrary vectors, given by the formula

$$B(\underline{v}, \underline{v}') = E\, dx(\underline{v})\, dx(\underline{v}') + F\, dy(\underline{v})\, dy(\underline{v}').$$

Example 13.3.2. Setting $E = F = 1$ in the previous example, we obtain the standard scalar product in the plane:

$$B(v, v') = v \cdot v' = dx(v)\, dx(v') + dy(v)\, dy(v') = v_1 v_1' + v_2 v_2'.$$

13.4. First fundamental form

Recall that the first fundamental form g is a symmetric bilinear form on the tangent space at p:

$$g : T_p \times T_p \to \mathbb{R},$$

defined for all p and varying continuously and smoothly in p.

Recall that the basis for T_p in coordinates (u^i) is given by the tangent vectors

$$\frac{\partial}{\partial u^i}.$$

These are given by certain derivations (see Section 13.2).

The first fundamental form g is traditionally expressed by a matrix of coefficients called *metric coefficients* g_{ij}, given by the inner product of the i-th and the j-th vector in the basis:

$$g_{ij} = g\left(\frac{\partial}{\partial u^i}, \frac{\partial}{\partial u^j}\right),$$

where g is the first fundamental form. In particular, the coefficient

$$g_{ii} = \left\|\frac{\partial}{\partial u^i}\right\|^2$$

is the square length of the i-th vector.

We will express the first fundamental form in more intrinsic notation of quadratic forms built from 1-forms (covectors).

13.5. Dual bases in differential geometry

Let us now restrict attention to the case of 2 dimensions, i.e. the case of surfaces. At every point $p = (u^1, u^2)$, we have the metric coefficients $g_{ij} = g_{ij}(u^1, u^2)$. Each metric coefficient is thus a function of two variables.

We will only consider the case when the matrix is diagonal. This can always be achieved, in two dimensions, at a point by a suitable change of coordinates, by the uniformisation theorem (see Section 13.6). We set $x = u^1$ and $y = u^2$ to simplify notation. In the notation developed in Section 13.3, we can write the first fundamental form as follows:

$$g = g_{11}(x, y)(dx)^2 + g_{22}(x, y)(dy)^2. \tag{13.5.1}$$

For example, if the metric coefficients form an identity matrix: $g_{ij} = \delta_{ij}$, we obtain the standard flat metric

$$g = (dx)^2 + (dy)^2 . \tag{13.5.2}$$

or simply as

$$g = dx^2 + dy^2.$$

EXAMPLE 13.5.1 (Hyperbolic metric). Let $g_{11} = g_{22} = \frac{1}{y^2}$. The resulting hyperbolic metric in the upperhalf plane $\{y > 0\}$ is expressed by the quadratic form

$$\frac{1}{y^2}\left(dx^2 + dy^2\right).$$

(note that this expression is undefined for $y = 0$). The hyperbolic metric in the upper half plane is a complete metric.

13.6. Uniformisation theorem

Closely related results are the Riemann mapping theorem and the conformal representation theorem.

THEOREM 13.6.1 (Riemann mapping/uniformisation). *Every metric on a connected*[1] *surface is conformally equivalent to a metric of constant Gaussian curvature.*

From the complex analytic viewpoint, the uniformisation theorem states that every Riemann surface is covered by either the sphere, the plane, or the upper halfplane. Thus no notion of curvature is needed for the statement of the uniformisation theorem. However, from the differential geometric point of view, what is relevant is that every conformal class of metrics contains a metric of constant Gaussian curvature. See [**Ab81**] for a lively account of the history of the uniformisation theorem.

[1] kashir

13.7. Surfaces of revolution in isothermal coordinates

The uniformisation theorem as stated it in Section 13.6 is an existence theorem. It does not not provide an explicit recipee as to how one would go about finding a representation of the surface in terms of a metric of constant curvature, together with a suitable conformal factor.

In this chapter we describe an explicit presentation of this kind in the case of tori of revolution.

We will make use of the differential notation for metrics developed in earlier chapters.

The following lemma expresses the metric of a surface of revolution in isothermal coordinates. Recall that if a surface is obtained by revolving a curve $(f(\phi), g(\phi))$, we obtain metric coefficients $g_{11} = f^2$ and $g_{22} = \left(\frac{df}{d\phi}\right)^2 + \left(\frac{dg}{d\phi}\right)^2$.

In other words, the metric can be written as

$$f^2 d\theta^2 + \left(\left(\frac{df}{d\phi}\right)^2 + \left(\frac{dg}{d\phi}\right)^2\right) d\phi^2. \tag{13.7.1}$$

LEMMA 13.7.1. *Consider an arc length parametrisation $(f(\phi), g(\phi))$, where $f(\phi) > 0$, of the generating curve of a surface of revolution. Then the change of variable*

$$\psi = \int \frac{1}{f(\phi)} d\phi,$$

produces an isothermal *parametrisation in terms of variables (θ, ψ). With respect to the new coordinates, the first fundamental form is given by a scalar matrix $(g_{ij}) = (f(\phi(\psi))^2 \delta_{ij})$, i.e. the metric is*

$$f(\phi(\psi))^2 (d\theta^2 + d\psi^2).$$

In other words, we obtain an explicit conformal equivalence between the metric on the surface of revolution and the standard flat metric on the quotient of the (θ, ψ) plane. Such coordinates are referred to as "isothermal coordinates" in the literature. The existence of such a parametrisation is of course predicted by the uniformisation theorem (Theorem 13.6.1) in the case of a general surface.

PROOF. Consider an arbitrary change of parameter $\phi = \phi(\psi)$. By chain rule,

$$\frac{df}{d\psi} = \frac{df}{d\phi} \frac{d\phi}{d\psi}.$$

Now consider again the first fundamental form (13.7.1). To impose the condition $g_{11} = g_{22}$, we need to solve the equation $f^2 = \left(\frac{df}{d\psi}\right)^2 + \left(\frac{dg}{d\psi}\right)^2$, or

$$f^2 = \left(\left(\frac{df}{d\phi} \right)^2 + \left(\frac{dg}{d\phi} \right)^2 \right) \left(\frac{d\phi}{d\psi} \right)^2 .$$

In the case when the generating curve is parametrized by arclength, we are therefore reduced to the equation

$$f = \frac{d\phi}{d\psi},$$

or $\psi = \int \frac{d\phi}{f(\phi)}$. Substituting the new variable $\phi(\psi)$ in place of ϕ in the parametrisation of the surface, we obtain a new parametrisation of the surface of revolution in coordinates (θ, ψ), such that the matrix of metric coefficients is a scalar matrix by construction. $\square$

CHAPTER 14

Tori, residues

14.1. More on dual bases

Recall that if $(x_1, \ldots, x_n)$ is a basis for a vector space V then the dual vector space V^* possesses a basis called the *dual basis* and denoted $(y_1, \ldots, y_n)$ satisfying $\langle x_i, y_j\rangle = y_j(x_i) = \delta_{ij}$.

EXAMPLE 14.1.1. In $\mathbb{R}^2$ we have a basis $(x_1, x_2) = \left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}\right)$ in the tangent plane T_p at a point p. The dual basis of 1-forms (y_1, y_2) for T_p^* is denoted (dx, dy). Thus we have

$$\left\langle \frac{\partial}{\partial x}, dx \right\rangle = dx\left(\frac{\partial}{\partial x}\right) = 1$$

and

$$\left\langle \frac{\partial}{\partial y}, dy \right\rangle = dy\left(\frac{\partial}{\partial y}\right) = 1,$$

while

$$\left\langle \frac{\partial}{\partial x}, dy \right\rangle = dy\left(\frac{\partial}{\partial x}\right) = 0,$$

etcetera.

Similarly, in polar coordinates at a point $p \neq 0$ we have a basis $\left(\frac{\partial}{\partial r}, \frac{\partial}{\partial \theta}\right)$ for T_p, and a dual basis $(dr, d\theta)$ for T_p^*. Thus we have

$$\left\langle \frac{\partial}{\partial r}, dr \right\rangle = dr\left(\frac{\partial}{\partial r}\right) = 1$$

and

$$\left\langle \frac{\partial}{\partial \theta}, d\theta \right\rangle = d\theta\left(\frac{\partial}{\partial \theta}\right) = 1,$$

while

$$\left\langle \frac{\partial}{\partial r}, d\theta \right\rangle = d\theta\left(\frac{\partial}{\partial r}\right) = 0,$$

etcetera.

Now in polar coordinates we have a natural area element $r\,dr\,d\theta$. Areas are calculated by Fubini's theorem as

$$\iint_D r\,dr\,d\theta = \int \left(\int r dr \right) d\theta.$$

Thus we have a natural basis $(y_1, y_2) = (rdr, d\theta)$ in T_p^* when $p \neq 0$, i.e., $y_1 = rdr$ while $y_2 = d\theta$. Its dual basis (x_1, x_2) in T_p can be easily identified. It is

$$(x_1, x_2) = \left(\frac{1}{r} \frac{\partial}{\partial r}, \frac{\partial}{\partial \theta} \right).$$

Indeed, we have

$$\langle x_1, y_1 \rangle = \left\langle \frac{1}{r} \frac{\partial}{\partial r}, rdr \right\rangle = rdr \left(\frac{1}{r} \frac{\partial}{\partial r} \right) = r \frac{1}{r}\, dr \left(\frac{\partial}{\partial r} \right) = 1,$$

etcetera.

14.2. Conformal parameter τ of tori of revolution

The results of Section 13.7 have the following immediate consequence.

COROLLARY 14.2.1. *Consider a torus of revolution in $\mathbb{R}^3$ formed by rotating a Jordan curve of length $L > 0$, with unit speed parametisation $(f(\phi), g(\phi))$ where $\phi \in [0, L]$. Then the torus is conformally equivalent to a flat torus*

$$\mathbb{R}^2 / L_{c,d}.$$

Here $\mathbb{R}^2$ is the (θ, ψ)-plane, where ψ is the antiderivative of $\frac{1}{f(\phi)}$ as in Section 13.7; while the rectangular lattice $L_{c,d} \subset \mathbb{R}^2$ is spanned by the orthogonal vectors $c\frac{\partial}{\partial \theta}$ and $d\frac{\partial}{\partial \psi}$, so that

$$L_{c,d} = \text{Span} \left(c \frac{\partial}{\partial \theta}\ ,\ d \frac{\partial}{\partial \psi} \right) = c\mathbb{Z} \oplus d\mathbb{Z},$$

where $c = 2\pi$ and $d = \int_0^L \frac{d\phi}{f(\phi)}$.

In Section 6.2 we showed that every flat torus $\mathbb{C}/L$ is similar to the torus spanned by $\tau \in \mathbb{C}$ and $1 \in \mathbb{C}$, where τ is in the standard fundamental domain

$$D = \{ z = x + iy \in \mathbb{C} : |x| \leq \tfrac{1}{2}, y > 0, |z| \geq 1 \}.$$

Then τ is called the *conformal parameter* of the torus.

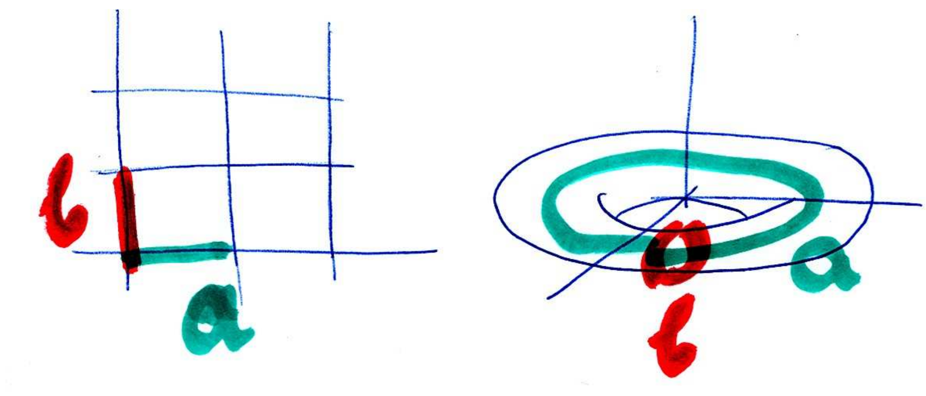

FIGURE 14.2.1. Torus: lattice (left) and imbedding (right)

COROLLARY 14.2.2. *The conformal parameter τ of a torus of revolution is pure imaginary:*

$$\tau = i\sigma^2$$

of absolute value

$$\sigma^2 = \max\left\{\frac{c}{d}\, ,\ \frac{d}{c}\right\} \geq 1.$$

PROOF. The proof is immediate from the fact that the lattice is rectangular. □

14.3. θ-loops and ϕ-loops on tori of revolution

Consider a torus of revolution (T^2, g) generated by a Jordan curve C in the (x, z)-plane, i.e., by a simple loop C, parametrized by a pair of functions $f(\phi)$, $g(\phi)$, so that $x = f(\phi)$ and $z = g(\phi)$.

DEFINITION 14.3.1. A ϕ-loop on the torus is a simple loop obtained by fixing the coordinate θ (i.e., the variable ϕ is changing). A θ-loop on the torus is a simple loop obtained by fixing the coordinate ϕ (i.e., the variable θ is changing).

PROPOSITION 14.3.2. *All ϕ-loops on the torus of revolution have the same length equal to the length L of the generating curve C (see Corollary 14.2.1).*

PROOF. The surface is rotationally invariant. In other words, all rotations around the z-axis are isometries. Therefore all ϕ-loops have the same length. □

Proposition 14.3.3. *The θ-loops on the torus of revolution have variable length, depending on the ϕ-coordinate of the loop. Namely, the length is $2\pi x = 2\pi f(\phi)$.*

Proof. The proof is immediate from the fact that the function $f(\phi)$ gives the distance r to the z-axis. □

Definition 14.3.4. We denote by λ_ϕ the (common) length of all ϕ-loops on a torus of revolution.

Definition 14.3.5. We denote by $\lambda_{\theta_{\min}}$ the least length of a θ-loop on a torus of revolution, and by $\lambda_{\theta_{\max}}$ the maximal length of such a θ-loop.

14.4. Tori generated by round circles

Let $a, b > 0$. We assume $a > b$ so as to obtain tori that are imbedded in 3-space. We consider the 2-parameter family $g_{a,b}$ of tori of revolution in 3-space with circular generating loop. The torus of revolution $g_{a,b}$ generated by a round circle is the locus of the equation

$$(r - a)^2 + z^2 = b^2, \tag{14.4.1}$$

where $r = \sqrt{x^2 + y^2}$. Note that the angle θ of the cylindrical coordinates (r, θ, z) does not appear in the equation (14.4.1). The torus is obtained by rotating the circle

$$(x - a)^2 + z^2 = b^2 \tag{14.4.2}$$

around the z-axis in $\mathbb{R}^3$. The torus admits a parametrisation in terms of the functions[1] $f(\phi) = a + b\cos\phi$ and $g(\phi) = b\sin\phi$. Namely, we have

$$x(\theta, \phi) = ((a + b\cos\phi)\cos\theta, (a + b\cos\phi)\sin\theta, b\sin\phi). \tag{14.4.3}$$

Here the θ-loop (see Section 14.3) has length $2\pi(a + b\cos\phi)$. The shortest θ-loop is therefore of length

$$\lambda_{\theta_{\min}} = 2\pi(a - b),$$

and the longest one is

$$\lambda_{\theta_{\max}} = 2\pi(a + b).$$

Meanwhile, the ϕ-loop has length

$$\lambda_\phi = 2\pi b.$$

[1]We use ϕ here and φ for the modified arclength parameter in the next section

14.5. Conformal parameter of tori of revolution, residues

We would like to compute the conformal parameter τ of the standard tori as in (14.4.3). We first modify the parametrisation so as to obtain a generating curve parametrized by arclength:

$$f(\varphi) = a + b\cos\frac{\varphi}{b}, \quad g(\varphi) = b\sin\frac{\varphi}{b}, \tag{14.5.1}$$

where $\varphi \in [0, L]$ with $L = 2\pi b$.

THEOREM 14.5.1. *The corresponding flat torus is given by the lattice L in the (θ, ψ) plane of the form*

$$L = \mathrm{Span}_{\mathbb{Z}}\left(c\frac{\partial}{\partial\theta}\ ,\ d\frac{\partial}{\partial\phi}\right)$$

where $c = 2\pi$ and

$$d = \frac{2\pi}{\sqrt{(a/b)^2 - 1}}.$$

Thus the conformal parameter τ of the flat torus satisfies

$$\tau = i\max\left(((a/b)^2 - 1)^{-1/2}, ((a/b)^2 - 1)^{1/2}\right).$$

PROOF. By Corollary 14.2.2, replacing φ by $\varphi(\psi)$ produces isothermal coordinates (θ, ψ) for the torus generated by (14.5.1), where

$$\psi = \int \frac{d\varphi}{f(\varphi)} = \int \frac{d\varphi}{a + b\cos\frac{\varphi}{b}},$$

and therefore the flat metric is defined by a lattice in the (θ, ψ) plane with $c = 2\pi$ and

$$d = \int_0^{L=2\pi b} \frac{d\varphi}{a + b\cos\frac{\varphi}{b}}.$$

Changing the variable to to $\phi = \frac{\varphi}{b}$ we obtain

$$d = \int_0^{2\pi} \frac{d\phi}{(a/b) + \cos\phi},$$

where $a/b > 1$. Let $a' = a/b$. Now the integral is the real part Re of the complex integral

$$d = \int_0^{2\pi} \frac{d\phi}{a' + \cos\phi} = \int \frac{d\phi}{a' + \mathrm{Re}(e^{i\phi})}.$$

Thus

$$d = \int \frac{2d\phi}{2a' + e^{i\phi} + e^{-i\phi}}. \tag{14.5.2}$$

The change of variables $z = e^{i\phi}$ yields

$$d\phi = \frac{-idz}{z}$$

and along the circle we have

$$d = \oint \frac{-2idz}{z(2a' + z + z^{-1})} = \oint \frac{-2idz}{z^2 + 2a'z + 1} = \oint \frac{-2idz}{(z-\lambda_1)(z-\lambda_2)},$$

where $\lambda_1 = -a' + \sqrt{a'^2 - 1}$ and $\lambda_2 = -a' - \sqrt{a'^2 - 1}$. The root λ_2 is outside the unit circle. Hence we need the residue at λ_1 to apply the residue theorem. The residue at λ_1 equals

$$\mathrm{Res}_{\lambda_1} = \frac{-2i}{\lambda_1 - \lambda_2} = \frac{-2i}{2\sqrt{a'^2 - 1}} = \frac{-i}{\sqrt{a'^2 - 1}}.$$

The integral is determined by the residue theorem in terms of the residue at the pole $z = \lambda_1$. Therefore the lattice parameter d from Corollary 14.2.2 can be computed from (14.5.2) as

$$d = \mathrm{Re}\big(2\pi i \,\mathrm{Res}_{\lambda_1}\big) = \frac{2\pi}{\sqrt{(a')^2 - 1}}.$$

proving the theorem. □

CHAPTER 15

Loewner's systolic inequality

15.1. Definition of systole

The unit circle $S^1 \subset \mathbb{C}$ bounds the unit disk $\mathbb{D}$. A loop on a surface M is a continuous map $S^1 \to M$.

A loop $S^1 \to M$ is called contractible if the map f extends from S^1 to the disk D by means of a continuous map $F : \mathbb{D} \to M$. Thus the restriction of F to S^1 is f.

The notions of contractible loops and simply connected spaces were reviewed in more detail in Section 16.18.

A loop is called non-contractible if it is not contractible.

Given a metric g on M, we will denote by $\mathrm{sys}_1(g)$, the infimum of lengths, referred to as the "systole" of g, of a noncontractible loop β in a compact, non-simply-connected Riemannian manifold (M, g):

$$\mathrm{sys}_1(g) = \inf_{\beta} \mathrm{length}(\beta), \tag{15.1.1}$$

where the infimum is over all noncontractible loops β in M. In graph theory, a similar invariant is known as the *girth* [**Tu47**].[1] See Section 10.3 for a further discussion of the term, by its promulgator.

It can be shown that for a compact Riemannian manifold, the infimum is always attained, *cf.* Theorem 16.10.1. A loop realizing the minimum is necessarily a simple closed geodesic.

In systolic questions about surfaces, integral-geometric identities play a particularly important role. Roughly speaking, there is an integral identity relating area on the one hand, and an average of energies of a suitable family of loops, on the other. By the Cauchy-Schwarz inequality, there is an inequality relating energy and length squared, hence one obtains an inequality between area and the square of the systole.

Such an approach works both for the Loewner inequality (15.2.1) and Pu's inequality (15.1.6) (biographical notes on C. Loewner and

[1]The notion of systole expressed by (15.1.1) is unrelated to the systolic arrays of [**Ku78**].

P. Pu appear in respectively). One can prove an inequality for the Möbius band this way, as well [**Bl61b**].

Here we prove the two classical results of systolic geometry, namely Loewner's torus inequality as well as Pu's inequality for the real projective plane.

15.1.1. Three systolic invariants. The material in this section is optional.

Let M be a Riemannian manifold. We define the homology 1-systole

$$\mathrm{sys}_1(M) \tag{15.1.2}$$

by minimizing vol(α) over all nonzero homology classes. Namely, $\mathrm{sys}_1(M)$ is the least length of a loop C representing a nontrivial homology class $[C]$ in $H_1(M;\mathbb{Z})$.

We also define the stable homology systole

$$\mathrm{stsys}_1(M) = \lambda_1\left(H_1(M)/T_1,\ \|\ \|\right), \tag{15.1.3}$$

namely by minimizing the stable norm $\|\ \|$ of a class of infinite order (see Definition 16.30.1 for details).

Remark 15.1.1. For the real projective plane, these two systolic invariants are not the same. Namely, the homology systole sys_1 equals the least length of a noncontractible loop (which is also nontrivial homologically), while the stable systole is infinite being defined by a minimum over an empty set.

Recall the following example from the previous section:

Example 15.1.2. For an arbitrary metric on the 2-torus $\mathbb{T}^2$, the 1-systole and the stable 1-systole coincide by Theorem 16.22.3:

$$\mathrm{sys}_1(\mathbb{T}^2) = \mathrm{stsys}_1(\mathbb{T}^2),$$

for every metric on $\mathbb{T}^2$.

Using the notion of a noncontractible loop, we can define the homotopy 1-systole

$$\mathrm{sys}_1(M) \tag{15.1.4}$$

as the least length of a non-contractible loop in M.

In the case of the torus, the fundamental group $\mathbb{Z}^2$ is abelian and torsionfree, and therefore $\mathrm{sys}_1(\mathbb{T}^2) = \mathrm{sys}_1(\mathbb{T}^2)$, so that all three invariants coincide in this case.

15.1.2. Isoperimetric inequality and Pu's inequality. The material in this section is optional.

Pu's inequality can be thought of as an "opposite" isoperimetric inequality, in the following precise sense.

The classical isoperimetric inequality in the plane is a relation between two metric invariants: length L of a simple closed curve in the plane, and area A of the region bounded by the curve. Namely, every simple closed curve in the plane satisfies the inequality

$$\frac{A}{\pi} \leq \left(\frac{L}{2\pi}\right)^2.$$

This classical *isoperimetric inequality* is sharp, insofar as equality is attained only by a round circle.

In the 1950's, Charles Loewner's student P. M. Pu [**Pu52**] proved the following theorem. Let $\mathbb{RP}^2$ be the real projective plane endowed with an arbitrary metric, *i.e.* an imbedding in some $\mathbb{R}^n$. Then

$$\left(\frac{L}{\pi}\right)^2 \leq \frac{A}{2\pi}, \tag{15.1.5}$$

where A is its total area and L is the length of its shortest noncontractible loop. This *isosystolic inequality*, or simply *systolic inequality* for short, is also sharp, to the extent that equality is attained only for a metric of constant Gaussian curvature, namely antipodal quotient of a round sphere, *cf.* Section 16.11. In our systolic notation (15.1.1), Pu's inequality takes the following form:

$$\operatorname{sys}_1(g)^2 \leq \tfrac{\pi}{2} \operatorname{area}(g), \tag{15.1.6}$$

for every metric g on $\mathbb{RP}^2$. See Theorem 16.13.2 for a discussion of the constant. The inequality is proved in Section 18.5. Pu's inequality can be generalized as follows. We will say that a surface is *aspherical* if it is not a 2-sphere.

THEOREM 15.1.3. *Every aspherical surface* (Σ, g) *satisfies the optimal bound* (15.1.6)*, attained precisely when, on the one hand, the surface* Σ *is a real projective plane, and on the other, the metric* g *is of constant Gaussian curvature.*

The extension to aspherical surfaces follows from Gromov's inequality (15.1.7) below (by comparing the numerical values of the two constants). Namely, every aspherical compact surface (Σ, g) admits a metric ball

$$B = B_p\left(\tfrac{1}{2}\operatorname{sys}_1(g)\right) \subset \Sigma$$

of radius $\frac{1}{2}\operatorname{sys}_1(g)$ which satisfies [**Gro83**, Corollary 5.2.B]

$$\operatorname{sys}_1(g)^2 \le \frac{4}{3}\operatorname{area}(B). \tag{15.1.7}$$

15.1.3. Hermite and Bergé-Martinet constants. The material in this subsection is optional.

Most of the material in this section has already appeared in earlier chapters.

Let $b \in \mathbb{N}$. The Hermite constant γ_b is defined in one of the following two equivalent ways:

(1) γ_b is the *square* of the biggest first successive minimum, *cf.* Definition 16.14.1, among all lattices of unit covolume;
(2) γ_b is defined by the formula

$$\sqrt{\gamma_b} = \sup\left\{ \frac{\lambda_1(L)}{\operatorname{vol}(\mathbb{R}^b/L)^{1/b}} \,\middle|\, L \subseteq (\mathbb{R}^b, \|\ \|) \right\}, \tag{15.1.8}$$

where the supremum is extended over all lattices L in $\mathbb{R}^b$ with a Euclidean norm $\|\ \|$.

A lattice realizing the supremum is called a *critical* lattice. A critical lattice may be thought of as the one realizing the densest packing in $\mathbb{R}^b$ when we place balls of radius $\frac{1}{2}\lambda_1(L)$ at the points of L.

The existence of the Hermite constant, as well as the existence of critical lattices, are both nontrivial results [**Ca71**].

Theorem 6.3.1 provides the value for γ_2.

Example 15.1.4. In dimensions $b \ge 3$, the Hermite constants are harder to compute, but explicit values (as well as the associated critical lattices) are known for small dimensions (≤ 8), *e.g.* $\gamma_3 = 2^{\frac{1}{3}} = 1.2599\ldots$, while $\gamma_4 = \sqrt{2} = 1.4142\ldots$. Note that γ_n is asymptotically linear in n, *cf.* (15.1.11).

A related constant γ'_b is defined as follows, *cf.* [**BeM**].

Definition 15.1.5. The Bergé-Martinet constant γ'_b is defined by setting

$$\gamma'_b = \sup\left\{ \lambda_1(L)\lambda_1(L^*) \,\middle|\, L \subseteq (\mathbb{R}^b, \|\ \|) \right\}, \tag{15.1.9}$$

where the supremum is extended over all lattices L in $\mathbb{R}^b$.

Here L^* is the lattice dual to L. If L is the $\mathbb{Z}$-span of vectors (x_i), then L^* is the $\mathbb{Z}$-span of a dual basis (y_j) satisfying $\langle x_i, y_j\rangle = \delta_{ij}$, *cf.* relation (16.4.1).

Thus, the constant γ'_b is bounded above by the Hermite constant γ_b of (15.1.8). We have $\gamma'_1 = 1$, while for $b \geq 2$ we have the following inequality:

$$\gamma'_b \leq \gamma_b \leq \frac{2}{3}\, b \quad \text{for all } b \geq 2. \tag{15.1.10}$$

Moreover, one has the following asymptotic estimates:

$$\frac{b}{2\pi e}(1+o(1)) \leq \gamma'_b \leq \frac{b}{\pi e}\,(1+o(1)) \quad \text{for } b \to \infty, \tag{15.1.11}$$

cf. [**LaLS90**, pp. 334, 337]. Note that the lower bound of (15.1.11) for the Hermite constant and the Bergé-Martinet constant is nonconstructive, but see [**RT90**] and [**ConS99**].

Definition 15.1.6. A lattice L realizing the supremum in (15.1.9) or (15.1.9) is called *dual-critical.*

Remark 15.1.7. The constants γ'_b and the dual-critical lattices in $\mathbb{R}^b$ are explicitly known for $b \leq 4$, *cf.* [**BeM**, Proposition 2.13]. In particular, we have $\gamma'_1 = 1$, $\gamma'_2 = \frac{2}{\sqrt{3}}$.

Example 15.1.8. In dimension 3, the value of the Bergé-Martinet constant, $\gamma'_3 = \sqrt{\frac{3}{2}} = 1.2247\ldots$, is slightly below the Hermite constant $\gamma_3 = 2^{\frac{1}{3}} = 1.2599\ldots$. It is attained by the face-centered cubic lattice, which is not isodual [**MilH73**, p. 31], [**BeM**, Proposition 2.13(iii)], [**CoS94**].

This is the end of the three subsections containing optional material.

15.2. Loewner's torus inequality

Historically, the first lower bound for the volume of a Riemannian manifold in terms of a systole is due to Charles Loewner. In 1949, Loewner proved the first systolic inequality, in a course on Riemannian geometry at Syracuse University, *cf.* [**Pu52**]. Namely, he showed the following result, whose proof appears in Section 18.2.

Theorem 15.2.1 (C. Loewner). *Every Riemannian metric g on the torus $\mathbb{T}^2$ satisfies the inequality*

$$\operatorname{sys}_1(g)^2 \leq \gamma_2 \ \operatorname{area}(g), \tag{15.2.1}$$

where $\gamma_2 = \frac{2}{\sqrt{3}}$ is the Hermite constant (15.1.8). *A metric attaining the optimal bound* (15.2.1) *is necessarily flat, and is homothetic to the quotient of $\mathbb{C}$ by the Eisenstein integers, i.e. lattice spanned by the cube roots of unity, cf. Lemma 6.3.1.*

The result can be reformulated in a number of ways.[2] Loewner's torus inequality relates the total area, to the systole, *i.e.* least length of a noncontractible loop on the torus $(\mathbb{T}^2, g)$:

$$\operatorname{area}(g) - \tfrac{\sqrt{3}}{2}\operatorname{sys}_1(g)^2 \geq 0. \tag{15.2.2}$$

The boundary case of equality is attained if and only if the metric is homothetic to the flat metric obtained as the quotient of $\mathbb{R}^2$ by the lattice formed by the Eisenstein integers.

15.3. Loewner's inequality with defect

Loewner's torus inequality can be strengthened by introducing a "defect" (shegiya, she'erit) term, similar to Bonnesen's strengthening of the isoperimetic inequality. To write it down, we need to review the conformal representation theorem (uniformisation theorem).

By the conformal representation theorem, we can assume that the metric g on the torus $\mathbb{T}^2$ is of the form

$$f^2(dx^2 + dy^2),$$

with respect to a unit area flat metric $dx^2 + dy^2$ on the torus $\mathbb{T}^2$ viewed as a quotient $\mathbb{R}^2/L$ of the (x, y) plane by a lattice. The defect term in question is simply the variance (**shonut**) of the conformal factor f above. The inequality with the defect term looks as follows.

THEOREM 15.3.1. *Every metric on the torus satisfies the following strengthened form of Loewner's inequality:*

$$\operatorname{area}(g) - \tfrac{\sqrt{3}}{2}\operatorname{sys}(g)^2 \geq \operatorname{Var}(f). \tag{15.3.1}$$

Here the error term, or *isosystolic defect*, is given by the variance[3]

$$\operatorname{Var}(f) = \int_{\mathbb{T}^2} (f - m)^2 \tag{15.3.2}$$

of the conformal factor f of the metric $g = f^2(dx^2 + dy^2)$ on the torus, relative to the unit area flat metric $g_0 = dx^2 + dy^2$ in the same conformal class. Here

$$m = \int_{\mathbb{T}^2} f dxdy \tag{15.3.3}$$

[2] Thus, In the case of the torus $\mathbb{T}^2$, the fundamental group is abelian. Hence the systole can be expressed in this case as follows: $\operatorname{sys}_1(\mathbb{T}^2) = \lambda_1\left(H_1(\mathbb{T}^2;\mathbb{Z}), \|\ \|\right)$, where $\|\ \|$ is the stable norm.

[3] shonut

is the mean[4] of f. More concretely, if $(\mathbb{T}^2, g_0) = \mathbb{R}^2/L$ where L is a lattice of unit coarea, and D is a fundamental domain for the action of L on $\mathbb{R}^2$ by translations, then the integral (15.3.3) can be written as

$$m = \int_D f(x,y)dxdy$$

where $dxdy$ is the standard Lebesgue measure of $\mathbb{R}^2$.

15.4. Computational formula for the variance

The proof of inequalities with isosystolic defect relies upon the familiar computational formula for the variance[5] of a random variable[6] X in terms of expected values.

Namely, we have the formula

$$E_\mu(X^2) - (E_\mu(X))^2 = \mathrm{Var}(X), \tag{15.4.1}$$

where μ is a probability measure. Here the variance is

$$\mathrm{Var}(X) = E_\mu\left((X-m)^2\right),$$

where $m = E_\mu(X)$ is the expected value (*i.e.* the mean) (tochelet).

15.5. An application of the computational formula

Keeping our differential geometric application in mind, we will denote the random variable (mishtaneh akra'i) f.

Now consider a flat metric g_0 of unit area on the 2-torus $\mathbb{T}^2$. Denote the associated measure by μ. In other words, μ is given by the usual Lebesgue measure $dxdy$. Since μ is a probability measure, we can apply formula (15.4.1) to it. Consider a metric $g = f^2 g_0$ conformal to the flat one, with conformal factor $f > 0$. Then we have

$$E_\mu(f^2) = \int_{\mathbb{T}^2} f^2 dxdy = \mathrm{area}(g),$$

since $f^2 dx\, dy$ is precisely the area element of the metric g.

Equation (15.4.1) therefore becomes

$$\mathrm{area}(g) - (E_\mu(f))^2 = \mathrm{Var}(f). \tag{15.5.1}$$

Next, we will relate the expected value $E_\mu(f)$ to the systole of the metric g. Then we will then relate (15.4.1) to Loewner's torus inequality.

---

[4]tochelet (with "het")

[5]shonut

[6]mishtaneh mikri

By the uniformisation theorem, every metric on the torus $\mathbb{T}^2$ is conformally equivalent to a flat metric of unit area. Denoting such a flat metric by g_0, such equivalence can be written as

$$g = f^2 g_0.$$

15.6. Conformal invariant σ

Let us prove Loewner's torus inequality for the metric $g = f^2 g_0$ on $\mathbb{T}^2$, using the computational formula for the variance. We first analyze the expected value term

$$E_\mu(f) = \int_{\mathbb{T}^2} f dx dy$$

in (15.5.1).

By the proof of Lemma 6.3.1, the lattice of deck transformations of the flat torus g_0 admits a $\mathbb{Z}$-basis similar (domeh) to $\{\tau, 1\} \subset \mathbb{C}$, where τ belongs to the standard fundamental domain (6.2.1). In other words, the lattice is similar to

$$\mathbb{Z}\tau + \mathbb{Z}1 \subset \mathbb{C}.$$

Definition 15.6.1. We define the conformal invariant σ as follows: consider the imaginary part $\mathrm{Im}(\tau)$ and set

$$\sigma^2 := \mathrm{Im}(\tau) > 0.$$

Lemma 15.6.2. *We have $\sigma^2 \geq \frac{\sqrt{3}}{2}$, with equality if and only if τ is the primitive cube or sixth root of unity.*

Proof. This is immediate from the geometry of the fundamental domain. □

15.7. Fundamental domain and Loewner's torus inequality

Thinking of a unit area flat metric g_0 as a lattice quotient $\mathbb{R}^2/L$, we can write down the flat metric as $dx^2 + dy^2$. We think of the conformal factor $f(x, y) > 0$ of the metric g on $\mathbb{T}^2$ as a doubly periodic (i.e., L-periodic) function on $\mathbb{R}^2$.

Theorem 15.7.1. *Let σ be the conformal invariant as above. Then the metric $f^2(dx^2 + dy^2)$ satisfies*

$$\mathrm{area}(g) - \sigma^2 \mathrm{sys}(g)^2 \geq \mathrm{Var}(f). \tag{15.7.1}$$

Proof. Since g_0 is assumed to be of unit area, the basis for its group of deck tranformations can therefore be taken to be the pair

$$\left\{ \frac{\tau}{\sigma}, \frac{1}{\sigma} \right\}.$$

Here $\operatorname{Im}\left(\frac{\tau}{\sigma}\right) = \sigma$. Thus the torus $\mathbb{C}/\operatorname{Span}\left(\frac{\tau}{\sigma}\,,\ \frac{1}{\sigma}\right)$ has unit area:

$$\operatorname{area}\left(\mathbb{C}/\operatorname{Span}\left(\frac{\tau}{\sigma}\,,\ \frac{1}{\sigma}\right)\right) = 1.$$

With these normalisations, we see that the flat torus is ruled by a pencil of horizontal closed geodesics, denoted

$$\gamma_y = \gamma_y(x),$$

each of length σ^{-1}, where the "width" of the pencil equals σ, i.e. the parameter y ranges through the interval $[0,\sigma]$, with $\gamma_\sigma = \gamma_0$. Note that each of these closed geodesics is noncontractible in $\mathbb{T}^2$, and therefore by definition

$$\operatorname{length}(\gamma_y) \geq \operatorname{sys}_1(g).$$

By Fubini's theorem, we pass to the iterated integral (nishneh) to obtain the following lower bound for the expected value:

$$\begin{aligned} E_\mu(f) &= \int_0^\sigma \left(\int_{\gamma_y} f(x)dx \right) dy \\ &= \int_0^\sigma \operatorname{length}(\gamma_y) dy \\ &\geq \sigma \operatorname{sys}(g). \end{aligned}$$

See also [**Ka07**, p. 41, 44]. Substituting into (15.5.1), we obtain the required inequality

$$\operatorname{area}(g) - \sigma^2 \operatorname{sys}(g)^2 \geq \operatorname{Var}(f). \tag{15.7.2}$$

where f is the conformal factor of the metric g with respect to the unit area flat metric g_0. □

Since we have in general $\sigma^2 \geq \frac{\sqrt{3}}{2}$, we obtain in particular Loewner's torus inequality with isosystolic defect,

$$\operatorname{area}(g) - \tfrac{\sqrt{3}}{2} \operatorname{sys}(g)^2 \geq \operatorname{Var}(f). \tag{15.7.3}$$

cf. [**Pu52**].

15.8. Boundary case of equality

COROLLARY 15.8.1. *A metric satisfying the boundary case of equality in Loewner's torus inequality is necessarily flat and homothetic to the quotient of* $\mathbb{R}^2$ *by the lattice of Eisenstein integers.*

PROOF. If a metric $f^2(dx^2 + dy^2)$ satisfies the boundary case of equality $\operatorname{area}(g) - \frac{\sqrt{3}}{2}\operatorname{sys}(g)^2 = 0$, then the variance of the conformal factor f must vanish by (15.7.3). Hence f is a constant function. The proof is completed by applying Lemma 6.3.1 on the Hermite constant in dimension 2. □

Now suppose τ is pure imaginary, *i.e.* the lattice L is a rectangular lattice of coarea 1.

Corollary 15.8.2. *If τ is pure imaginary, then the metric $g = f^2 g_0$ satisfies the inequality*

$$\operatorname{area}(g) - \operatorname{sys}(g)^2 \geq \operatorname{Var}(f). \tag{15.8.1}$$

Proof. If τ is pure imaginary then $\sigma = \sqrt{\operatorname{Im}(\tau)} \geq 1$, and the inequality follows from (15.7.1). □

In particular, every surface of revolution satisfies (15.8.1), since its lattice is rectangular, *cf.* Corollary 14.2.1.

CHAPTER 16

Manifolds and global geometry

16.1. Global geometry of surfaces

Discussion of Local versus Global: The local behavior is by definition the behavior in an open neighborhood of a point. The local behavior of a smooth curve is well understood by the implicit function theorem. Namely, a smooth curve in the plane or in 3-space can be thought of as the graph of a smooth function. A curve in the plane is locally the graph of a scalar function. A curve in 3-space is locally the graph of a vector-valued function.

Example 16.1.1. The unit circle in the plane can be defined implicitly by

$$x^2 + y^2 = 1,$$

or parametrically by

$$t \mapsto (\cos t, \sin t).$$

Alternatively, it can be given locally as the graph of the function

$$f(x) = \sqrt{1 - x^2}.$$

Note that this presentation works only for points on the upper halfcircle. For points on the lower halfcircle we use the function

$$-\sqrt{1 - x^2}.$$

Both of these representations fail at the points $(1, 0)$ and $(-1, 0)$.

To overcome this difficulty, we must work with y as the independent variable, instead of x. Thus, we can parametrize a neighborhood of (1,0) by using the function

$$x = g(y) = \sqrt{1 - y^2}.$$

Example 16.1.2. The helix given in parametric form by

$$(x, y, z) = (\cos t, \sin t, t).$$

It can also be defined as the graph of the vector-valued function $f(z)$, with values in the (x, y)-plane, where

$$(x(z), y(z)) = f(z) = (\cos z, \sin z).$$

In this case the graph representation in fact works even globally.

EXAMPLE 16.1.3. The unit sphere in 3-space can be represented locally as the graph of the function of two variables

$$f(x,y) = \sqrt{1-x^2-y^2}.$$

As above, concerning the local nature of the presentation necessitates additional functions to represent neighborhoods of points not in the open northern hemisphere (this example is discussed in more detail in Section 16.3).

16.2. Definition of manifold

Motivated by the examples given in the previous section, we give a general definition as follows.

A manifold is defined as a subset of Euclidean space which is locally a graph of a function, possibly vector-valued. This is the original definition of Poincaré who invented the notion (see Arnold [**1**, p. 234]).

DEFINITION 16.2.1. By a 2-dimensional closed Riemannian manifold we mean a compact subset

$$\Sigma \subset \mathbb{R}^n$$

such that in an open neighborhood of every point $p \in \Sigma$ in $\mathbb{R}^n$, the compact subset Σ can be represented as the graph of a suitable smooth vector-valued function of two variables.

Here the function has values in $(n-2)$-dimensional vectors.

The usual parametrisation of the graph can then be used to calculate the coefficients g_{ij} (see Definition 11.4.3) of the first fundamental form, as, for example, in Theorem 16.13.2 and Example 7.6.1. The collection of all such data is then denoted by the pair (Σ, g), where

$$g = (g_{ij})$$

is referred to as "the metric".

REMARK 16.2.2. Differential geometers like the (Σ, g) notation, because it helps separate the topology Σ from the geometry g. Strictly speaking, the notation is redundant, since the object g already incorporates all the information, including the topology. However, geometers have found it useful to use g when one wants to emphasize the geometry, and Σ when one wants to emphasize the topology.

Note that, as far as the intrinsic geometry of a Riemannian manifold is concerned, the imbedding in $\mathbb{R}^n$ referred to in Definition 16.2.1 is irrelevant to a certain extent, all the more so since certain basic examples, such as flat tori, are difficult to imbed in a transparent way.

16.3. Sphere as a manifold

The round 2-sphere $S^2 \subset \mathbb{R}^3$ defined by the equation

$$x^2 + y^2 + z^2 = 1$$

is a closed Riemannian manifold. Indeed, consider the function $f(x, y)$ defined in the unit disk $x^2 + y^2 < 1$ by setting $f(x, y) = \sqrt{1 - x^2 - y^2}$. Define a coordinate chart

$$\underline{x}_1(u^1, u^2) = (u^1, u^2, f(u^1, u^2)).$$

Thus, each point of the open northern hemisphere admits a neighborhood diffeomorphic to a ball (and hence to $\mathbb{R}^2$). To cover the southern hemisphere, use the chart

$$\underline{x}_2(u^1, u^2) = (u^1, u^2, -f(u^1, u^2)).$$

To cover the points on the equator, use in addition charts $\underline{x}_3(u^1, u^2) = (u^1, f(u^1, u^2), u^2)$, $\underline{x}_4(u^1, u^2) = (u^1, -f(u^1, u^2), u^2)$, as well as the pair of charts $\underline{x}_5(u^1, u^2) = (f(u^1, u^2), u^1, u^2)$, $\underline{x}_6(u^1, u^2) = (-f(u^1, u^2), u^1, u^2)$.

16.4. Dual bases

Tangent space, cotangent space, and the notation for bases in these spaces were discussed in Section 13.2.

We will work with dual bases $\left(\frac{\partial}{\partial u^i}\right)$ for vectors, and (du^i) for covectors (*i.e.* elements of the dual space), such that

$$du^i \left(\frac{\partial}{\partial u^j} \right) = \delta^i_j, \tag{16.4.1}$$

where δ^i_j is the Kronecker delta.

Recall that the metric coefficients are defined by setting

$$g_{ij} = \langle \underline{x}_i, \underline{x}_j \rangle,$$

where $\underline{x}$ is the parametrisation of the surface. We will only work with metrics whose first fundamental form is diagonal. We can thus write the first fundamental form as follows:

$$g = g_{11}(u^1, u^2)(du^1)^2 + g_{22}(u^1, u^2)(du^2)^2. \tag{16.4.2}$$

Remark 16.4.1. Such data can be computed from a Euclidean imbedding as usual, or it can be given apriori without an imbedding, as we did in the case of the hyperlobic metric.

We will work with such data independently of any Euclidean imbedding, as discussed in Section 11.3. For example, if the metric coefficients form an identity matrix, we obtain

$$g = \left(du^1\right)^2 + \left(du^2\right)^2, \tag{16.4.3}$$

where the interior superscript denotes an index, while exterior superscript denotes the squaring operation.

16.5. Jacobian matrix

The Jacobian matrix of $v = v(u)$ is the matrix

$$\frac{\partial(v^1, v^2)}{\partial(u^1, u^2)},$$

which is the matrix of partial derivatives. Denote by

$$\mathrm{Jac}_v(u) = \det\left(\frac{\partial(v^1, v^2)}{\partial(u^1, u^2)}\right).$$

It is shown in advanced calculus that for any function $f(v)$ in a domain D, one has

$$\int_D f(v)dv^1dv^2 = \int_D g(u)\mathrm{Jac}_v(u)du^1du^2, \tag{16.5.1}$$

where $g(u) = f(v(u))$.

Example 16.5.1. Let $u^1 = r$, $u^2 = \theta$. Let $v^1 = x$, $v^2 = y$. We have $x = r\cos\theta$ and $y = r\sin\theta$. One easily shows that the Jacobian is $\mathrm{Jac}_v(u) = r$. The area elements are related by

$$dv^1dv^2 = \mathrm{Jac}_v(u)du^1du^2,$$

or

$$dxdy = rdrd\theta.$$

A similar relation holds for integrals.

16.6. Area of a surface, independence of partition

Partition[1] is what allows us to perform actual calculuations with area, but the result is independent of partition (see below).

Based on the local definition of area discussed in an earlier chapter, we will now deal with the corresponding global invariant.

Definition 16.6.1. The *area element* dA of the surface is the element

$$dA := \sqrt{\det(g_{ij})}du^1du^2,$$

where $\det(g_{ij}) = g_{11}g_{22} - g_{12}^2$ as usual.

[1] ritzuf or chaluka?

THEOREM 16.6.2. *Define the* area *of* (Σ, g) *by means of the formula*

$$\text{area} = \int_\Sigma dA = \sum_{\{U\}} \int_U \sqrt{\det(g_{ij})} du^1 du^2, \tag{16.6.1}$$

namely by choosing a partition $\{U\}$ *of* Σ *subordinate to a finite open cover as in Definition 16.2.1, performing a separate integration in each open set, and summing the resulting areas. Then the total area is independent of the partition and choice of coordinates.*

PROOF. Consider a change from a coordinate chart (u^i) to another coordinate chart, denoted (v^α). In the overlap of the two domains, the coordinates can be expressed in terms of each other, *e.g.* $v = v(u)$, and we have the 2 by 2 Jacobian matrix $\text{Jac}_v(u)$.

Denote by $\tilde{g}_{\alpha\beta}$ the metric coefficients with respect to the chart (v^α). Thus, in the case of a metric induced by a Euclidean imbedding defined by $x = x(u) = x(u^1, u^2)$, we obtain a new parametrisation

$$y(v) = x(u(v)).$$

Then we have

$$\begin{aligned}
\tilde{g}_{\alpha\beta} &= \left\langle \frac{\partial y}{\partial v^\alpha}, \frac{\partial y}{\partial v^\beta} \right\rangle \\
&= \left\langle \frac{\partial x}{\partial u^i}\frac{\partial u^i}{\partial v^\alpha}, \frac{\partial x}{\partial u^j}\frac{\partial u^j}{\partial v^\beta} \right\rangle \\
&= \frac{\partial u^i}{\partial v^\alpha}\frac{\partial u^j}{\partial v^\beta} \left\langle \frac{\partial x}{\partial u^i}, \frac{\partial x}{\partial u^j} \right\rangle \\
&= g_{ij} \frac{\partial u^i}{\partial v^\alpha}\frac{\partial u^j}{\partial v^\beta}.
\end{aligned}$$

The right hand side is a product of *three* square matrices:

$$\frac{\partial u^i}{\partial v^\alpha} \, g_{ij} \, \frac{\partial u^j}{\partial v^\beta}.$$

The matrices on the left and on the right are both Jacobian matrices. Since determinant is multiplicative, we obtain

$$\det(\tilde{g}_{\alpha\beta}) = \det(g_{ij}) \det\left(\frac{\partial(u^1, u^2)}{\partial(v^1, v^2)}\right)^2.$$

Hence using equation (16.5.1), we can write the area element as

$$\begin{aligned} dA &= \det{}^{\frac{1}{2}}(\tilde{g}_{\alpha\beta}) dv^1 dv^2 \\ &= \det{}^{\frac{1}{2}}(\tilde{g}_{\alpha\beta}) \operatorname{Jac}_v(u) du^1 du^2 \\ &= \det{}^{\frac{1}{2}}(g_{ij}) \det\left(\frac{\partial(u^1, u^2)}{\partial(v^1, v^2)}\right) \operatorname{Jac}_v(u) du^1 du^2 \\ &= \det{}^{\frac{1}{2}}(g_{ij})\, du^1 du^2 \end{aligned}$$

since inverse maps have reciprocal Jacobians by chain rule. Thus the integrand is unchanged and the area element is well defined. □

16.7. Conformal equivalence

Definition 16.7.1. Two metrics, $g = g_{ij} du^i du^j$ and $h = h_{ij} du^i du^j$, on Σ are called *conformally equivalent*, or *conformal* for short, if there exists a function $f = f(u^1, u^2) > 0$ such that

$$g = f^2 h,$$

in other words,

$$g_{ij} = f^2 h_{ij} \quad \forall i, j. \tag{16.7.1}$$

Definition 16.7.2. The function f above is called the *conformal factor* (note that sometimes it is more convenient to refer, instead, to the function $\lambda = f^2$ as the conformal factor).

Theorem 16.7.3. *Note that the length of every vector at a given point (u^1, u^2) is multiplied precisely by $f(u^1, u^2)$.*

Proof. More speficially, a vector $v = v^i \frac{\partial}{\partial u^i}$ which is a unit vector for the metric h, is "stretched" by a factor of f, *i.e.* its length with respect to g equals f. Indeed, the new length of v is

$$\begin{aligned} \sqrt{g(v, v)} &= g\left(v^i \frac{\partial}{\partial u^i}, v^j \frac{\partial}{\partial u^j}\right)^{\frac{1}{2}} \\ &= \left(g\left(\frac{\partial}{\partial u^i}, \frac{\partial}{\partial u^j}\right) v^i v^j\right)^{\frac{1}{2}} \\ &= \sqrt{g_{ij} v^i v^j} \\ &= \sqrt{f^2 h_{ij} v^i v^j} \\ &= f\sqrt{h_{ij} v^i v^j} \\ &= f, \end{aligned}$$

proving the theorem. □

Definition 16.7.4. An equivalence class of metrics on Σ conformal to each other is called a *conformal structure* on Σ (mivneh conformi).

16.8. Geodesic equation

The material in this section has already been dealt with in an earlier chapter.

Perhaps the simplest possible definition of a geodesic β on a surface in 3-space is in terms of the orthogonality of its second derivative β'' to the surface. The nonlinear second order ordinary differential equation defining a geodesic is, of course, the "true" if complicated definition. We will now prove the equivalence of the two definitions. Consider a plane curve

$$\underset{s}{\mathbb{R}} \underset{\alpha}{\longrightarrow} \underset{(u^1,u^2)}{\mathbb{R}^2}$$

where $\alpha = (\alpha^1(s), \alpha^2(s))$. Let $x : \mathbb{R}^2 \to \mathbb{R}^3$ be a regular parametrisation of a surface in 3-space. Then the composition

$$\underset{s}{\mathbb{R}} \underset{\alpha}{\longrightarrow} \underset{(u^1,u^2)}{\mathbb{R}^2} \underset{x}{\longrightarrow} \mathbb{R}^3$$

yields a curve

$$\beta = x \circ \alpha.$$

Definition 16.8.1. A curve $\beta = x \circ \alpha$ is a geodesic on the surface x if one of the following two equivalent conditions is satisfied:

(a) we have for each $k = 1, 2$,

$$(\alpha^k)'' + \Gamma^k_{ij}(\alpha^i)'(\alpha^j)' = 0 \quad \text{where} \quad ' = \frac{d}{ds}\ , \tag{16.8.1}$$

meaning that

$$(\forall k) \quad \frac{d^2\alpha^k}{ds^2} + \Gamma^k_{ij}\frac{d\alpha^i}{ds}\frac{d\alpha^j}{ds} = 0;$$

(b) the vector β'' is perpendicular to the surface and one has

$$\beta'' = L_{ij}\alpha^{i'}\alpha^{j'}n. \tag{16.8.2}$$

To prove the equivalence, we write $\beta = x \circ \alpha$, then $\beta' = x_i\alpha^{i'}$ by chain rule. Furthermore,

$$\beta'' = \frac{d}{ds}(x_i \circ \alpha)\alpha^{i'} + x_i\alpha^{i''} = x_{ij}\alpha^{j'}\alpha^{i'} + x_k\alpha^{k''}.$$

Since $x_{ij} = \Gamma^k_{ij}x_k + L_{ij}n$ holds, we have

$$\beta'' - L_{ij}\alpha^{i'}\alpha^{j'}n = x_k\left(\alpha^{k''} + \Gamma^k_{ij}\alpha^{i'}\alpha^{j'}\right).$$

16.9. Closed geodesic

DEFINITION 16.9.1. A *closed geodesic* in a Riemannian 2-manifold Σ is defined equivalently as

(1) a periodic curve $\beta : \mathbb{R} \to \Sigma$ satisfying the geodesic equation $\alpha^{k''} + \Gamma^k_{ij}\alpha^{i'}\alpha^{j'} = 0$ in every chart $\underline{x} : \mathbb{R}^2 \to \Sigma$, where, as usual, $\beta = \underline{x} \circ \alpha$ and $\alpha(s) = (\alpha^1(s), \alpha^2(s))$ where s is arclength. Namely, there exists a period $T > 0$ such that $\beta(s+T) = \beta(s)$ for all s.

(2) A unit speed map from a circle $\mathbb{R}/L_T \to \Sigma$ satisfying the geodesic equation at each point, where $L_T = T\mathbb{Z} \subset \mathbb{R}$ is the rank one lattice generated by $T > 0$.

DEFINITION 16.9.2. The *length* $L(\beta)$ of a path $\beta : [a, b] \to \Sigma$ is calculated using the formula

$$L(\beta) = \int_a^b \|\beta'(t)\| dt,$$

where $\|v\| = \sqrt{g_{ij}v^i v^j}$ whenever $v = v^i \underline{x}_i$. The energy is defined by $E(\beta) = \int_a^b \|\beta'(t)\|^2 dt$.

A closed geodesic as in Definition 16.9.1, item 2 has length T.

REMARK 16.9.3. The geodesic equation (16.8.1) is the Euler-Lagrange equation of the first variation of arc length. Therefore when a path minimizes arc length among all neighboring paths connecting two fixed points, it must be a geodesic. A corresponding statement is valid for closed loops, *cf.* proof of Theorem 16.10.1. See also Section 15.1. In Sections 16.11 and 16.15 we will give a complete description of the geodesics for the constant curvature sphere, as well as for flat tori.

16.10. Existence of closed geodesic

THEOREM 16.10.1. *Every free homotopy class of loops in a closed manifold contains a closed geodesic.*

PROOF. We sketch a proof for the benefit of a curious reader, who can also check that the construction is independent of the choices involved. The relevant topological notions are defined in Section 16.18 and [**Hat02**]. A free homotopy class α of a manifold M corresponds to a conjugacy class $g_\alpha \subset \pi_1(M)$. Pick an element $g \in g_\alpha$. Thus g acts on the universal cover $\tilde{M}$ of M. Let $f_g : M \to \mathbb{R}$ be the displacement function of g, *i.e.*

$$f_g(x) = d(\tilde{x}, g.\tilde{x}).$$

Let $x_0 \in M$ be a minimum of f_g. A first variation argument shows that any length-minimizing path between $\tilde{x}_0$ and $g.\tilde{x}_0$ descends to a closed geodesic in M representing α, *cf.* [**Car92, Ch93, GaHL04**]. □

16.11. Surfaces of constant curvature

By the uniformisation theorem 13.6.1, all surfaces fall into three types, according to whether they are conformally equivalent to metrics that are:

(1) flat (*i.e.* have zero Gaussian curvature $K \equiv 0$);
(2) spherical ($K \equiv +1$);
(3) hyperbolic ($K \equiv -1$).

For closed surfaces, the sign of the Gaussian curvature K is that of its Euler characteristic, *cf.* formula (16.26.1).

Theorem 16.11.1 (Constant positive curvature). *There are only two compact surfaces, up to isometry, of constant Gaussian curvature $K = +1$. They are the round sphere S^2 of Example 16.3; and the real projective plane, denoted $\mathbb{RP}^2$.*

16.12. Real projective plane

Intuitively, one thinks of the real projective plane as the quotient surface obtained if one starts with the northern hemisphere of the 2-sphere, and "glues" together pairs of opposite points of the equatorial circle (the boundary of the hemisphere).

More formally, the real projective plane can be defined as follows. Let

$$m : S^2 \to S^2$$

denote the antipodal map of the sphere, *i.e.* the restriction of the map

$$v \mapsto -v$$

in $\mathbb{R}^3$. Then m is an involution. In other words, if we consider the action of the group $\mathbb{Z}_2 = \{e, m\}$ on the sphere, each orbit of the $\mathbb{Z}_2$ action on S^2 consists of a pair of antipodal points

$$\{\pm p\} \subset S^2. \tag{16.12.1}$$

Definition 16.12.1. On the set-theoretic level, the real projective plane $\mathbb{RP}^2$ is the set of orbits of type (16.12.1), *i.e.* the quotient of S^2 by the $\mathbb{Z}_2$ action.

Denote by $Q : S^2 \to \mathbb{RP}^2$ the quotient map. The smooth structure and metric on $\mathbb{RP}^2$ are induced from S^2 in the following sense. Let $\underline{x}$:

$\mathbb{R}^2 \to S^2$ be a chart on S^2 not containing any pair of antipodal points. Let g_{ij} be the metric coefficients with respect to this chart.

Let $\underline{y} = m \circ \underline{x} = -\underline{x}$ denote the "opposite" chart, and denote by h_{ij} its metric coefficients. Then

$$h_{ij} = \langle \frac{\partial \underline{y}}{\partial u^i}, \frac{\partial \underline{y}}{\partial u^j} \rangle = \langle -\frac{\partial \underline{x}}{\partial u^i}, -\frac{\partial \underline{x}}{\partial u^j} \rangle = \langle \frac{\partial \underline{x}}{\partial u^i}, \frac{\partial \underline{x}}{\partial u^j} \rangle = g_{ij}. \tag{16.12.2}$$

Thus the opposite chart defines the identical metric coefficients. The composition $Q \circ \underline{x}$ is a chart on $\mathbb{RP}^2$, and the same functions g_{ij} form the metric coefficients for $\mathbb{RP}^2$ relative to this chart.

We can summarize the preceeding discussion by means of the following definition.

Definition 16.12.2. The real projective plane $\mathbb{RP}^2$ is defined in the following two equivalent ways:

(1) the quotient of the round sphere S^2 by (the restriction to S^2 of) the antipodal map $v \mapsto -v$ in $\mathbb{R}^3$. In other words, a typical point of $\mathbb{RP}^2$ can be thought of as a pair of opposite points of the round sphere.
(2) the northern hemisphere of S^2, with opposite points of the equator identified.

The smooth structure of $\mathbb{RP}^2$ is induced from the round sphere. Since the antipodal map preserves the metric coefficients by the calculation (16.12.2), the metric structure of constant Gaussian curvature $K = +1$ descends to $\mathbb{RP}^2$, as well.

16.13. Simple loops for surfaces of positive curvature

Definition 16.13.1. A loop $\alpha : S^1 \to X$ of a space X is called *simple* if the map α is one-to-one, *cf.* Definition 16.18.1.

Theorem 16.13.2. *The basic properties of the geodesics on surfaces of constant positive curvature as as follows:*

(1) *all geodesics are closed;*
(2) *the simple closed geodesics on S^2 have length 2π and are defined by the great circles;*
(3) *the simple closed geodesics on $\mathbb{RP}^2$ have length π;*
(4) *the simple closed geodesics of $\mathbb{RP}^2$ are parametrized by half-great circles on the sphere.*

Proof. We calculate the length of the equator of S^2. Here the sphere $\rho = 1$ is parametrized by

$$x(\theta, \varphi) = (\sin\varphi\cos\theta, \sin\varphi\sin\theta, \cos\varphi)$$

in spherical coordinates (θ, φ). The equator is the curve $x \circ \alpha$ where $\alpha(s) = (s, \frac{\pi}{2})$ with $s \in [0, 2\pi]$. Thus $\alpha^1(s) = \theta(s) = s$. Recall that the metric coefficients are given by $g_{11}(\theta, \varphi) = \sin^2 \varphi$, while $g_{22} = 1$ and $g_{12} = 0$. Thus

$$\|\beta'(s)\| = \sqrt{g_{ij}\left(s, \tfrac{\pi}{2}\right) \alpha^{i\prime} \alpha^{j\prime}} = \sqrt{\left(\sin \tfrac{\pi}{2}\right) \left(\tfrac{d\theta}{ds}\right)^2} = 1.$$

Thus the length of β is

$$\int_0^{2\pi} \|\beta'(s)\| ds = \int_0^{2\pi} 1 ds = 2\pi.$$

A geodesic on $\mathbb{RP}^2$ is twice as short as on S^2, since the antipodal points are identified, and therefore the geodesic "closes up" sooner than (*i.e.* twice as fast as) on the sphere. For example, a longitude of S^2 is not a closed curve, but it descends to a closed curve on $\mathbb{RP}^2$, since its endpoints (north and south poles) are antipodal, and are therefore identified with each other. □

16.14. Successive minima

The material in this section has already been dealt with in an earlier chapter.

Let B be a finite-dimensional Banach space, *i.e.* a vector space together with a norm $\| \ \|$. Let $L \subset (B, \| \ \|)$ be a lattice of maximal rank, *i.e.* satisfying $\operatorname{rank}(L) = \dim(B)$. We define the notion of successive minima of L as follows, *cf.* [**GruL87**, p. 58].

Definition 16.14.1. For each $k = 1, 2, \ldots, \operatorname{rank}(L)$, define the *k-th successive minimum* λ_k of the lattice L by

$$\lambda_k(L, \| \ \|) = \inf \left\{ \lambda \in \mathbb{R} \;\middle|\; \begin{array}{c} \exists \text{ lin. indep. } v_1, \ldots, v_k \in L \\ \text{with } \|v_i\| \leq \lambda \text{ for all } i \end{array} \right\}. \quad (16.14.1)$$

Thus the first successive minimum, $\lambda_1(L, \| \ \|)$ is the least length of a nonzero vector in L.

16.15. Flat surfaces

A metric is called *flat* if its Gaussian curvature K vanishes at every point.

Theorem 16.15.1. *A closed surface of constant Gaussian curvature $K = 0$ is topologically either a torus $\mathbb{T}^2$ or a Klein bottle.*

Let us give a precise description in the former case.

EXAMPLE 16.15.2 (Flat tori). Every flat torus is isometric to a quotient $\mathbb{T}^2 = \mathbb{R}^2/L$ where L is a lattice, *cf.* [**Lo71**, Theorem 38.2]. In other words, a point of the torus is a coset of the additive action of the lattice in $\mathbb{R}^2$. The smooth structure is inherited from $\mathbb{R}^2$. Meanwhile, the additive action of the lattice is isometric. Indeed, we have $\operatorname{dist}(p,q) = \|q-p\|$, while for any $\ell \in L$, we have

$$\operatorname{dist}(p+\ell, q+\ell) = \|q+\ell-(p+\ell)\| = \|q-p\| = \operatorname{dist}(p,q).$$

Therefore the flat metric on $\mathbb{R}^2$ descends to $\mathbb{T}^2$.

Note that locally, all flat tori are indistinguishable from the flat plane itself. However, their global geometry depends on the metric invariants of the lattice, *e.g.* its successive minima, *cf.* Definition 16.14.1. Thus, we have the following.

THEOREM 16.15.3. *The least length of a nontrivial closed geodesic on a flat torus $\mathbb{T}^2 = \mathbb{R}^2/L$ equals the first successive minimum $\lambda_1(L)$.*

PROOF. The geodesics on the torus are the projections of straight lines in $\mathbb{R}^2$. In order for a straight line to close up, it must pass through a pair of points x and $x+\ell$ where $\ell \in L$. The length of the corresponding closed geodesic on $\mathbb{T}^2$ is precisely $\|\ell\|$, where $\|\ \|$ is the Euclidean norm. By choosing a shortest element in the lattice, we obtain a shortest closed geodesic on the corresponding torus. □

16.16. Hyperbolic surfaces

Most closed surfaces admit neither flat metrics nor metrics of positive curvature, but rather hyperbolic metrics. A hyperbolic surface is a surface equipped with a metric of constant Gaussian curvature $K = -1$. This case is far richer than the other two.

EXAMPLE 16.16.1. The pseudosphere (so called because its Gaussian curvature is constant, and equals -1) is the surface of revolution

$$(f(\varphi)\cos\theta, f(\varphi)\sin\theta, g(\varphi))$$

in $\mathbb{R}^3$ defined by the functions $f(\phi) = e^{\phi}$ and

$$g(\phi) = \int_0^{\phi} \left(1-e^{2\psi}\right)^{1/2} d\psi,$$

where ϕ ranges through the interval $-\infty < \phi \leq 0$. The usual formulas $g_{11} = f^2$ as well as $g_{22} = \left(\frac{df}{d\phi}\right)^2 + \left(\frac{dg}{d\phi}\right)^2$ yield in our case $g_{11} = e^{2\phi}$, while

$$\begin{aligned} g_{22} &= \left(e^{\phi}\right)^2 + \left(\sqrt{1-e^{2\phi}}\right)^2 \\ &= e^{2\phi} + 1 - e^{2\phi} \\ &= 1. \end{aligned}$$

Thus $(g_{ij}) = \begin{pmatrix} e^{2\phi} & 0 \\ 0 & 1 \end{pmatrix}$. The pseudosphere has constant Gaussian curvature -1, but it is not a closed surface (as it is unbounded in $\mathbb{R}^3$).

16.17. Hyperbolic plane

The metric

$$g_{\mathcal{H}^2} = \frac{1}{y^2}(dx^2 + dy^2) \qquad (16.17.1)$$

in the upperhalf plane

$$\mathcal{H}^2 = \{(x, y) \mid y > 0\}$$

is called the hyperbolic metric of the upper half plane.

THEOREM 16.17.1. *The metric* (16.17.1) *has constant Gaussian curvature* $K = -1$.

PROOF. By Theorem 11.7.3, we have

$$K = -\Delta_{LB} \log f = \Delta_{LB} \log y = y^2 \left(-\frac{1}{y^2}\right) = -1,$$

as required. □

In coordinates (u^1, u^2), we can write it, a bit awkwardly, as

$$g_{\mathcal{H}^2} = \frac{1}{(u^2)^2}\left(\left(du^1\right)^2 + \left(du^2\right)^2\right).$$

The Riemannian manifold $(\mathcal{H}^2, g_{\mathcal{H}^2})$ is referred to as the *Poincaré upperhalf plane*. Its significance resides in the following theorem.

THEOREM 16.17.2. *Every closed hyperbolic surface* Σ *is isometric to the quotient of the Poincaré upperhalf plane by the action of a suitable group* Γ*:*

$$\Sigma = \mathcal{H}^2/\Gamma.$$

Here the nonabelian group Γ is a discrete subgroup $\Gamma \subset PSL(2, \mathbb{R})$, where a matrix $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ acts on $\mathcal{H}^2 = \{z \in \mathbb{C} | \Im(z) > 0\}$ by

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

called fractional linear transformations, of Mobius transformations. All such transformations are isometries of the hyperbolic metric. The following theorem is proved, for example, in [**Kato92**].

THEOREM 16.17.3. *Every geodesic in the Poincaré upperhalf plane is either a vertical ray, or a semicircle perpendicular to the x-axis.*

The foundational significance of this model in the context of the parallel postulate of Euclid has been discussed by numerous authors.

EXAMPLE 16.17.4. The length of a vertical interval joining i to ci can be calculated as follows. Recall that the conformal factor is $f(x, y) = \frac{1}{y}$. The length is therefore given by

$$\left| \int_1^c \frac{1}{y} dy \right| = |\log c|.$$

Here the substitution $y = e^s$ gives an arclength parametrisation.

16.18. Loops, simply connected spaces

We would like to provide a self-contained explanation of the topological ingredient which is necessary so as to understand Loewner's torus inequality, *i.e.* essentially the notion of a noncontractible loop and the fundamental group of a topological space X. See [**Hat02**, Chapter 1] for a more detailed account.

DEFINITION 16.18.1. A *loop* in X can be defined in one of two equivalent ways:

(1) a continuous map $\beta : [a, b] \to X$ satisfying $\beta(a) = \beta(b)$;
(2) a continuous map $\lambda : S^1 \to X$ from the circle S^1 to X.

LEMMA 16.18.2. *The two definitions of a loop are equivalent.*

PROOF. Consider the unique increasing linear function

$$f : [a, b] \to [0, 2\pi]$$

which is one-to-one and onto. Thus, $f(t) = \frac{2\pi(t-a)}{b-a}$. Given a map

$$\lambda(e^{is}) : S^1 \to X,$$

we associate to it a map $\beta(t) = \lambda\left(e^{if(t)}\right)$, and vice versa. □

DEFINITION 16.18.3. A loop $S^1 \to X$ is said to be *contractible* if the map of the circle can be extended to a continuous map of the unit disk $\mathbb{D} \to X$, where $S^1 = \partial\mathbb{D}$.

DEFINITION 16.18.4. A space X is called *simply connected* if every loop in X is contractible.

THEOREM 16.18.5. *The sphere $S^n \subset \mathbb{R}^{n+1}$ which is the solution set of $x_0^2 + \ldots + x_n^2 = 1$ is simply connected for $n \geq 2$. The circle S^1 is not simply connected.*

16.19. Orientation on loops and surfaces

Let $S^1 \subset \mathbb{C}$ be the unit circle. The choice of an orientation on the circle is an arrow pointing clockwise or counterclockwise. The standard choice is to consider S^1 as an oriented manifold with orientation chosen counterclockwise.

If a surface is imbedded in 3-space, one can choose a continuous unit normal vector n at every point. Then an orientation is defined by the right hand rule with respect to n thought of as the thumb (agudal).

16.20. Cycles and boundaries

The singular homology groups with integer coefficients, $H_k(M;\mathbb{Z})$ for $k = 0, 1, \ldots$ of M are abelian groups which are homotopy invariants of M. Developing the singular homology theory is time-consuming. The case that we will be primarily interested in as far as these notes are concerned, is that of the 1-dimensional homology group:

$$H_1(M;\mathbb{Z}).$$

In this case, the homology groups can be characterized easily without the general machinery of singular simplices.

Let $S^1 \subset \mathbb{C}$ be the unit circle, which we think of as a 1-dimensional manifold with an orientation given by the counterclockwise direction.

DEFINITION 16.20.1. A 1-cycle α on a manifold M is an integer linear combination

$$\alpha = \sum_i n_i f_i$$

where $n_i \in \mathbb{Z}$ is called the *multiplicity* (ribui), while each

$$f_i : S^1 \to M$$

is a loop given by a smooth map from the circle to M, and each loop is endowed with the orientation coming from S^1.

DEFINITION 16.20.2. The space of 1-cycles on M is denoted

$$Z_1(M;\mathbb{Z}).$$

Let $(\Sigma_g, \partial\Sigma_g)$ be a surface with boundary $\partial\Sigma_g$, where the genus g is irrelevant for the moment and is only added so as to avoid confusion with the summation symbol $\sum$.

The boundary $\partial\Sigma_g$ is a disjoint union of circles. Now assume the surface Σ_g is oriented.

PROPOSITION 16.20.3. *The orientation of the surface induces an orientation on each boundary circle.*

Thus we obtain an orientation-preserving identification of each boundary component with the standard unit circle $S^1 \subset \mathbb{C}$ (with its counterclockwise orientation).

Given a map $\Sigma_g \to M$, its restriction to the boundary therefore produces a 1-cycle

$$\partial\Sigma_g \in Z_1(M;\mathbb{Z}).$$

DEFINITION 16.20.4. The space

$$B_1(M;\mathbb{Z}) \subset Z_1(M;\mathbb{Z})$$

of 1-boundaries in M is the space of all cycles

$$\sum_i n_i f_i \in Z_1(M;\mathbb{Z})$$

such that there exists a map of an oriented surface $\Sigma_g \to M$ (for some g) satisfying

$$\partial\Sigma_g = \sum_i n_i f_i.$$

EXAMPLE 16.20.5. Consider the cylinder

$$x^2 + y^2 = 1, \quad 0 \le z \le 1$$

of unit height. The two boundary components correspond to the two circles: the "bottom" circle C_{bottom} defined by $z = 0$, and the "top" circle C_{top} defined by and $z = 1$. Consider the orientation on the cylinder defined by the outward pointing normal vector. It induces the counterclockwise orientation on C_{bottom}, and a clockwise orientation on C_{top}.

Now let C_0 and C_1 be the same circles with the following choice of orientation: we choose a standard counterclockwise parametrisation on both circles, i.e., parametrize them by means of $(\cos\theta, \sin\theta)$. Then the boundary of the cylinder is the *difference* of the two circles: $C_0 - C_1$, or $C_1 - C_0$, depending on the choice of orientation.

EXAMPLE 16.20.6. Cutting up a circle of genus 2 into two onceholed tori shows that the separating curve is a 1-boundary.

THEOREM 16.20.7. *On a closed orientable surface, a separating curve is a boundary, while a non-separating loop is never a boundary.*

16.21. First singular homology group

DEFINITION 16.21.1. The 1-dimensional homology group of M with integer coefficients is the quotient group

$$H_1(M;\mathbb{Z}) = Z_1(M;\mathbb{Z})/B_1(M;\mathbb{Z}).$$

DEFINITION 16.21.2. Given a cycle $C \in Z_1(M;\mathbb{Z})$, its homology class will be denoted $[C] \in H_1(M;\mathbb{Z})$.

EXAMPLE 16.21.3. A non-separating loop on a closed surface represents a non-trivial homology class of the surface.

THEOREM 16.21.4. *The 1-dimensional homology group $H_1(M;\mathbb{Z})$ is the abelianisation of the fundamental group $\pi_1(M)$:*

$$H_1(M;\mathbb{Z}) = (\pi_1 M)^{ab}.$$

Note that a significant difference between the fundamental group and the first homology group is the following. While only based loops participate in the definition of the fundamental group, the definition of $H_1(M;\mathbb{Z})$ involves free (not based) loops.

EXAMPLE 16.21.5. The fundamental groups of the real projective plane $\mathbb{RP}^2$ and the 2-torus $\mathbb{T}^2$ are already abelian. Therefore one obtains

$$H_1(\mathbb{RP}^2;\mathbb{Z}) = \mathbb{Z}/2\mathbb{Z},$$

and

$$H_1(T^2;\mathbb{Z}) = \mathbb{Z}^2.$$

EXAMPLE 16.21.6. The fundamental group of an orientable closed surface Σ_g of genus g is known to be a group on $2g$ generators with a single relation which is a product of g commutators. Therefore one has

$$H_1(\Sigma_g;\mathbb{Z}) = \mathbb{Z}^{2g}.$$

16.22. Stable norm in 1-dimensional homology

Assume the manifold M has a Riemannian metric. Given a smooth loop $f : S^1 \to M$, we can measure its volume (length) with respect to the metric of M. We will denote this length by

$$\mathrm{vol}(f)$$

with a view to higher-dimensional generalisation.

DEFINITION 16.22.1. The volume (length) of a 1-cycle $C = \sum_i n_i f_i$ is defined as

$$\mathrm{vol}(C) = \sum_i |n_i| \,\mathrm{vol}(f_i).$$

Definition 16.22.2. Let $\alpha \in H_1(M;\mathbb{Z})$ be a 1-dimensional homology class. We define the volume of α as the infimum of volumes of representative 1-cycles:

$$\mathrm{vol}(\alpha) = \inf \{\mathrm{vol}(C) \mid C \in \alpha\},$$

where the infimum is over all cycles $C = \sum_i n_i f_i$ representing the class $\alpha \in H_1(M;\mathbb{Z})$.

The following phenomenon occurs for orientable surfaces.

Theorem 16.22.3. *Let M be an orientable surface, i.e. 2-dimensional manifold. Let $\alpha \in H_1(M;\mathbb{Z})$. For all $j \in \mathbb{N}$, we have*

$$\mathrm{vol}(j\alpha) = j\,\mathrm{vol}(\alpha), \tag{16.22.1}$$

where $j\alpha$ denotes the class $\alpha + \alpha + \ldots + \alpha$, with j summands.

Proof. To fix ideas, let $j = 2$. By Lemma 16.22.4 below, a minimizing loop C representing a multiple class 2α will necessarily intersect itself in a suitable point p. Then the 1-cycle represented by C can be decomposed into the sum of two 1-cycles (where each can be thought of as a loop based at p). The shorter of the two will then give a minimizing loop in the class α which proves the identity (16.22.1) in this case. The general case follows similarly. □

Lemma 16.22.4. *A loop going around a cylinder twice necessarily has a point of self-intersection.*

Proof. We think of the loop as the graph of a 4π-periodic function $f(t)$ (or alternatively a function on $[0, 4\pi]$ with equal values at the endpoints). Consider the difference $g(t) = f(t) - f(t + 2\pi)$. Then g takes both positive and negative values. By the intermediate function theorem, the function g must have a zero t_0. Then $f(t_0) = f(t_0 + 2\pi)$ hence t_0 is a point of self-intersection of the loop. □

16.23. The degree of a map

An example of a degree d map is most easily produced in the case of a circle. A self-map of a circle given by

$$e^{i\theta} \mapsto e^{id\theta}$$

has degree d.

We will discuss the degree in the context of surfaces only.

Definition 16.23.1. The degree

$$d_f$$

of a map f between closed surfaces is the algebraic number of points in the inverse image of a generic point of the target surface.

We can use the 2-dimensional homology groups defined elsewhere in these notes, so as to calculate the degree as follows. Recall that

$$H_2(\Sigma;\mathbb{Z}) = \mathbb{Z},$$

where the generator is represented by the identity self-map of the surface.

THEOREM 16.23.2. *A map*

$$f : \Sigma_1 \to \Sigma_2$$

induces a homomorphism

$$f_* : H_2(\Sigma_1;\mathbb{Z}) \to H_2(\Sigma_2;\mathbb{Z}),$$

corresponding to multiplication by the degree d_f *once the groups are identified with* $\mathbb{Z}$*:*

$$\mathbb{Z} \to \mathbb{Z}, \quad n \mapsto d_f\, n.$$

16.24. Degree of normal map of an imbedded surface

THEOREM 16.24.1. *Let* $\Sigma \subset \mathbb{R}^3$ *be an imbedded surface. Let* p *be its genus. Consider the normal map*

$$f_n : \Sigma \to S^2$$

defined by sending each point $x \in \Sigma$ *to the normal vector* $n = n_x$ *at* x*. Then the degree of the normal map is precisely* $1 - p$*.*

EXAMPLE 16.24.2. For the unit sphere, the normal map is the identity map. The genus is 0, while the degree of the normal map is $1 - p = 1$.

EXAMPLE 16.24.3. For the torus, the normal map is harder to visualize. The genus is 1, while the degree of the normal map is 0.

EXAMPLE 16.24.4. For a genus 2 surface imbedded in $\mathbb{R}^3$, the degree of the normal map is $1 - 2 = -1$. This means that if the surface is oriented by the outward-pointing normal vector, the normal map is orientation-reversing.

16.25. Euler characteristic of an orientable surface

The Euler characteristic $\chi(\Sigma)$ is even for closed orientable surfaces, and the integer $p = p(\Sigma) \geq 0$ defined by

$$\chi(\Sigma) = 2 - 2p$$

is called the *genus* of Σ.

EXAMPLE 16.25.1. We have $p(S^2) = 0$, while $p(\mathbb{T}^2) = 1$.

In general, the genus can be understood intuitively as the number of "holes", *i.e.* "handles", in a familiar 3-dimensional picture of a pretzel. We see from formula (16.26.1) that the only compact orientable surface admitting flat metrics is the 2-torus. See [**Ar83**] for a friendly topological introduction to surfaces, and [**Hat02**] for a general definition of the Euler characteristic.

16.26. Gauss-Bonnet theorem

Every imbedded closed surface in 3-space admits a continuous choice of a unit normal vector $n = n_x$ at every point x. Note that no such choice is possible for an imbedding of the Mobius band, see [**Ar83**] for more details on orientability and imbeddings.

Closed imbedded surfaces in $\mathbb{R}^3$ are called *orientable.*

REMARK 16.26.1. The integrals of type

$$\int_\Sigma$$

will be understood in the sense of Theorem 16.6.2, namely using an implied partition subordinate to an atlas, and calculating the integral using coordinates (u^1, u^2) in each chart, so that we can express the metric in terms of metric coefficients

$$g_{ij} = g_{ij}(u^1, u^2)$$

and similarly the Gaussian curvature

$$K = K(u^1, u^2).$$

THEOREM 16.26.2 (Gauss-Bonnet theorem). *Every closed surface* Σ *satisfies the identity*

$$\int_\Sigma K(u^1, u^2)\sqrt{\det(g_{ij})}du^1du^2 = 2\pi\chi(\Sigma), \qquad (16.26.1)$$

where K *is the Gaussian curvature function on* Σ*, whereas* $\chi(\Sigma)$ *is its Euler characteristic.*

16.27. Change of metric exploiting Gaussian curvature

We will use the term pseudometric for a quadratic form (or the associated bilinear form), possibly degenerate.

We would like to give an indication of a proof of the Gauss-Bonnet theorem. We will have to avoid discussing some technical points. Consider the normal map

$$F : \Sigma \to S^2, \quad x \mapsto n_x.$$

Consider a neigborhood in Σ where the map F is a homeomorphism (this is not always possible, and is one of the technical points we are avoiding).

Definition 16.27.1. Let g_Σ the metric of Σ, and h the standard metric of S^2.

Given a point $x \in \Sigma$ in such a neighborhood, we can calculate the curvature $K(x)$. We can then consider a new metric in the conformal class of the metric g_Σ, defined as follows.

Definition 16.27.2. We define a new pseudometric, denoted $\hat{g}_\Sigma$, on Σ by multiplying by the conformal factor $K(x)$ at the point x. Namely, $\hat{g}_\Sigma$ is the pseudometric which at the point x is given by the quadratic form

$$\hat{g}_x = K(x) g_x.$$

If $K \geq 0$ then the length of vectors is multiplied by $\sqrt{K}$.

The key to understanding the proof of the Gauss-Bonnet theorem in the case of imbedded surfaces is the following theorem.

Theorem 16.27.3. *Consider the restriction of the normal map F to a neighborhood as above. We modify the metric on the source Σ by the conformal factor given by the Gaussian curvature, as above. Then the map*

$$F : (\Sigma, \hat{g}_\Sigma) \to (S^2, h)$$

preserves areas: the area of the neighborhood in Σ (with respect to the modified metric) equals to the "area" of its image on the sphere.

16.28. Gauss map

Definition 16.28.1. The Gauss map is the map

$$F : M \to S^2, \quad p \mapsto N_p$$

defined by sending a point p of M the unit normal vector $N = N_p$ thought of as a point of S^2.

The map F sends an infinitesimal parallelogram on the surface, to an infinitesimal parallelogram on the sphere.

We may identify the tangent space to M at p and the tangent space to S^2 at $F(p) \in S^2$. Then the differential of the map F is the Weingarten map

$$W : T_pM \to T_{F(p)}S^2.$$

The element of area KdA of the surface is mapped to the element of area of the sphere. In other words, we modify the element of area by multiplying by the determinant (Jacobian) of the Weingarten map, namely the Gaussian curvature $K(p)$. Hence the image of the area element KdA is precisely area 2-form h on the sphere, as discussed in the previous section.

It remains to be checked that the map has topological degree given by half the Euler characteristic of the surface M, proving the theorem in the case of imbedded surfaces. Since degree is invariant under continuous deformations, the result can be checked for a particular standard imbedding of a surface of arbitrary genus in $\mathbb{R}^3$.

16.29. An identity

Another way of writing identity (16.22.1) is as follows:

$$\mathrm{vol}(\alpha) = \frac{1}{j}\,\mathrm{vol}(j\alpha).$$

This phenomenon is no longer true for higher-dimensional manifolds. Namely, the volume of a homology class is no longer multiplicative. However, the limit as $j \to \infty$ exists and is called the stable norm.

Definition 16.29.1. Let M be a compact manifold of arbitrary dimension. The *stable norm* $\|\ \|$ of a class $\alpha \in H_1(M;\mathbb{Z})$ is the limit

$$\|\alpha\| = \lim_{j\to\infty} \frac{1}{j}\,\mathrm{vol}(j\alpha). \tag{16.29.1}$$

It is obvious from the definition that one has $\|\alpha\| \le \mathrm{vol}(\alpha)$. However, the inequality may be strict in general. As noted above, for 2-dimensional manifolds we have $\|\alpha\| = \mathrm{vol}(\alpha)$.

Proposition 16.29.2. *The stable norm vanishes for a class of finite order.*

Proof. If $\alpha \in H_1(M,\mathbb{Z})$ is a class of finite order, one has finitely many possibilities for $\mathrm{vol}(j\alpha)$ as j varies. The factor of $\frac{1}{j}$ in (16.29.1) shows that $\|\alpha\| = 0$. □

Similarly, if two classes differ by a class of finite order, their stable norms coincide. Thus the stable norm passes to the quotient lattice defined below.

DEFINITION 16.29.3. The torsion subgroup of $H_1(M;\mathbb{Z})$ will be denoted $T_1(M) \subset H_1(M;\mathbb{Z})$. The quotient lattice $L_1(M)$ is the lattice

$$L_1(M) = H_1(M;\mathbb{Z})/T_1(M).$$

PROPOSITION 16.29.4. *The lattice $L_1(M)$ is isomorphic to $\mathbb{Z}^{b_1(M)}$, where b_1 is called the first Betti number of M.*

PROOF. This is a general result in the theory of finitely generated abelian groups. □

16.30. Stable systole

DEFINITION 16.30.1. Let M be a manifold endowed with a Riemannian metric, and consider the associated stable norm $\| \ \|$. The stable 1-systole of M, denoted $\mathrm{stsys}_1(M)$, is the least norm of a 1-homology class of infinite order:

$$\begin{aligned} \mathrm{stsys}_1(M) &= \inf\left\{\|\alpha\| \mid \alpha \in H_1(M,\mathbb{Z}) \setminus T_1(M)\right\} \\ &= \lambda_1\left(L_1(M), \| \ \|\right). \end{aligned}$$

EXAMPLE 16.30.2. For an arbitrary metric on the 2-torus $\mathbb{T}^2$, the 1-systole and the stable 1-systole coincide by Theorem 16.22.3:

$$\mathrm{sys}_1(\mathbb{T}^2) = \mathrm{stsys}_1(\mathbb{T}^2),$$

for every metric on $\mathbb{T}^2$.

16.31. Free loops, based loops, and fundamental group

One can refine the notion of simple connectivity by introducing a group, denoted

$$\pi_1(X) = \pi_1(X, x_0),$$

and called the fundamental group of X relative to a fixed "base" point $x_0 \in X$.

DEFINITION 16.31.1. A *based loop* is a loop $\alpha : [0,1] \to X$ satisfying the condition $\alpha(0) = \alpha(1) = x_0$.

In terms of the second item of Definition 16.18.1, we choose a fixed point $s_0 \in S^1$. For example, we can choose $s_0 = e^{i0} = 1$ for the usual unit circle $S^1 \subset \mathbb{C}$, and require that $\alpha(s_0) = x_0$. Then the group $\pi_1(X)$ is the quotient of the space of all based loops modulo the equivalence equivalence relation defined by homotopies fixing the basepoint, *cf.* Definition 16.31.2.

The equivalence class of the identity element is precisely the set of contractible loops based at x_0. The equivalence relation can be described as follows for a pair of loops in terms of the second item of Definition 16.18.1.

Definition 16.31.2. Two based loops $\alpha, \beta : S^1 \to X$ are equivalent, or *homotopic*, if there is a continuous map of the cylinder $S^1 \times [c, d] \to X$ whose restriction to $S^1 \times \{c\}$ is α, whose restriction to $S^1 \times \{d\}$ is β, while the map is constant on the segment $\{s_0\} \times [c, d]$ of the cylinder, *i.e.* the basepoint does not move during the homotopy.

Definition 16.31.3. An equivalence class of based loops is called a *based homotopy class.* Removing the basepoint restriction (as well as the constancy condition of Definition 16.31.2), we obtain a larger class called a *free homotopy class (of loops).*

Definition 16.31.4. *Composition* of two loops is defined most conveniently in terms of item 1 of Definition 16.18.1, by concatenating their domains.

In more detail, the product of a pair of loops, $\alpha : [-1, 0] \to X$ and $\beta : [0, +1] \to X$, is a loop $\alpha.\beta : [-1, 1] \to X$, which coincides with α and β in their domains of definition. The product loop $\alpha.\beta$ is continuous since $\alpha(0) = \beta(0) = x_0$. Then Theorem 16.18.5 can be refined as follows.

16.32. Fundamental groups of surfaces

Theorem 16.32.1. *We have* $\pi_1(S^1) = \mathbb{Z}$, *while* $\pi_1(S^n)$ *is the trivial group for all* $n \geq 2$.

Definition 16.32.2. The 2-torus $\mathbb{T}^2$ is defined to be the following Cartesian product: $\mathbb{T}^2 = S^1 \times S^1$, and can thus be realized as a subset

$$\mathbb{T}^2 = S^1 \times S^1 \subset \mathbb{R}^2 \times \mathbb{R}^2 = \mathbb{R}^4.$$

We have $\pi_1(\mathbb{T}^2) = \mathbb{Z}^2$. The familiar doughnut picture realizes $\mathbb{T}^2$ as a subset in Euclidean space $\mathbb{R}^3$.

Definition 16.32.3. A 2-dimensional closed Riemannian manifold (*i.e.* surface) Σ is called *orientable* if it can be realized by a subset of $\mathbb{R}^3$.

Theorem 16.32.4. *The fundamental group of a surface of genus* g *is isomorphic to a group on* $2g$ *generators* $a_1, b_1, \ldots, a_g, b_g$ *satisfying* *the unique relation*

$$\prod_{i=1}^{g} a_i b_i a_i^{-1} b_i^{-1} = 1.$$

Note that this is not the only possible presentation of the group in terms of a single relation.

CHAPTER 17

Pu's inequality

17.1. Hopf fibration h

To prove Pu's inequality, we will need to study the Hopf fibration more closely.

The circle action in $\mathbb{C}^2$ restricts to the unit sphere $S^3 \subset \mathbb{C}^2$, which therefore admits a fixed point free circle action. Namely, the circle

$$S^1 = \{e^{i\theta} : \theta \in \mathbb{R}\} \subset \mathbb{C}$$

acts on a point $(z_1, z_2) \in \mathbb{C}^2$ by

$$e^{i\theta}.(z_1, z_2) = (e^{i\theta} z_1, e^{i\theta} z_2). \tag{17.1.1}$$

The quotient manifold (space of orbits) S^3/S^1 is the 2-sphere S^2 (from the projective viewpoint, the quotient space is the complex projective line $\mathbb{CP}^1$).

Definition 17.1.1. The quotient map

$$h : S^3 \to S^2, \tag{17.1.2}$$

called the *Hopf fibration*.

17.2. Tangent map

Proposition 17.2.1. *Consider a smooth map $f : M \to N$ between differentiable manifolds. Then f defines a natural map*

$$df : TM \to TN$$

called the tangent map.

Proof. We represent a tangent vector $v \in T_pM$ by a curve

$$c(t) : I \to M,$$

such that $c'(o) = v$. The composite map $\sigma = f \circ c : I \to N$ is a curve in N. Then the vector $\sigma'(0)$ is the image of v under df:

$$df(v) = \sigma'(0)$$

by chain rule, proving the proposition. □

17.3. Riemannian submersion

DEFINITION 17.3.1. A projection $(\mathbb{R}^2, g) \to \mathbb{R}$ to the u^1-axis is called a *Riemannian submersion* if the metric coefficients in the horizontal direction are independent of u^2, or more precisely,

$$g = h(u^1)\left(du^1\right)^2 + k(u^1, u^2)\left(du^2\right)^2, \qquad (17.3.1)$$

for suitable functions h and k (note that the matrix of metric coefficients is diagonal but not necessarily scalar).

Replacing the coordinates by (x, y), we obtain a somewhat more readable formula

$$g = h(x)dx^2 + k(x, y)dy^2.$$

Being a Riemannian submersion is, of course, a local property, and we stated it for Euclidean space for convenience. We will use it mostly in the context of maps between compact manifolds.

EXAMPLE 17.3.2. The metric (16.17.1) of the Poincaré upperhalf plane admits a Riemannian submersion to (the positive ray of) the y-axis but not to the x-axis, as the conformal factor depends on y.

17.4. Riemannian submersions

Riemannian submersions in the 2-dimensional case were dealt with in an Section 17.3.

Consider a fiber bundle map $f : M \to N$ between closed manifolds, i.e., a smooth map whose df is onto. Let $F \subset M$ be the fiber over a point $p \in N$. The differential $df : TM \to TN$ vanishes on the subspace $T_xF \subset TM$ at every point $x \in F$. More precisely, we have the vertical space[1]

$$\ker(df_x) = T_xF.$$

Given a Riemannian metric on M, we can consider the orthogonal complement of T_xF in T_xM, denoted

$$H_x \subset T_xM.$$

Thus

$$T_xM = T_xF \oplus H_x$$

is an orthogonal decomposition.

DEFINITION 17.4.1. A fiber bundle map $f : M \to N$ between Riemannian manifolds is called a Riemannian submersion if at every point $x \in M$, the restriction of $df : H_x \to T_{f(x)}N$ is an isometry.

[1] anchi

PROPOSITION 17.4.2. *Let S^3 be the 3-sphere of constant sectional curvature +1, i.e., the unit sphere in $\mathbb{R}^4$. The Hopf fibration*

$$h : S^3 \to S^2$$

is a Riemannian submersion, for which the natural metric on the base S^2 is a metric of constant Gaussian curvature +4.

The latter is explained by the fact that the maximal distance between a pair of S^1 orbits is $\frac{\pi}{2}$ rather than π, in view of the fact that a pair of antipodal points of S^3 lies in a common orbit and therefore descends to the same point of the quotient space.

A pair of points at maximal distance is defined by a pair $v, w \in S^3$ such that w is orthogonal to $\mathbb{C}v$.

17.5. Hamilton quaternions

The quaternionic viewpoint will be applied in Section ?The algebra of the Hamilton quaternions is the real 4-dimensional vector space with real basis $\{1, i, j, k\}$, so that

$$\mathbb{H} = \mathbb{R}1 + \mathbb{R}i + \mathbb{R}j + \mathbb{R}k$$

equipped with an associative (chok kibutz), non-commutative product operation. This operation has the following properties:

(1) the center of $\mathbb{H}$ is $\mathbb{R}1$;
(2) the operation satisfies the relations $i^2 = j^2 = k^2 = -1$ and $ij = -ji = k$, $jk = -kj = i$, $ki = -ik = j$.

LEMMA 17.5.1. *The algebra $\mathbb{H}$ is naturally isomorphic to the complex vector space $\mathbb{C}^2$ via the identification $\mathbb{R}1 + \mathbb{R}i \simeq \mathbb{C}$.*

PROOF. The subspace $\mathbb{R}j + \mathbb{R}k$ can be thought of as

$$\mathbb{R}j + \mathbb{R}ij = (\mathbb{R}1 + \mathbb{R}i)j,$$

and we therefore obtain a natural isomorphism

$$\mathbb{H} = \mathbb{C}1 + \mathbb{C}j,$$

showing that $\{1, j\}$ is a complex basis for $\mathbb{H}$. □

THEOREM 17.5.2. *Nonzero quaternions form a group under quaternionic multiplication.*

PROOF. Given $q = a + bi + cj + dk \in \mathbb{H}$, let $N(q) = a^2 + b^2 + c^2 + d^2$, and $\bar{q} = a - bi - cj - dk$. Then $q\bar{q} = N(q)$. Therefore we have a multiplicative inverse

$$q^{-1} = \frac{1}{N(q)}\bar{q},$$

proving the theorem. □

17.6. Complex structures on the algebra $\mathbb{H}$

CHAPTER 18

Approach using energy-area identity

18.1. An integral-geometric identity

Let $\mathbb{T}^2$ be a torus with an arbitrary metric. Let $\mathbb{T}_0 = \mathbb{R}^2/L$ be the flat torus conformally equivalent to $\mathbb{T}^2$. Let $\ell_0 = \ell_0(x)$ be a simple closed geodesic of $\mathbb{T}_0$. Thus ℓ_0 is the projection of a line $\tilde{\ell}_0 \subset \mathbb{R}^2$. Let $\tilde{\ell}_y \subset \mathbb{R}^2$ be the line parallel to $\tilde{\ell}_0$ at distance $y > 0$ from ℓ_0 (here we must "choose sides", *e.g.* by orienting $\tilde{\ell}_0$ and requiring $\tilde{\ell}_y$ to lie to the left of $\tilde{\ell}_0$). Denote by $\ell_y \subset \mathbb{T}_0$ the closed geodesic loop defined by the image of $\tilde{\ell}_y$. Let $y_0 > 0$ be the *smallest* number such that $\ell_{y_0} = \ell_0$, *i.e.* the lines $\tilde{\ell}_{y_0}$ and $\tilde{\ell}_0$ both project to ℓ_0.

Note that the loops in the family $\{\ell_y\} \subset \mathbb{T}^2$ are not necessarily geodesics with respect to the metric of $\mathbb{T}^2$. On the other hand, the family satisfies the following identity.

Lemma 18.1.1 (An elementary integral-geometric identity). *The metric on $\mathbb{T}^2$ satisfies the following identity:*

$$\operatorname{area}(\mathbb{T}^2) = \int_0^{y_0} E(\ell_y)dy, \tag{18.1.1}$$

where E is the energy of a loop with respect to the metric of $\mathbb{T}^2$, see Definition 16.9.2.

Proof. Denote by f^2 the conformal factor of $\mathbb{T}^2$ with respect to the flat metric $\mathbb{T}_0$. Thus the metric on $\mathbb{T}^2$ can be written as

$$f^2(x,y)(dx^2 + dy^2).$$

By Fubini's theorem applied to a rectangle with sides $\operatorname{length}_{\mathbb{T}_0}(\ell_0)$ and y_0, combined with Theorem 16.7.3, we obtain

$$\begin{aligned}\operatorname{area}(\mathbb{T}^2) &= \int_{\mathbb{T}_0} f^2 dxdy \\ &= \int_0^{y_0} \left(\int_{\ell_y} f^2(x,y)dx \right) dy \\ &= \int_0^{y_0} E(\ell_y)dy,\end{aligned}$$

proving the lemma. □

REMARK 18.1.2. The identity (18.1.1) can be thought of as the simplest integral-geometric identity, *cf.* equation (18.4.3).

18.2. Two proofs of the Loewner inequality

We give a slightly modified version of M. Gromov's proof [**Gro96**], using conformal representation and the Cauchy-Schwarz inequality, of the Loewner inequality (15.2.1) for the 2-torus, see also [**CK03**]. We present the following slight generalisation: there exists a pair of closed geodesics on $(\mathbb{T}^2, g)$, of respective lengths λ_1 and λ_2, such that

$$\lambda_1 \lambda_2 \leq \gamma_2 \operatorname{area}(g), \tag{18.2.1}$$

and whose homotopy classes form a generating set for $\pi_1(\mathbb{T}^2) = \mathbb{Z} \times \mathbb{Z}$.

PROOF. The proof relies on the conformal representation

$$\phi : \mathbb{T}_0 \to (\mathbb{T}^2, g),$$

where $\mathbb{T}_0$ is flat, *cf.* uniformisation theorem 13.6.1. Here the map ϕ may be chosen in such a way that $(\mathbb{T}^2, g)$ and $\mathbb{T}_0$ have the same area. Let f be the conformal factor, so that

$$g = f^2 \left((du^1)^2 + (du^2)^2 \right)$$

as in formula (16.7.1), where $(du^1)^2 + (du^2)^2$ (locally) is the flat metric.

Let ℓ_0 be any closed geodesic in $\mathbb{T}_0$. Let $\{\ell_s\}$ be the family of geodesics parallel to ℓ_0. Parametrize the family $\{\ell_s\}$ by a circle S^1 of length σ, so that

$$\sigma \ell_0 = \operatorname{area}(\mathbb{T}_0).$$

Thus $\mathbb{T}_0 \to S^1$ is a Riemannian submersion, *cf.* Definition 17.3.1. Then

$$\operatorname{area}(\mathbb{T}^2) = \int_{\mathbb{T}_0} f^2.$$

By Fubini's theorem, we have $\operatorname{area}(\mathbb{T}^2) = \int_{S^1} ds \int_{\ell_s} f^2 dt$. Therefore by the Cauchy-Schwarz inequality,

$$\operatorname{area}(\mathbb{T}^2) \geq \int_{S^1} ds \frac{\left(\int_{\ell_s} f dt \right)^2}{\ell_0} = \frac{1}{\ell_0} \int_{S^1} ds \left(\operatorname{length} \phi(\ell_s) \right)^2.$$

Hence there is an s_0 such that $\operatorname{area}(\mathbb{T}^2) \geq \frac{\sigma}{\ell_0} \operatorname{length} \phi(\ell_{s_0})^2$, so that

$$\operatorname{length} \phi(\ell_{s_0}) \leq \ell_0.$$

This reduces the proof to the flat case. Given a lattice in $\mathbb{C}$, we choose a shortest lattice vector λ_1, as well as a shortest one λ_2 not proportional to λ_1. The inequality now follows from Lemma 6.3.1. In the boundary case of equality, one can exploit the equality in the Cauchy-Schwarz inequality to prove that the conformal factor must be constant. □

Alternative proof. Let ℓ_0 be any simple closed geodesic in $\mathbb{T}_0$. Since the desired inequality (18.2.1) is scale-invariant, we can assume that the loop has unit length:

$$\mathrm{length}_{\mathbb{T}_0}(\ell_0) = 1,$$

and, moreover, that the corresponding covering transformation of the universal cover $\mathbb{C} = \tilde{\mathbb{T}}_0$ is translation by the element $1 \in \mathbb{C}$. We complete 1 to a basis $\{\tau, 1\}$ for the lattice of covering transformations of $\mathbb{T}_0$. Note that the rectangle defined by

$$\{z = x + iy \in \mathbb{C} \mid 0 < x < 1,\ 0 < y < \mathrm{Im}(\tau)]\}$$

is a fundamental domain for $\mathbb{T}_0$, so that $\mathrm{area}(\mathbb{T}_0) = \mathrm{Im}(\tau)$. The maps

$$\ell_y(x) = x + iy,\ x \in [0, 1]$$

parametrize the family of geodesics parallel to ℓ_0 on $\mathbb{T}_0$. Recall that the metric of the torus $\mathbb{T}$ is $f^2(dx^2 + dy^2)$.

Lemma 18.2.1. *We have the following relation between the length and energy of a loop:*

$$\mathrm{length}(\ell_y)^2 \leq E(\ell_y).$$

Proof. By the Cauchy-Schwarz inequality,

$$\int_0^1 f^2(x, y)dx \geq \left(\int_0^1 f(x, y)dx\right)^2,$$

proving the lemma. □

Now by Lemma 18.1.1 and Lemma 18.2.1, we have

$$\mathrm{area}(\mathbb{T}^2) \geq \int_0^{\mathrm{Im}(\tau)} \left(\mathrm{length}(\ell_y)\right)^2 dy.$$

Hence there is a y_0 such that

$$\mathrm{area}(\mathbb{T}^2) \geq \mathrm{Im}(\tau)\,\mathrm{length}(\ell_{y_0})^2,$$

so that $\mathrm{length}(\ell_{y_0}) \leq 1$. This reduces the proof to the flat case. Given a lattice in $\mathbb{C}$, we choose a shortest lattice vector λ_1, as well as a shortest one λ_2 not proportional to λ_1. The inequality now follows from Lemma 6.3.1. □

REMARK 18.2.2. Define a conformal invariant called the *capacity* of an annulus as follows. Consider a right circular cylinder

$$\zeta_\kappa = \mathbb{R}/\mathbb{Z} \times [0, \kappa]$$

based on a unit circle $\mathbb{R}/\mathbb{Z}$. Its capacity $C(\zeta_\kappa)$ is defined to be its height, $C(\zeta_\kappa) = \kappa$. Recall that every annular region in the plane is conformally equivalent to such a cylinder, and therefore we have defined a conformal invariant of an arbitrary annular region. Every annular region R satisfies the inequality $\operatorname{area}(R) \geq C(R) \operatorname{sys}_1(R)^2$. Meanwhile, if we cut a flat torus along a shortest loop, we obtain an annular region R with capacity at least $C(R) \geq \gamma_2^{-1} = \frac{\sqrt{3}}{2}$. This provides an alternative proof of the Loewner theorem. In fact, we have the following identity:

$$\operatorname{confsys}_1(g)^2 C(g) = 1, \tag{18.2.2}$$

where confsys is the conformally invariant generalisation of the homology systole, while $C(g)$ is the largest capacity of a cylinder obtained by cutting open the underlying conformal structure on the torus.

QUESTION 18.2.3. Is the Loewner inequality (15.2.1) satisfied by *every* orientable nonsimply connected compact surface? Inspite of its elementary nature, and considerable research devoted to the area, the question is still open. Recently the case of genus 2 was settled as well as genus $g \geq 20$

18.3. Hopf fibration and the Hamilton quaternions

The circle action in $\mathbb{C}^2$ restricts to the unit sphere $S^3 \subset \mathbb{C}^2$, which therefore admits a fixed point free circle action. Namely, the circle $S^1 = \{e^{i\theta} : \theta \in \mathbb{R}\} \subset \mathbb{C}$ acts on (z_1, z_2) by

$$e^{i\theta}.(z_1, z_2) = (e^{i\theta} z_1, e^{i\theta} z_2).$$

It is possible to show that the quotient manifold (space of orbits) S^3/S^1 is the 2-sphere S^2 (from the projective viewpoint, the quotient is the complex projective line $\mathbb{CP}^1$). Moreover, the quotient map

$$S^3 \to S^2, \tag{18.3.1}$$

called the *Hopf fibration*, is a Riemannian submersion, for which the natural metric on the base is a metric of constant Gaussian curvature $+4$ rather than $+1$. The latter is explained by the fact that the maximal distance between a pair of S^1 orbits is $\frac{\pi}{2}$ rather than π, in view of the fact that a pair of antipodal points of S^3 lies in a common orbit.

Furthermore, the sphere S^3 can be thought of as the Lie group formed of the unit (Hamilton) quaternions in $\mathbb{H} = \mathbb{C}^2$. The Lie group $SO(3)$

can be identified with $S^3/\{\pm 1\}$. Therefore the fibration (18.3.1) descends to a fibration

$$SO(3) \to S^2, \tag{18.3.2}$$

exploited in the next section.

REMARK 18.3.1. The group $G = SO(3)$ can be identified with the unit tangent bundle of S^2, as follows. Choose a fixed unit tangent vector v to S^2, and send an element $g \in G$, acting on S^2, to the image of v under $dg : TS^2 \to TS^2$.

REMARK 18.3.2. Let us elaborate a bit further on the quaternionic viewpoint, with an eye to aplying it in Section 18.4. A choice of a purely imaginary Hamilton quaternion in $\mathbb{H} = \mathbb{R}1 + \mathbb{R}i + \mathbb{R}j + \mathbb{R}k$, *i.e.* a linear combination of i, j, and k, specifies a complex structure on $\mathbb{H}$. Such a complex structure leads to a fibration of the unit sphere, as in (18.3.1). Choosing a pair of purely imaginary quaternions which are orthogonal, results in a pair of circle fibrations of S^3 such that a fiber of one fibration is perpendicular to a fiber of the other fibration whenever two such fibers meet. The construction of two "perpendicular" fibrations also descends to the quotient by the natural $\mathbb{Z}_2$ action on S^3.

18.4. Double fibration of $SO(3)$ and integral geometry on S^2

In this section, we present a self-contained account of an integral-geometric identity, used in the proof of Pu's inequality (15.1.6) in Section 18.5, as well as in our argument in Lemma The identity has its origin in results of P. Funk [**Fu16**] determining a symmetric function on the two-sphere from its great circle integrals; see [**Hel99**, Proposition 2.2, p. 59], as well as Preface therein.

The sphere S^2 is the homogeneous space of the Lie group $SO(3)$, *cf.* (18.3.2), so that $S^2 = SO(3)/SO(2)$, *cf.* [**Ar83**, p. 82]. Denote by $SO(2)_\sigma$ the fiber over (stabilizer of) a typical point $\sigma \in S^2$. The projection

$$p : SO(3) \to S^2 \tag{18.4.1}$$

is a Riemannian submersion for the standard metric of constant sectional curvature $\frac{1}{4}$ on $SO(3)$. The total space $SO(3)$ admits another Riemannian submersion, which we denote

$$q : SO(3) \to \widetilde{\mathbb{RP}}^2, \tag{18.4.2}$$

whose typical fiber ν is an orbit of the geodesic flow on $SO(3)$ when the latter is viewed as the unit tangent bundle of S^2, *cf.* Remark 18.3.1. Here $\widetilde{\mathbb{RP}}^2$ denotes the double cover of $\mathbb{RP}^2$, homeomorphic to the 2-sphere, *cf.* Definition 16.12.2. A fiber of fibration q is the collection of

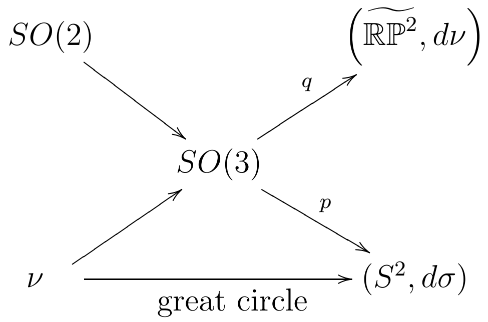

FIGURE 18.4.1. Integral geometry on S^2, *cf.* (18.4.1) and (18.4.2)

unit vectors tangent to a given directed closed geodesic (great circle) on S^2. Thus, each orbit ν projects under p to a great circle on the sphere. We think of the base space $\widetilde{\mathbb{RP}}^2$ of q as the configuration space of oriented great circles on the sphere, with measure $d\nu$. While fibration q may seem more "mysterious" than fibration p, the two are actually equivalent from the quaternionic point of view, *cf.* Remark 18.3.2.

The diagram of Figure 18.4.1 illustrates the maps defined so far.

Denote by $E_g(\nu)$ the energy, and by $L_g(\nu)$ the length, of a curve ν with respect to a (possibly singular) metric $g = f^2 g_0$.

THEOREM 18.4.1. *We have the following integral-geometric identity:*

$$\operatorname{area}(g) = \frac{1}{2\pi}\int_{\widetilde{\mathbb{RP}}^2} E_g(\nu)d\nu.$$

PROOF. We apply Fubini's theorem [**Ru87**, p. 164] twice, to both q and p, to obtain

$$\begin{aligned} \int_{\widetilde{\mathbb{RP}}^2} E_g(\nu)d\nu &= \int_{\widetilde{\mathbb{RP}}^2} d\nu \left(\int_\nu f^2\circ p\circ \nu(t)dt\right) \\ &= \int_{SO(3)} f^2\circ p \\ &= \int_{S^2} f^2\, d\sigma \left(\int_{SO(2)_\sigma} 1\right) \\ &= 2\pi\ \operatorname{area}(g), \end{aligned} \tag{18.4.3}$$

completing the proof of the theorem. □

PROPOSITION 18.4.2. *Let f be a square integrable function on S^2, which is positive and continuous except possibly for a finite number of* *points where f either vanishes or has a singularity of type $\frac{1}{\sqrt{r}}$, Then there is a great circle ν such that the following inequality is satisfied:*

$$\left(\int_\nu f(\nu(t))dt \right)^2 \leq \pi \int_{S^2} f^2(\sigma)d\sigma, \tag{18.4.4}$$

where t is the arclength parameter, and $d\sigma$ is the Riemannian measure of the metric g_0 of the standard unit sphere. In other words, there is a great circle of length L in the metric $f^2 g_0$, so that $L^2 \leq \pi A$, where A is the Riemannian surface area of the metric $f^2 g_0$. In the boundary case of equality in (18.4.4)*, the function f must be constant.*

Proof. The proof is an averaging argument and shows that the average length of great circles is short. Comparing length and energy yields the inequality

$$\frac{\left(\int_{\widetilde{\mathbb{RP}^2}} L_g(\nu)^2 \right)}{2\pi} \leq \int_{\widetilde{\mathbb{RP}^2}} E_g(\nu)d\nu. \tag{18.4.5}$$

The formula now follows from Theorem 18.4.1, since $\text{area}(\widetilde{\mathbb{R}P^2}) = 4\pi$.

In the boundary case of equality in (18.4.4), one must have equality also in inequality (18.4.5) relating length and energy. It follows that the conformal factor f must be constant along every great circle. Hence f is constant everywhere on S^2. □

18.5. Proof of Pu's inequality

A metric g on $\mathbb{RP}^2$ lifts to a centrally symmetric metric $\tilde{g}$ on S^2. Applying Proposition 18.4.2 to $\tilde{g}$, we obtain a great circle of $\tilde{g}$-length L satisfying $L^2 \leq \pi A$, where A is the Riemannian surface area of $\tilde{g}$. Thus we have $\left(\frac{L}{2}\right)^2 \leq \frac{\pi}{2}\left(\frac{A}{2}\right)$, where $\text{sys}_1(g) \leq \frac{L}{2}$ while $\text{area}(g) = \frac{A}{2}$, proving inequality (15.1.6). See [**Iv02**] for an alternative proof.

18.6. A table of optimal systolic ratios of surfaces

Denote by $\text{SR}(\Sigma)$ the supremum of the systolic ratios,

$$\text{SR}(\Sigma) = \sup_g \frac{\text{sys}_1(g)^2}{\text{area}(g)},$$

ranging over all metrics g on a surface Σ. The known values of the optimal systolic ratio are tabulated in Figure 18.6.1. It is interesting to note that the optimal ratio for the Klein bottle $\mathbb{RP}^2 \# \mathbb{RP}^2$ is achieved by a singular metric, described in the references listed in the table.

	$\mathrm{SR}(\Sigma)$	numerical value	where to find it
$\Sigma = \mathbb{RP}^2$	$= \dfrac{\pi}{2}$ [**Pu52**]	≈ 1.5707	Section 18.5
infinite $\pi_1(\Sigma)$	$< \dfrac{4}{3}$ [**Gro83**]	$< 1.3333\ldots$	(15.1.7)
$\Sigma = \mathbb{T}^2$	$= \frac{2}{\sqrt{3}}$ (Loewner)	≈ 1.1547	(15.2.1)
$\Sigma = \mathbb{RP}^2 \# \mathbb{RP}^2$	$= \dfrac{\pi}{2^{3/2}}$	≈ 1.1107	[**Bav86, Bav06, Sak88**]
Σ of genus 2	$> \frac{1}{3}\left(\sqrt{2}+1\right)$	> 0.8047	?
Σ of genus 3	$\geq \frac{8}{7\sqrt{3}}$	> 0.6598	[**Cal96**]

FIGURE 18.6.1. Values for optimal systolic ratio SR of surface Σ

CHAPTER 19

A primer on surfaces

In this Chapter, we collect some classical facts on Riemann surfaces. More specifically, we deal with hyperelliptic surfaces, real surfaces, and Katok's optimal bound for the entropy of a surface.

19.1. Hyperelliptic involution

Let Σ be an orientable closed Riemann surface which is not a sphere. By a Riemann surface, we mean a surface equipped with a fixed conformal structure, *cf.* Definition 16.7.4, while all maps are angle-preserving.

Furthermore, we will assume that the genus is at least 2.

DEFINITION 19.1.1. A *hyperelliptic involution* of a Riemann surface Σ of genus p is a holomorphic (conformal) map, $J : \Sigma \to \Sigma$, satisfying $J^2 = 1$, with $2p + 2$ fixed points.

DEFINITION 19.1.2. A surface Σ admitting a hyperelliptic involution will be called a *hyperelliptic surface*.

REMARK 19.1.3. The involution J can be identified with the nontrivial element in the center of the (finite) automorphism group of Σ (*cf.* [**FK92**, p. 108]) when it exists, and then such a J is unique, *cf.* [**Mi95**, p.204] (recall that the genus is at least 2).

It is known that the quotient of Σ by the involution J produces a conformal branched 2-fold covering

$$Q : \Sigma \to S^2 \tag{19.1.1}$$

of the sphere S^2.

DEFINITION 19.1.4. The $2p + 2$ fixed points of J are called *Weierstrass points*. Their images in S^2 under the ramified double cover Q of formula (19.1.1) will be referred to as *ramification points*.

A notion of a Weierstrass point exists on any Riemann surface, but will only be used in the present text in the hyperelliptic case.

EXAMPLE 19.1.5. In the case $p = 2$, topologically the situation can be described as follows. A simple way of representing the figure 8 contour in the (x, y) plane is by the reducible curve

$$(((x-1)^2+y^2)-1)(((x+1)^2+y^2)-1)=0 \tag{19.1.2}$$

(or, alternatively, by the lemniscate $r^2 = \cos 2\theta$ in polar coordinates, *i.e.* the locus of the equation $(x^2+y^2)^2 = x^2-y^2$).

Now think of the figure 8 curve of (19.1.2) as a subset of $\mathbb{R}^3$. The boundary of its tubular neighborhood in $\mathbb{R}^3$ is a genus 2 surface. Rotation by π around the x-axis has six fixed points on the surface, namely, a pair of fixed points near each of the points -2, 0, and $+2$ on the x-axis. The quotient by the rotation can be seen to be homeomorphic to the sphere.

A similar example can be repackaged in a metrically more precise way as follows.

Example 19.1.6. We start with a round metric on $\mathbb{RP}^2$. Now attach a small handle. The orientable double cover Σ of the resulting surface can be thought of as the unit sphere in $\mathbb{R}^3$, with two little handles attached at north and south poles, *i.e.* at the two points where the sphere meets the z-axis. Then one can think of the hyperelliptic involution J as the rotation of Σ by π around the z-axis. The six fixed points are the six points of intersection of Σ with the z-axis. Furthermore, there is an orientation reversing involution τ on Σ, given by the restriction to Σ of the antipodal map in $\mathbb{R}^3$. The composition $\tau \circ J$ is the reflection fixing the xy-plane, in view of the following matrix identity:

$$\begin{pmatrix}-1&0&0\\0&-1&0\\0&0&-1\end{pmatrix}\begin{pmatrix}-1&0&0\\0&-1&0\\0&0&1\end{pmatrix}=\begin{pmatrix}1&0&0\\0&1&0\\0&0&-1\end{pmatrix}. \tag{19.1.3}$$

Meanwhile, the induced orientation reversing involution τ_0 on S^2 can just as well be thought of as the reflection in the xy-plane. This is because, at the level of the 2-sphere, it is "the same thing as" the composition $\tau \circ J$. Thus the fixed circle of τ_0 is precisely the equator, *cf.* formula (19.3.3). Then one gets a quotient metric on S^2 which is roughly that of the western hemisphere, with the boundary longitude folded in two, The metric has little bulges along the z-axis at north and south poles, which are leftovers of the small handle.

19.2. Hyperelliptic surfaces

For a treatment of hyperelliptic surfaces, see [**Mi95**, p. 60-61]. By [**Mi95**, Proposition 4.11, p. 92], the affine part of a hyperelliptic surface Σ is defined by a suitable equation of the form

$$w^2 = f(z) \tag{19.2.1}$$

in $\mathbb{C}^2$, where f is a polynomial. On such an affine part, the map J is given by $J(z, w) = (z, -w)$, while the hyperelliptic quotient map $Q : \Sigma \to S^2$ is represented by the projection onto the z-coordinate in $\mathbb{C}^2$.

A slight technical problem here is that the map

$$\Sigma \to \mathbb{CP}^2, \tag{19.2.2}$$

whose image is the compactification of the solution set of (19.2.1), is not an imbedding. Indeed, there is only one point at infinity, given in homogeneous coordinates by $[0 : w : 0]$. This point is a singularity. A way of desingularizing it using an explicit change of coordinates at infinity is presented in [**Mi95**, p. 60-61]. The resulting smooth surface is unique [**DaS98**, Theorem, p. 100].

Remark 19.2.1. To explain what happens "geometrically", note that there are two points on our affine surface "above infinity". This means that for a large circle $|z| = r$, there are two circles above it satisfying equation (19.2.1) where f has even degree $2p+2$ (for a Weierstrass point we would only have one circle). To see this, consider $z = re^{ia}$. As the argument a varies from 0 to 2π, the argument of $f(z)$ will change by $(2p + 2)2\pi$. Thus, if $(re^{ia}, w(a))$ represents a continuous curve on our surface, then the argument of w changes by $(2p+2)\pi$, and hence we end up where we started, and not at $-w$ (as would be the case were the polynomial of odd degree). Thus there are *two* circles on the surface over the circle $|z| = r$. We conclude that to obtain a smooth compact surface, we will need to add two points at infinity, *cf.* discussion around [**FK92**, formula (7.4.1), p. 102].

Thus, the affine part of Σ, defined by equation (19.2.1), is a Riemann surface with a pair of punctures p_1 and p_2. A neighborhood of each p_i is conformally equivalent to a punctured disk. By replacing each punctured disk by a full one, we obtain the desired compact Riemann surface Σ. The point at infinity $[0 : w : 0] \in \mathbb{CP}^2$ is the image of both p_i under the map (19.2.2).

19.3. Ovalless surfaces

Denote by Σ^ι the fixed point set of an involution ι of a Riemann surface Σ. Let Σ be a hyperelliptic surface of even genus p. Let $J : \Sigma \to$

Σ be the hyperelliptic involution, *cf.* Definition 19.1.1. Let $\tau : \Sigma \to \Sigma$ be a fixed point free, antiholomorphic involution.

LEMMA 19.3.1. *The involution τ commutes with J and descends to S^2. The induced involution $\tau_0 : S^2 \to S^2$ is an inversion in a circle $C_0 = Q(\Sigma^{\tau\circ J})$. The set of ramification points is invariant under the action of τ_0 on S^2.*

PROOF. By the uniqueness of J, *cf.* Remark 19.1.3, we have the commutation relation

$$\tau \circ J = J \circ \tau, \tag{19.3.1}$$

cf. relation (19.1.3). Therefore τ descends to an involution τ_0 on the sphere. There are two possibilities, namely, τ is conjugate, in the group of fractional linear transformations, either to the map $z \mapsto \bar{z}$, or to the map $z \mapsto -\frac{1}{\bar{z}}$. In the latter case, τ is conjugate to the antipodal map of S^2.

In the case of even genus, there is an odd number of Weierstrass points in a hemisphere. Hence the inverse image of a great circle is a connected loop. Thus we get an action of $\mathbb{Z}_2 \times \mathbb{Z}_2$ on a loop, resulting in a contradiction.

In more detail, the set of the $2p + 2$ ramification points on S^2 is centrally symmetric. Since there is an odd number, $p + 1$, of ramification points in a hemisphere, a generic great circle $A \subset S^2$ has the property that its inverse image $Q^{-1}(A) \subset \Sigma$ is connected. Thus both involutions τ and J, as well as $\tau \circ J$, act fixed point freely on the loop $Q^{-1}(A) \subset \Sigma$, which is impossible. Therefore τ_0 must fix a point. It follows that τ_0 is an inversion in a circle. □

Suppose a hyperelliptic Riemann surface Σ admits an antiholomorphic involution τ. In the literature, the components of the fixed point set Σ^τ of τ are sometimes referred to as “ovals”. When τ is fixed point free, we introduce the following terminology.

DEFINITION 19.3.2. A hyperelliptic surface (Σ, J) of even positive genus $p > 0$ is called *ovalless real* if one of the following equivalent conditions is satisfied:

(1) Σ admits an imaginary reflection, *i.e.* a fixed point free, antiholomorphic involution τ;
(2) the affine part of Σ is the locus in $\mathbb{C}^2$ of the equation

$$w^2 = -P(z), \tag{19.3.2}$$

where P is a monic polynomial, of degree $2p + 2$, with real coefficients, no real roots, and with distinct roots.

Lemma 19.3.3. *The two ovalless reality conditions of Definition 19.3.2 are equivalent.*

Proof. A related result appears in [**GroH81**, p. 170, Proposition 6.1(2)]. To prove the implication (2) $\Longrightarrow$ (1), note that complex conjugation leaves the equation invariant, and therefore it also leaves invariant the locus of (19.3.2). A fixed point must be real, but P is positive hence (19.3.2) has no real solutions. There is no real solution at infinity, either, as there are two points at infinity which are not Weierstrass points, since P is of even degree, as discussed in Remark 19.2.1. The desired imaginary reflection τ switches the two points at infinity, and, on the affine part of the Riemann surface, coincides with complex conjugation $(z, w) \mapsto (\bar{z}, \bar{w})$ in $\mathbb{C}^2$.

To prove the implication (1) $\Longrightarrow$ (2), note that by Lemma 19.3.1, the induced involution τ_0 on $S^2 = \Sigma/J$ may be thought of as complex conjugation, by choosing the fixed circle of τ_0 to be the circle

$$\mathbb{R} \cup \{\infty\} \subset \mathbb{C} \cup \{\infty\} = S^2. \tag{19.3.3}$$

Since the surface is hyperelliptic, it is the smooth completion of the locus in $\mathbb{C}^2$ of *some* equation of the form (19.3.2), *cf.* (19.2.1). Here P is of degree $2p+2$ with distinct roots, but otherwise to be determined. The set of roots of P is the set of (the z-coordinates of) the Weierstrass points. Hence the set of roots must be invariant under τ_0. Thus the roots of the polynomial either come in conjugate pairs, or else are real. Therefore P has real coefficients. Furthermore, the leading cofficient of P may be absorbed into the w-coordinate by extracting a square root. Here we may have to rotate w by i, but at any rate the coefficients of P remain real, and thus P can be assumed monic.

If P had a real root, there would be a ramification point fixed by τ_0. But then the corresponding Weierstrass point must be fixed by τ, as well! This contradicts the fixed point freeness of τ. Thus all roots of P must come in conjugate pairs. □

19.4. Katok's entropy inequality

Let (Σ, g) be a closed surface with a Riemannian metric. Denote by $(\tilde{\Sigma}, \tilde{g})$ the universal Riemannian cover of (Σ, g). Choose a point $\tilde{x}_0 \in \tilde{\Sigma}$.

Definition 19.4.1. The volume entropy (or asymptotic volume) $h(M, g)$ of a surface (Σ, g) is defined by setting

$$h(\Sigma, g) = \lim_{R \to +\infty} \frac{\log\left(\operatorname{vol}_{\tilde{g}} B(\tilde{x}_0, R)\right)}{R}, \tag{19.4.1}$$

where $\operatorname{vol}_{\tilde{g}} B(\tilde{x}_0, R)$ is the volume (area) of the ball of radius R centered at $\tilde{x}_0 \in \tilde{\Sigma}$.

Since Σ is compact, the limit in (19.4.1) exists, and does not depend on the point $\tilde{x}_0 \in \tilde{\Sigma}$ [**Ma79**]. This asymptotic invariant describes the exponential growth rate of the volume in the universal cover.

DEFINITION 19.4.2. The minimal volume entropy, MinEnt, of Σ is the infimum of the volume entropy of metrics of unit volume on Σ, or equivalently

$$\operatorname{MinEnt}(\Sigma) = \inf_g h(\Sigma, g) \operatorname{vol}(\Sigma, g)^{\frac{1}{2}} \tag{19.4.2}$$

where g runs over the space of all metrics on Σ. For an n-dimensional manifold in place of Σ, one defines MinEnt similarly, by replacing the exponent of vol by $\frac{1}{n}$.

The classical result of A. Katok [**Kato83**] states that every metric g on a closed surface Σ with negative Euler characteristic $\chi(\Sigma)$ satisfies the optimal inequality

$$h(g)^2 \geq \frac{2\pi|\chi(\Sigma)|}{\operatorname{area}(g)}. \tag{19.4.3}$$

Inequality (19.4.3) also holds for hom ent(g) [**Kato83**], as well as the topological entropy, since the volume entropy bounds from below the topological entropy (see [**Ma79**]). We recall the following well-known fact, *cf.* [**KatH95**, Proposition 9.6.6, p. 374].

LEMMA 19.4.3. *Let (M, g) be a closed Riemannian manifold. Then,*

$$h(M, g) = \lim_{T \to +\infty} \frac{\log(P'(T))}{T} \tag{19.4.4}$$

where $P'(T)$ is the number of homotopy classes of loops based at some fixed point x_0 which can be represented by loops of length at most T.

PROOF. Let $x_0 \in M$ and choose a lift $\tilde{x}_0 \in \tilde{M}$. The group

$$\Gamma := \pi_1(M, x_0)$$

acts on $\tilde{M}$ by isometries. The orbit of $\tilde{x}_0$ under Γ is denoted $\Gamma.\tilde{x}_0$. Consider a fundamental domain Δ for the action of Γ, containing $\tilde{x}_0$. Denote by D the diameter of Δ. The cardinal of $\Gamma.\tilde{x}_0 \cap B(\tilde{x}_0, R)$ is bounded from above by the number of translated fundamental domains $\gamma.\Delta$, where $\gamma \in \Gamma$, contained in $B(\tilde{x}_0, R + D)$. It is also bounded from below by the number of translated fundamental domains $\gamma.\Delta$

contained in $B(\tilde{x}_0, R)$. Therefore, we have

$$\frac{\mathrm{vol}(B(\tilde{x}_0, R))}{\mathrm{vol}(M, g)} \le \mathrm{card}\left(\Gamma.\tilde{x}_0 \cap B(\tilde{x}_0, R)\right) \le \frac{\mathrm{vol}(B(\tilde{x}_0, R + D))}{\mathrm{vol}(M, g)}. \tag{19.4.5}$$

Take the log of these terms and divide by R. The lower bound becomes

$$\begin{aligned} &\frac{1}{R}\log\left(\frac{\mathrm{vol}(B(R))}{\mathrm{vol}(g)}\right) = \\ &\quad = \frac{1}{R}\log\left(\mathrm{vol}(B(R))\right) - \frac{1}{R}\log\left(\mathrm{vol}(g)\right), \end{aligned} \tag{19.4.6}$$

and the upper bound becomes

$$\begin{aligned} &\frac{1}{R}\log\left(\frac{\mathrm{vol}(B(R+D))}{\mathrm{vol}(g)}\right) = \\ &\quad = \frac{R+D}{R}\frac{1}{R+D}\log(\mathrm{vol}(B(R+D))) - \frac{1}{R}\log(\mathrm{vol}(g)). \end{aligned} \tag{19.4.7}$$

Hence both bounds tend to $h(g)$ when R goes to infinity. Therefore,

$$h(g) = \lim_{R \to +\infty} \frac{1}{R}\log\left(\mathrm{card}(\Gamma.\tilde{x}_0 \cap B(\tilde{x}_0, R))\right). \tag{19.4.8}$$

This yields the result since $P'(R) = \mathrm{card}(\Gamma.\tilde{x}_0 \cap B(\tilde{x}_0, R))$. $\square$

Bibliography

[Ab81] Abikoff, W.: The uniformization theorem. *Amer. Math. Monthly* **88** (1981), no. 8, 574–592.

[Ac60] Accola, R.: Differentials and extremal length on Riemann surfaces. *Proc. Nat. Acad. Sci. USA* **46** (1960), 540–543.

[Am04] Ammann, B.: Dirac eigenvalue estimates on two-tori. *J. Geom. Phys.* **51** (2004), no. 3, 372–386.

[Ar83] Armstrong, M.: Basic topology. Corrected reprint of the 1979 original. *Undergraduate Texts in Math.* Springer-Verlag, New York-Berlin, 1983.

[Ar74] Arnold, V.: Mathematical methods of classical mechanics. Translated from the 1974 Russian original by K. Vogtmann and A. Weinstein. Corrected reprint of the second (1989) edition. Graduate Texts in Mathematics, 60. Springer-Verlag, New York, 199?.

[1] Arnol'd, V. I. On the teaching of mathematics.(Russian) Uspekhi Mat. Nauk 53 (1998), no. 1(319), 229–234; translation in Russian Math. Surveys 53 (1998), no. 1, 229–236.

[Ba93] Babenko, I.: Asymptotic invariants of smooth manifolds. *Russian Acad. Sci. Izv. Math.* **41** (1993), 1–38.

[Ba02a] Babenko, I. Forte souplesse intersystolique de variétés fermées et de polyèdres. *Annales de l'Institut Fourier* **52**, 4 (2002), 1259-1284.

[Ba02b] Babenko, I. Loewner's conjecture, Besicovich's example, and relative systolic geometry. [Russian]. *Mat. Sbornik* **193** (2002), 3-16.

[Ba04] Babenko, I.: Géométrie systolique des variétés de groupe fondamental $\mathbb{Z}_2$, *Sémin. Théor. Spectr. Géom. Grenoble* **22** (2004), 25-52.

[BB05] Babenko, I.; Balacheff, F.: Géométrie systolique des sommes connexes et des revêtements cycliques, *Mathematische Annalen* (to appear).

[BaK98] Babenko, I.; Katz, M.: Systolic freedom of orientable manifolds, *Annales Scientifiques de l'E.N.S. (Paris)* **3**1 (1998), 787-809.

[Bana93] Banaszczyk, W. New bounds in some transference theorems in the geometry of numbers. *Math. Ann.* **296** (1993), 625–635.

[Bana96] Banaszczyk, W. Inequalities for convex bodies and polar reciprocal lattices in $\mathbb{R}^n$ II: Application of K-convexity. *Discrete Comput. Geom.* **16** (1996), 305–311.

[BCIK04] Bangert, V; Croke, C.; Ivanov, S.; Katz, M.: Boundary case of equality in optimal Loewner-type inequalities, *Trans. Amer. Math. Soc.*, to appear. See `arXiv:math.DG/0406008`

[BCIK05] Bangert, V; Croke, C.; Ivanov, S.; Katz, M.: Filling area conjecture and ovalless real hyperelliptic surfaces, *Geometric and Functional Analysis (GAFA)* **15** (2005) no. 3. See `arXiv:math.DG/0405583`

[BK03] Bangert, V.; Katz, M.: Stable systolic inequalities and cohomology products, *Comm. Pure Appl. Math.* **56** (2003), 979–997. Available at `arXiv:math.DG/0204181`

[BK04] Bangert, V; Katz, M.: An optimal Loewner-type systolic inequality and harmonic one-forms of constant norm. *Comm. Anal. Geom.* **12** (2004), number 3, 703-732. See `arXiv:math.DG/0304494`

[Bar57] Barnes, E. S.: On a theorem of Voronoi. *Proc. Cambridge Philos. Soc.* **53** (1957), 537–539.

[Bav86] Bavard, C.: Inégalité isosystolique pour la bouteille de Klein. *Math. Ann.* **274** (1986), no. 3, 439–441.

[Bav88] Bavard, C.: Inégalités isosystoliques conformes pour la bouteille de Klein. *Geom. Dedicata* **27** (1988), 349–355.

[Bav92] Bavard, C.: La systole des surfaces hyperelliptiques, *Prepubl. Ec. Norm. Sup. Lyon* **71** (1992).

[Bav06] Bavard, C.: Une remarque sur la géométrie systolique de la bouteille de Klein. *Arch. Math. (Basel)* **87** (2006), no. 1, 72–74.

[BeM] Bergé, A.-M.; Martinet, J. Sur un problème de dualité lié aux sphères en géométrie des nombres. *J. Number Theory* **32** (1989), 14–42.

[Ber72] Berger, M. A l'ombre de Loewner. *Ann. Scient. Ec. Norm. Sup. (Sér. 4)* **5** (1972), 241–260.

[Ber80] Berger, M.: Une borne inférieure pour le volume d'une variété riemannienne en fonction du rayon d'injectivité. *Ann. Inst. Fourier* **30** (1980), 259–265.

[Ber93] Berger, M. Systoles et applications selon Gromov. Séminaire N. Bourbaki, exposé 771, *Astérisque* **216** (1993), 279–310.

[Ber76] Berstein, I.: On the Lusternik–Schnirelmann category of Grassmannians, *Proc. Camb. Phil. Soc.* **79** (1976), 129-134.

[Bes87] Besse, A.: Einstein manifolds. *Ergebnisse der Mathematik und ihrer Grenzgebiete (3)*, **10**. Springer-Verlag, Berlin, 1987.

[BesCG95] Besson, G.; Courtois, G.;, Gallot, S.: Volume et entropie minimale des espaces localement symétriques. *Invent. Math.* **103** (1991), 417–445.

[BesCG95] Besson, G.; Courtois, G.;, Gallot, S.: Hyperbolic manifolds, amalgamated products and critical exponents. *C. R. Math. Acad. Sci. Paris* **336** (2003), no. 3, 257–261.

[Bl61] Blatter, C. Über Extremallängen auf geschlossenen Flächen. *Comment. Math. Helv.* **35** (1961), 153–168.

[Bl61b] Blatter, C.: Zur Riemannschen Geometrie im Grossen auf dem Möbiusband. *Compositio Math.* **15** (1961), 88–107.

[Bo94] Borzellino, J. E.: Pinching theorems for teardrops and footballs of revolution. *Bull. Austral. Math. Soc.* **49** (1994), no. 3, 353–364.

[BG76] Bousfield, A.; Gugenheim, V.: On PL de Rham theory and rational homotopy type. *Mem. Amer. Math. Soc.* **8**, 179, Providence, R.I., 1976.

[BI94] Burago, D.; Ivanov, S.: Riemannian tori without conjugate points are flat. *Geom. Funct. Anal.* **4** (1994), no. 3, 259–269.

[BI95] Burago, D.; Ivanov, S.: On asymptotic volume of tori. *Geom. Funct. Anal.* **5** (1995), no. 5, 800–808.

[BuraZ80] Burago, Y.; Zalgaller, V.: Geometric inequalities. *Grundlehren der Mathematischen Wissenschaften* **285**. Springer Series in Soviet Mathematics. Springer-Verlag, Berlin, 1988 (original Russian edition: 1980).

[BS94] Buser, P.; Sarnak, P.: On the period matrix of a Riemann surface of large genus. With an appendix by J. H. Conway and N. J. A. Sloane. *Invent. Math.* **117** (1994), no. 1, 27–56.

[Cal96] Calabi, E.: Extremal isosystolic metrics for compact surfaces, Actes de la table ronde de geometrie differentielle, *Sem. Congr.* **1** (1996), Soc. Math. France, 167–204.

[Ca76] Manfredo do Carmo, Differential geometry of curves and surfaces. Prentice-Hall, 1976.

[Car92] do Carmo, M.: Riemannian geometry. Translated from the second Portuguese edition by Francis Flaherty. Mathematics: Theory & Applications. Birkhäuser Boston, Inc., Boston, MA, 1992.

[Ca71] Cassels, J. W. S.: An introduction to the geometry of numbers. Second printing. *Grundlehren der mathematischen Wissenschaften*, **9**9. Springer-Verlag, Berlin-Heidelberg-New York, 1971.

[2] Cauchy, A. L.: Leọns sur les applications du calcul infinitésimal à la géométrie. Paris: Imprimérie Royale, 1826.

[Ch93] Chavel, I.: Riemannian geometry—a modern introduction. *Cambridge Tracts in Mathematics*, **1**08. Cambridge University Press, Cambridge, 1993.

[CoS94] Conway, J.; Sloane, N.: On lattices equivalent to their duals. J. Number Theory **48** (1994), no. 3, 373–382.

[ConS99] Conway, J. H.; Sloane, N. J. A.: Sphere packings, lattices and groups. Third edition. Springer, 1999.

[CoT03] Cornea, O.; Lupton, G.; Oprea, J.; Tanré, D.: Lusternik-Schnirelmann category. *Mathematical Surveys and Monographs*, **103**. American Mathematical Society, Providence, RI, 2003.

[Cr80] Croke, C.: Some isoperimetric inequalities and eigenvalue estimates. *Ann. Sci. École Norm. Sup.* **13** (1980), 519–535.

[Cr88] Croke, C.: An isoembolic pinching theorem. *Invent. Math.* **92** (1988), 385–387.

[CK03] Croke, C.; Katz, M.: Universal volume bounds in Riemannian manifolds, *Surveys in Differential Geometry* **V**III, Lectures on Geometry and Topology held in honor of Calabi, Lawson, Siu, and Uhlenbeck at Harvard University, May 3- 5, 2002, edited by S.T. Yau (Somerville, MA: International Press, 2003.) pp. 109 - 137. See `arXiv:math.DG/0302248`

[DaS98] Danilov, V. I.; Shokurov, V. V.: Algebraic curves, algebraic manifolds and schemes. Translated from the 1988 Russian original by D. Coray and V. N. Shokurov. Reprint of the original English edition from the series Encyclopaedia of Mathematical Sciences [Algebraic geometry. I, Encyclopaedia Math. Sci., 23, Springer, Berlin, 1994]. Springer-Verlag, Berlin, 1998.

[Do79] Dombrowski, P.: 150 years after Gauss' "Disquisitiones generales circa superficies curvas". With the original text of Gauss. *Astérisque*, **62**. Société Mathématique de France, Paris, 1979.

[DR03] Dranishnikov, A.; Rudyak, Y.: Examples of non-formal closed $(k-1)$-connected manifolds of dimensions $\geq 4k-1$, *Proc. Amer. Math. Soc.* **133** (2005), no. 5, 1557-1561. See `arXiv:math.AT/0306299`.

[FK92] Farkas, H. M.; Kra, I.: Riemann surfaces. Second edition. Graduate Texts in Mathematics, 71. Springer-Verlag, New York, 1992.

[Fe69] Federer, H.: Geometric Measure Theory. *Grundlehren der mathematischen Wissenschaften*, **1**53. Springer-Verlag, Berlin, 1969.

[Fe74] Federer, H. Real flat chains, cochains, and variational problems. *Indiana Univ. Math. J.* **24** (1974), 351–407.

[FHL98] Felix, Y.; Halperin, S.; Lemaire, J.-M.: The rational LS category of products and of Poincaré duality complexes. *Topology* **37** (1998), no. 4, 749–756.

[Fer77] Ferrand, J.: Sur la regularite des applications conformes, *C.R. Acad.Sci. Paris Ser.A-B* **2**84 (1977), no.1, A77-A79.

[Fr99] Freedman, M.: $\mathbb{Z}_2$-Systolic-Freedom. *Geometry and Topology Monographs* **2**, Proceedings of the Kirbyfest (J. Hass, M. Scharlemann eds.), 113–123, Geometry & Topology, Coventry, 1999.

[Fu16] Funk, P.: Uber eine geometrische Anwendung der Abelschen Integralgleichung. *Math. Ann.* **7**7 (1916), 129-135.

[GaHL04] Gallot, S.; Hulin, D.; Lafontaine, J.: Riemannian geometry. Third edition. Universitext. Springer-Verlag, Berlin, 2004.

[Ga71] Ganea, T.: Some problems on numerical homotopy invariants. Symposium on Algebraic Topology (Battelle Seattle Res. Center, Seattle Wash., 1971), pp. 23–30. *Lecture Notes in Math.,* **249**, Springer, Berlin, 1971.

[Go05] Goldman, Ron: Curvature formulas for implicit curves and surfaces. *Comput. Aided Geom. Design* **22** (2005), no. 7, 632–658. See http://u.math.biu.ac.il/ katzmik/goldman05.pdf

[GG92] Gómez-Larrañaga, J.; González-Acuña, F.: Lusternik-Schnirel'mann category of 3-manifolds. *Topology* **31** (1992), no. 4, 791–800.

[Gre67] Greenberg, M. Lectures on Algebraic Topology. W.A. Benjamin, Inc., New York, Amsterdam, 1967.

[GHV76] Greub, W.; Halperin, S.; Vanstone, R.: Connections, curvature, and cohomology. Vol. III: Cohomology of principal bundles and homogeneous spaces, *Pure and Applied Math.*, **47**, Academic Press, New York-London, 1976.

[GriH78] Griffiths, P.; Harris, J.: Principles of algebraic geometry. Pure and Applied Mathematics. Wiley-Interscience [John Wiley & Sons], New York, 1978 [1994].

[Gro81a] Gromov, M.: Structures métriques pour les variétés riemanniennes. (French) [Metric structures for Riemann manifolds] Edited by J. Lafontaine and P. Pansu. *Textes Mathématiques*, **1**. CEDIC, Paris, 1981.

[Gro81b] Gromov, M.: Volume and bounded cohomology. *Publ. IHES.* **56** (1981), 213–307.

[Gro83] Gromov, M.: Filling Riemannian manifolds, *J. Diff. Geom.* **18** (1983), 1-147.

[Gro96] Gromov, M.: Systoles and intersystolic inequalities, Actes de la Table Ronde de Géométrie Différentielle (Luminy, 1992), 291–362, *Sémin.*

Congr., **1**, Soc. Math. France, Paris, 1996.
www.emis.de/journals/SC/1996/1/ps/smf_sem-cong_1_291-362.ps.gz

[Gro99] Gromov, M.: Metric structures for Riemannian and non-Riemannian spaces, *Progr. in Mathematics*, **152**, Birkhäuser, Boston, 1999.

[GroH81] Gross, B. H.; Harris, J.: Real algebraic curves. *Ann. Sci. Ecole Norm. Sup.* (4) **14** (1981), no. 2, 157–182.

[GruL87] Gruber, P.; Lekkerkerker, C.: *Geometry of numbers.* Second edition. North-Holland Mathematical Library, **37**. North-Holland Publishing Co., Amsterdam, 1987.

[Hat02] Hatcher, A.: Algebraic topology. Cambridge University Press, Cambridge, 2002.

[He86] Hebda, J. The collars of a Riemannian manifold and stable isosystolic inequalities. *Pacific J. Math.* **121** (1986), 339-356.

[Hel99] Helgason, S.: The Radon transform. 2nd edition. Progress in Mathematics, 5. Birkhäuser, Boston, Mass., 1999.

[Hem76] Hempel, J.: 3-Manifolds. *Ann. of Math. Studies* **8**6, Princeton Univ. Press, Princeton, New Jersey 1976.

[Hil03] Hillman, J. A.: Tits alternatives and low dimensional topology. *J. Math. Soc. Japan* **55** (2003), no. 2, 365–383.

[Hi04] Hillman, J. A.: An indecomposable PD_3-complex: II, *Algebraic and Geometric Topology* **4** (2004), 1103-1109.

[Iv02] Ivanov, S.: On two-dimensional minimal fillings, *St. Petersbg. Math. J.* **13** (2002), no. 1, 17–25.

[IK04] Ivanov, S.; Katz, M.: Generalized degree and optimal Loewner-type inequalities, *Israel J. Math.* **141** (2004), 221-233. arXiv:math.DG/0405019

[Iw02] Iwase, N.: A_∞-method in Lusternik-Schnirelmann category. *Topology* **41** (2002), no. 4, 695–723.

[Jo48] John, F. Extremum problems with inequalities as subsidiary conditions. Studies and Essays Presented to R. Courant on his 60th Birthday, January 8, 1948, 187–204. Interscience Publishers, Inc., New York, N. Y., 1948.

[Kato83] Katok, A.: Entropy and closed geodesics, *Ergo. Th. Dynam. Sys.* **2** (1983), 339–365.

[Kato92] Katok, S.: Fuchsian groups. Chicago Lectures in Mathematics. University of Chicago Press, Chicago, IL, 1992.

[KatH95] Katok, A.; Hasselblatt, B.: Introduction to the modern theory of dynamical systems. With a supplementary chapter by Katok and L. Mendoza. *Encyclopedia of Mathematics and its Applications*, **54**. Cambridge University Press, Cambridge, 1995.

[Ka83] Katz, M.: The filling radius of two-point homogeneous spaces, *J. Diff. Geom.* **18** (1983), 505-511.

[Ka88] Katz, M.: The first diameter of 3-manifolds of positive scalar curvature. *Proc. Amer. Math. Soc.* **104** (1988), no. 2, 591–595.

[Ka02] Katz, M. Local calibration of mass and systolic geometry. *Geom. Funct. Anal. (GAFA)* **12**, issue 3 (2002), 598-621. See arXiv:math.DG/0204182

[Ka03] Katz, M.: Four-manifold systoles and surjectivity of period map, *Comment. Math. Helv.* **78** (2003), 772-786. See arXiv:math.DG/0302306

[Ka05] Katz, M.: Systolic inequalities and Massey products in simply-connected manifolds, *Geometriae Dedicata*, to appear.

[Ka07] Katz, M.: Systolic geometry and topology. With an appendix by Jake P. Solomon. *Mathematical Surveys and Monographs*, **137**. American Mathematical Society, Providence, RI, 2007.

[KL04] Katz, M.; Lescop, C.: Filling area conjecture, optimal systolic inequalities, and the fiber class in abelian covers. Proceedings of conference and workshop in memory of R. Brooks, held at the Technion, *Israel Mathematical Conference Proceedings (IMCP), Contemporary Math.*, Amer. Math. Soc., Providence, R.I. (to appear). See `arXiv:math.DG/0412011`

[KR04] Katz, M.; Rudyak, Y.: Lusternik-Schnirelmann category and systolic category of low dimensional manifolds. *Communications on Pure and Applied Mathematics*, to appear. See `arXiv:math.DG/0410456`

[KR05] Katz, M.; Rudyak, Y.: Bounding volume by systoles of 3-manifolds. See `arXiv:math.DG/0504008`

[KatzRS06] Katz, M.; Rudyak, Y.; Sabourau, S.: Systoles of 2-complexes, Reeb graph, and Grushko decomposition. *International Math. Research Notices* 2006 (2006). Art. ID 54936, pp. 1–30. See `arXiv:math.DG/0602009`

[KS04] Katz, M.; Sabourau, S.: Hyperelliptic surfaces are Loewner, *Proc. Amer. Math. Soc.*, to appear. See `arXiv:math.DG/0407009`

[KS05] Katz, M.; Sabourau, S.: Entropy of systolically extremal surfaces and asymptotic bounds, *Ergodic Theory and Dynamical Systems*, **25** (2005). See `arXiv:math.DG/0410312`

[KS06] Katz, M.; Sabourau, S.: An optimal systolic inequality for CAT(0) metrics in genus two, preprint. See `arXiv:math.DG/0501017`

[KSV05] Katz, M.; Schaps, M.; Vishne, U.: Logarithmic growth of systole of arithmetic Riemann surfaces along congruence subgroups. See `arXiv:math.DG/0505007`

[Ke74] Keisler, H. J.: Elementary calculus: An infinitesimal approach. Available online at http://www.math.wisc.edu/∼keisler/calc.html

[Kl82] Klein, F.: Über Riemanns Theorie der algebraischen Funktionen und ihrer Integrale. – Eine Ergänzung der gewöhnlichen Darstellungen. Leipzig: B.G. Teubner 1882.

[Ko87] Kodani, S.: On two-dimensional isosystolic inequalities, *Kodai Math. J.* **10** (1987), no. 3, 314–327.

[Kon03] Kong, J.: Seshadri constants on Jacobian of curves. *Trans. Amer. Math. Soc.* **355** (2003), no. 8, 3175–3180.

[Ku78] Kung, H. T.; Leiserson, C. E.: Systolic arrays (for VLSI). Sparse Matrix Proceedings 1978 (Sympos. Sparse Matrix Comput., Knoxville, Tenn., 1978), pp. 256–282, SIAM, Philadelphia, Pa., 1979.

[LaLS90] Lagarias, J.C.; Lenstra, H.W., Jr.; Schnorr, C.P.: Bounds for Korkin-Zolotarev reduced bases and successive minima of a lattice and its reciprocal lattice. *Combinatorica* **10** (1990), 343–358.

[La75] Lawson, H.B. The stable homology of a flat torus. *Math. Scand.* **36** (1975), 49–73.

[Leib] Leibniz, G. W.: La naissance du calcul différentiel. (French) [The birth of differential calculus] 26 articles des Acta Eruditorum. [26 articles of the Acta Eruditorum] Translated from the Latin and with an introduction and notes by Marc Parmentier. With a preface by Michel Serres. Mathesis. Librairie Philosophique J. Vrin, Paris, 1989.

[Li69] Lichnerowicz, A.: Applications harmoniques dans un tore. *C.R. Acad. Sci., Sér. A* **2**69, 912–916 (1969).

[Lo71] Loewner, C.: Theory of continuous groups. Notes by H. Flanders and M. Protter. *Mathematicians of Our Time* **1**, The MIT Press, Cambridge, Mass.-London, 1971.

[Ma69] Macbeath, A. M.: Generators of the linear fractional groups. 1969 Number Theory (Proc. Sympos. Pure Math., Vol. XII, Houston, Tex., 1967) pp. 14–32 Amer. Math. Soc., Providence, R.I.

[Mi95] Miranda, R.: Algebraic curves and Riemann surfaces. *Graduate Studies in Mathematics*, **5**. American Mathematical Society, Providence, RI, 1995.

[Ma79] Manning, A.: Topological entropy for geodesic flows, *Ann. of Math.* (2) **110** (1979), 567–573.

[Ma69] May, J.: Matric Massey products. *J. Algebra* **12** (1969) 533–568.

[MP77] Millman, R.; Parker, G. Elements of differential geometry. Prentice-Hall Inc., Englewood Cliffs, N. J., 1977.

[MS86] Milman, V. D.; Schechtman, G.: Asymptotic theory of finite-dimensional normed spaces. With an appendix by M. Gromov. *Lecture Notes in Mathematics*, **1200**. Springer-Verlag, Berlin, 1986.

[MilH73] Milnor, J.; Husemoller, D.: Symmetric bilinear forms. Springer, 1973.

[MiTW73] Misner, Charles W.; Thorne, Kip S.; Wheeler, John Archibald: Gravitation. W. H. Freeman and Co., San Francisco, Calif., 1973.

[Mo86] Moser, J. Minimal solutions of variational problems on a torus. *Ann. Inst. H. Poincaré-Analyse non linéaire* **3** (1986), 229–272.

[MS92] Moser, J.; Struwe, M. On a Liouville-type theorem for linear and nonlinear elliptic differential equations on a torus. *Bol. Soc. Brasil. Mat.* **23** (1992), 1–20.

[NR02] Nabutovsky, A.; Rotman, R.: The length of the shortest closed geodesic on a 2-dimansional sphere, *Int. Math. Res. Not.* **2002:23** (2002), 1211-1222.

[OR01] Oprea, J.; Rudyak, Y.: Detecting elements and Lusternik-Schnirelmann category of 3-manifolds. Lusternik-Schnirelmann category and related topics (South Hadley, MA, 2001), 181–191, *Contemp. Math.*, **316**, Amer. Math. Soc., Providence, RI, 2002. `arXiv:math.AT/0111181`

[Pa04] Parlier, H.: Fixed-point free involutions on Riemann surfaces. See `arXiv:math.DG/0504109`

[PP03] Paternain, G. P.; Petean, J.: Minimal entropy and collapsing with curvature bounded from below. *Invent. Math.* **151** (2003), no. 2, 415–450.

[Pu52] Pu, P.M.: Some inequalities in certain nonorientable Riemannian manifolds, *Pacific J. Math.* **2** (1952), 55–71.

[Ri1854] B. Riemann, Hypothèses qui servent de fondement à la géométrie, in Oeuvres mathématiques de Riemann. J. Gabay, 1990.

[RT90] Rosenbloom, M.; Tsfasman, M.: Multiplicative lattices in global fields. *Invent. Math.* **101** (1990), no. 3, 687–696.

[Ru87] Rudin, W.: Real and complex analysis. Third edition. McGraw-Hill Book Co., New York, 1987.

[Ru99] Rudyak, Y.: On category weight and its applications. *Topology* **38**, 1, (1999), 37–55.

[RT00] Rudyak, Y.; Tralle, A.: On Thom spaces, Massey products, and nonformal symplectic manifolds. *Internat. Math. Res. Notices* **10** (2000), 495–513.

[Sa04] Sabourau, S.: Systoles des surfaces plates singulières de genre deux, *Math. Zeitschrift* **247** (2004), no. 4, 693–709.

[Sa05] Sabourau, S.: Entropy and systoles on surfaces, preprint.

[Sa06] Sabourau, S.: Systolic volume and minimal entropy of aspherical manifolds, preprint.

[Sak88] Sakai, T.: A proof of the isosystolic inequality for the Klein bottle. *Proc. Amer. Math. Soc.* **104** (1988), no. 2, 589–590.

[SS03] Schmidt, Thomas A.; Smith, Katherine M.: Galois orbits of principal congruence Hecke curves. J. London Math. Soc. (2) 67 (2003), no. 3, 673–685.

[Sen91a] Senn, W. Strikte Konvexität für Variationsprobleme auf dem n-dimensionalen Torus. *manuscripta math.* **71** (1991), 45–65.

[Sen91b] Senn, W. Über Mosers regularisiertes Variationsproblem für minimale Blätterungen des n-dimensionalen Torus. *J. Appl. Math. Physics (ZAMP)* **42** (1991), 527–546.

[Ser73] Serre, J.-P.: A course in arithmetic. Translated from the French. Graduate Texts in Mathematics, No. 7. Springer-Verlag, New York-Heidelberg, 1973.

[Sik05] Sikorav, J.: Bounds on primitives of differential forms and cofilling inequalities. See `arXiv:math.DG/0501089`

[Sin78] Singhof, W.: Generalized higher order cohomology operations induced by the diagonal mapping. *Math. Z.* **162**, 1, (1978), 7–26.

[Sm62] Smale, S.: On the structure of 5-manifolds. *Ann. of Math.* (2) **75** (1962), 38–46.

[Sp79] Spivak, M.: A comprehensive introduction to differential geometry. Vol. 4. Second edition. Publish or Perish, Inc., Wilmington, Del., 1979.

[St00] Streit, M.: Field of definition and Galois orbits for the Macbeath-Hurwitz curves. Arch. Math. (Basel) 74 (2000), no. 5, 342–349.

[Tr93] Tralle, A.: On compact homogeneous spaces with nonvanishing Massey products, in *Differential geometry and its applications (Opava, 1992)*, pp. 47–50, Math. Publ. **1**, Silesian Univ. Opava, 1993.

[Tu47] Tutte, W.: A family of cubical graphs. *Proc. Cambridge Philos. Soc.* **43** (1947), 459–474.

[Wa83] Warner, F.W.: Foundations of Differentiable Manifolds and Lie Groups. *Graduate Texts in Mathematics*, **9**4. Springer-Verlag, New York-Heidelberg-Berlin, 1983.

[We55] Weatherburn, C. E.: Differential geometry of three dimensions. Vol. 1. Cambridge University Press, 1955.

[50] White, B.: Regularity of area-minimizing hypersurfaces at boundaries with multiplicity. Seminar on minimal submanifolds (E. Bombieri

ed.), 293-301. *Annals of Mathematics Studies* **103**. Princeton University Press, Princeton N.J., 1983.

[Wr74] Wright, Alden H.: Monotone mappings and degree one mappings between PL manifolds. Geometric topology (Proc. Conf., Park City, Utah, 1974), pp. 441–459. Lecture Notes in Math., Vol. 438, Springer, Berlin, 1975.

[Ya88] Yamaguchi, T.: Homotopy type finiteness theorems for certain precompact families of Riemannian manifolds, *Proc. Amer. Math. Soc.* **102** (1988), 660–666.