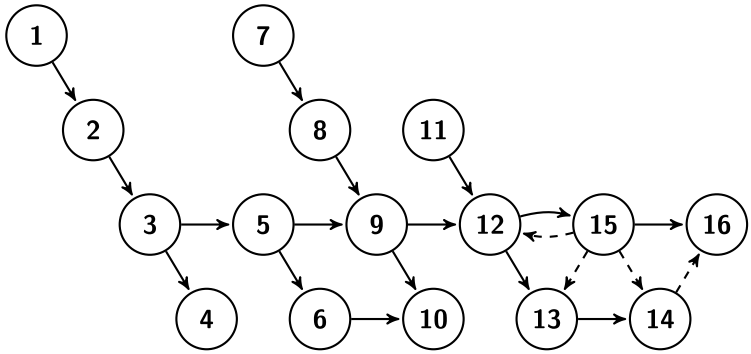
What is differential geometry: curves and surfaces

Anton Petrunin and
Sergio Zamora Barrera

Contents

CHAPTER DEPENDENCY GRAPH



These notes are designed for those who either plan to work in differential geometry, or at least want to have a good reason *not* to do it. It should be more than sufficient for a semester-long course.

Differential geometry exploits several branches of mathematics including real analysis, measure theory, calculus of variations, differential equations, elementary and convex geometry, topology, and more. This subject is wide even at the beginning. For that reason, it is fun and painful both to teach and to study.

In this book, we discuss smooth curves and surfaces — the main gate to differential geometry. This subject provides a collection of examples and ideas critical for further study. It is wise to become a master in this subject before making further steps — there is no need to rush.

We give a general overview of the subject, keeping it problem-centered, elementary, visual, and virtually rigorous; we allow gaps that belong to other branches of mathematics (mostly to the subjects discussed briefly in the preliminaries).

We focus on the techniques that are absolutely essential for further study. For that reason we omit a number of topics that traditionally included in the introductory texts; for example, we do not touch minimal surfaces and Peterson–Codazzi formulas. At the same time, we get to applications that are not in the scope of typical introductory texts.

The first example is the theorem of Vladimir Ionin and Herman Pestov about *the Moon in a puddle* (6.13). This theorem might be the simplest meaningful example of the so-called *local to global theorems* which lies in the heart of differential geometry; by that reason it is a good answer to the main question of this book — "What is differential geometry?".

Other examples include the theorem of Sergei Bernstein's on saddle graphs, and two theorems of Stephan Cohn-Vossen on existence of simple two-sided infinite geodesic on an open convex surface and the splitting theorem.

We extensively used two textbooks: by Aleksei Chernavskii [22] and by Victor Toponogov [77]. Both of these books are based on extensive teaching experience. These notes are based on the lectures given at the MASS program (Mathematics Advanced Study Semesters at Pennsylvania State University) Fall semester 2018. We want to thank the students in our class, Yurii Burago, and Nina Lebedeva for help. A large number of these topics were presented by Yurii Burago in his lectures teaching the first author at the Leningrad University.

Anton Petrunin and
Sergio Zamora Barrera.

Chapter 0

Preliminaries

This chapter should be used as a quick reference while reading the rest of the book; it also contains all necessary references with complete proof.

The fist section on metric spaces is an exception; we suggest to read in before going further.

A Metric spaces

We assume that the reader is familiar with the notion of distance in the Euclidean space. In this chapter we briefly discuss its generalization and fix notations that will be used further.

The introductory part of the book by Dmitri Burago, Yuri Burago, and Sergei Ivanov [13] contains all the needed material.

Definitions

Metric is a function that returns a real value $\mathrm{dist}(x,y)$ for any pair x, y in a given nonempty set $\mathcal{X}$ and satisfies the following axioms for any triple x, y, z:

(a) Positiveness:
$$\mathrm{dist}(x,y) \geqslant 0.$$

(b) $x = y$ if and only if
$$\mathrm{dist}(x,y) = 0.$$

(c) Symmetry:
$$\mathrm{dist}(x,y) = \mathrm{dist}(y,x).$$

(d) Triangle inequality:
$$\mathrm{dist}(x,z) \leqslant \mathrm{dist}(x,y) + \mathrm{dist}(y,z).$$

A set with a metric is called *metric space* and the elements of the set are called *points*.

Shortcut for distance

Usually we consider only one metric on a set, therefore we can denote the metric space and its underlying set by the same letter, say $\mathcal{X}$. In this case we also use the shortcut notations $|x-y|$ or $|x-y|_{\mathcal{X}}$ for the *distance* $\mathrm{dist}(x,y)$ *from x to y in $\mathcal{X}$*. For example, the triangle inequality can be written as

$$|x-z|_{\mathcal{X}} \leqslant |x-y|_{\mathcal{X}} + |y-z|_{\mathcal{X}}.$$

The Euclidean space and plane as well as the real line will be the most important examples of metric spaces for us. In these examples the introduced notation $|x-y|$ for the distance from x to y has perfect sense as the norm of the vector $x-y$. However, in a general metric space the expression $x-y$ has no meaning, but we use expression $|x-y|$ for the distance anyway.

More examples

Usually, if we say *plane* or *space* we mean the *Euclidean* plane or space. However the plane (as well as the space) admits many other metrics, for example the so-called *Manhattan metric* from the following exercise.

0.1. Exercise. *Consider the function*

$$\mathrm{dist}(p,q) = |x_1 - x_2| + |y_1 - y_2|,$$

where $p=(x_1,y_1)$ and $q=(x_2,y_2)$ are points in the coordinate plane $\mathbb{R}^2$. Show that dist *is a metric on $\mathbb{R}^2$.*

Another example: the *discrete space* — an arbitrary nonempty set $\mathcal{X}$ with the metric defined as $|x-y|_{\mathcal{X}}=0$ if $x=y$ and $|x-y|_{\mathcal{X}}=1$ otherwise.

Subspaces

Any subset of a metric space is also a metric space, by restricting the original metric to the subset; the obtained metric space is called a *subspace*. In particular, all subsets of the Euclidean space are metric spaces.

Balls

Given a point p in a metric space $\mathcal{X}$ and a real number $R\geqslant 0$, the set of points x on the distance less then R (at most R) from p is called the *open* (respectively *closed*) *ball* of radius R with center p. The *open ball* is denoted as $B(p, R)$ or $B(p, R)_{\mathcal{X}}$; the second notation is used if we need to emphasize that the ball lies in the metric space $\mathcal{X}$. Formally speaking

$$B(p, R) = B(p, R)_{\mathcal{X}} = \{\, x \in \mathcal{X} \, : \, |x - p|_{\mathcal{X}} < R \,\}.$$

Analogously, the *closed ball* is denoted as $\bar{B}[p, R]$ or $\bar{B}[p, R]_{\mathcal{X}}$ and

$$\bar{B}[p, R] = \bar{B}[p, R]_{\mathcal{X}} = \{\, x \in \mathcal{X} \, : \, |x - p|_{\mathcal{X}} \leqslant R \,\}.$$

0.2. Exercise. *Let $\mathcal{X}$ be a metric space.*

(a) *Show that if $\bar{B}[p, 2] \subset \bar{B}[q, 1]$ for some points $p, q \in \mathcal{X}$, then $\bar{B}[p, 2] = \bar{B}[q, 1]$.*

(b) *Construct a metric space $\mathcal{X}$ with two points p and q such that the strict inclusion $B(p, \frac{3}{2}) \subset B(q, 1)$ holds.*

Continuity

0.3. Definition. *Let $\mathcal{X}$ be a metric space. A sequence of points $x_1, x_2, \ldots$ in $\mathcal{X}$ is called* convergent *if there is $x_\infty \in \mathcal{X}$ such that $|x_\infty - x_n| \to 0$ as $n \to \infty$. That is, for every $\varepsilon > 0$, there is a natural number N such that for all $n \geqslant N$, we have*

$$|x_\infty - x_n|_{\mathcal{X}} < \varepsilon.$$

In this case we say that the sequence (x_n) converges *to x_∞, or x_∞ is the* limit *of the sequence (x_n). Notationally, we write $x_n \to x_\infty$ as $n \to \infty$ or $x_\infty = \lim_{n\to\infty} x_n$.*

0.4. Definition. *Let $\mathcal{X}$ and $\mathcal{Y}$ be metric spaces. A map $f \colon \mathcal{X} \to \mathcal{Y}$ is called* continuous *if for any convergent sequence $x_n \to x_\infty$ in $\mathcal{X}$, we have $f(x_n) \to f(x_\infty)$ in $\mathcal{Y}$.*

Equivalently, $f \colon \mathcal{X} \to \mathcal{Y}$ is continuous if for any $x \in \mathcal{X}$ and any $\varepsilon > 0$, there is $\delta > 0$ such that

$$|x - y|_{\mathcal{X}} < \delta \quad \textit{implies that} \quad |f(x) - f(y)|_{\mathcal{Y}} < \varepsilon.$$

0.5. Exercise. *Let $\mathcal{X}$ and $\mathcal{Y}$ be metric spaces $f \colon \mathcal{X} \to \mathcal{Y}$ is* distance non-expanding map*; that is,*

$$|f(x) - f(y)|_{\mathcal{Y}} \leqslant |x - y|_{\mathcal{X}}$$

for any $x, y \in \mathcal{X}$. Show that f is continuous.

Homeomorphisms

0.6. Definition. *Let $\mathcal{X}$ and $\mathcal{Y}$ be metric spaces. A continuous bijection $f\colon \mathcal{X} \to \mathcal{Y}$ is called a* homeomorphism *if its inverse $f^{-1}\colon \mathcal{Y} \to \mathcal{X}$ is also continuous.*

If there exists a homeomorphism $f\colon \mathcal{X} \to \mathcal{Y}$, we say that $\mathcal{X}$ is homeomorphic *to $\mathcal{Y}$, or $\mathcal{X}$ and $\mathcal{Y}$ are* homeomorphic.

If a metric space $\mathcal{X}$ is homeomorphic to a known space, for example plane, sphere, disc, circle and so on, we may also say that $\mathcal{X}$ is a *topological* plane, sphere, disc, circle and so on.

Closed and open sets

0.7. Definition. *A subset C of a metric space $\mathcal{X}$ is called* closed *if whenever a sequence (x_n) of points from C converges in $\mathcal{X}$, we have that $\lim_{n\to\infty} x_n \in C$.*

A set $\Omega \subset \mathcal{X}$ is called open *if for any $z \in \Omega$, there is $\varepsilon > 0$ such that $B(z,\varepsilon) \subset \Omega$.*

0.8. Exercise. *Let Q be a subset of a metric space $\mathcal{X}$. Show that Q is closed if and only if its complement $\Omega = \mathcal{X} \backslash Q$ is open.*

An open set Ω that contains a given point p is called a *neighborhood of p*. A closed subset C that contains p together with its neighborhood is called a *closed neighborhood of p*

B Elementary geometry

Internal angles

Polygon is defined as a compact set bounded by a closed polygonal line. Recall that the internal angle of a polygon P at a vertex v is defined as angular measure of the intersection of P with a small circle centered at v.

0.9. Theorem. *The sum of all the internal angles of a simple n-gon is $(n-2)\cdot\pi$.*

While this theorem is well known, it is not easy to find a reference with a proof without cheating. A clean proof was given by Gary Meisters [58]. It uses induction on n and is based on the following:

0.10. Claim. *Suppose P is an n-gon with $n \geqslant 4$. Then a diagonal of P lies completely in P.*

Angle monotonicity

The *measure* of angle with sides $[p,x]$ and $[p,y]$ will be denoted by $\measuredangle[p\,{}^{x}_{y}]$; it takes a value in the interval $[0,\pi]$.

The following lemma is very simple and very useful. It says that the angle of a triangle monotonically depends on the opposite side, assuming we keep the other two sides fixed. It follows directly from the cosine rule.

0.11. Monotonicity lemma. *Let x, y, z, x^*, y^* and z^* be 6 points such that $|x-y| = |x^*-y^*| > 0$ and $|y-z| = |y^*-z^*| > 0$. Then*

$$\measuredangle[y\,{}^{x}_{z}] \geqslant \measuredangle[y^*\,{}^{x^*}_{z^*}] \quad \textit{if and only if} \quad |x-z| \geqslant |x^*-z^*|.$$

Spherical triangle inequality

The following theorem says that the triangle inequality holds for angles between half-lines from a fixed point. In particular it implies that a sphere with the angle metric is a metric space.

0.12. Theorem. *The following inequality holds for any three line segments $[o,a]$, $[o,b]$ and $[o,c]$ in the Euclidean space:*

$$\measuredangle[o\,{}^{a}_{b}] + \measuredangle[o\,{}^{b}_{c}] \geqslant \measuredangle[o\,{}^{a}_{c}]$$

Most of authors use this theorem without mentioning, but the proof is not that simple. A short elementary proof can be found in the classical textbook in Euclidean geometry by Andrey Kiselyov [45, §47].

Area of spherical triangle

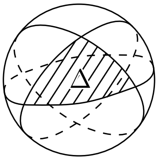
0.13. Lemma. *Let Δ be a spherical triangle; that is, Δ is the intersection of three closed half-spheres in the unit sphere $\mathbb{S}^2$. Then*

❶ $$\operatorname{area}\Delta = \alpha + \beta + \gamma - \pi,$$

where α, β and γ are the angles of Δ.

The value $\alpha+\beta+\gamma-\pi$ is called *excess of the triangle* Δ, so the lemma says that area of a spherical triangle equals its excess.

This lemma appears in many texts. We give its proof here since it is very important in our intuitive proof of Gauss–Bonnet formula.



Proof. Recall that

❷ $$\operatorname{area}\mathbb{S}^2 = 4{\cdot}\pi.$$

Note that the area of a spherical slice S_α between two meridians meeting at angle α is proportional to α. Since for S_π is a half-sphere, from ❷, we get $\operatorname{area} S_\pi = 2{\cdot}\pi$. Therefore the coefficient is 2; that is,

❸ $$\operatorname{area} S_\alpha = 2{\cdot}\alpha.$$

Extending the sides of Δ we get 6 slices: two S_α, two S_β and two S_γ which cover most of the sphere once, but the triangle Δ and its centrally symmetric copy Δ^* are covered 3 times. It follows that

$$2{\cdot}\operatorname{area} S_\alpha + 2{\cdot}\operatorname{area} S_\beta + 2{\cdot}\operatorname{area} S_\gamma = \operatorname{area}\mathbb{S}^2 + 4{\cdot}\operatorname{area}\Delta.$$

Substituting ❷ and ❸ and simplifying, we get ❶. □

C Convex geometry

A set X in the Euclidean space is called *convex* if for any two points $x, y \in X$, any point z between x and y lies in X. It is called *strictly convex* if for any two points $x, y \in X$, any point z between x and y lies in the interior of X.

From the definition, it is easy to see that the intersection of an arbitrary family of convex sets is convex. The intersection of all convex sets containing X is called the *convex hull* of X; it is the minimal convex set containing the set X.

We will use the following corollary of the so-called *hyperplane separation theorem*:

0.14. Lemma. *Let $K \subset \mathbb{R}^3$ be a closed convex set. Then for any point $p \notin K$ there is a plane Π that separates K from p; that is, K and p lie on opposite open half-spaces separated by Π.*

These definitions and hyperplane separation should appear on fist few pages of any introductory text in convex geometry; see for example the book of Roger Webster [82].

D Linear algebra

The following theorem can be found in any textbook in linear algebra; the book of Sergei Treil [80] will do.

0.15. Spectral theorem. *Any symmetric matrix is diagonalizable by an orthogonal matrix.*

We will use this theorem only for $2{\times}2$ matrices. In this case it can be restated as follows: Consider a function

$$f(x,y) = \begin{pmatrix} x & y \end{pmatrix} \cdot \begin{pmatrix} \ell & m \\ m & n \end{pmatrix} \cdot \begin{pmatrix} x \\ y \end{pmatrix} = \ell{\cdot}x^2 + 2{\cdot}m{\cdot}x{\cdot}y + n{\cdot}y^2,$$

that is defined on a (x,y)-coordinate plane. Then after proper rotation of the coordinates, the expression for f in the new coordinates will be

$$\bar f(x,y) = \begin{pmatrix} x & y \end{pmatrix} \cdot \begin{pmatrix} k_1 & 0 \\ 0 & k_2 \end{pmatrix} \cdot \begin{pmatrix} x \\ y \end{pmatrix} = k_1{\cdot}x^2 + k_2{\cdot}y^2.$$

E Analysis

The following material is discussed in any course of real analysis, the classical book of Walter Rudin [71] is one of our favorites.

Lipschitz condition

Recall that a function f between metric spaces is called *Lipschitz* if there is a constant L such that

$$|f(x) - f(y)| \leqslant L{\cdot}|x - y|$$

for all values x and y in the domain of definition of f.

The following theorem makes it possible to extend a number of results about smooth functions to Lipschitz functions. Recall that *almost all* means all values, with the possible exceptions in a set of zero *Lebesgue measure.*

0.16. Rademacher's theorem. *Let $f\colon [a,b] \to \mathbb{R}$ be a Lipschitz function. Then the derivative f' of f is a bounded measurable function defined almost everywhere in $[a,b]$ and it satisfies the fundamental theorem of calculus; that is, the following identity*

$$f(b) - f(a) = \int\limits_a^b f'(x){\cdot}dx,$$

holds if the integral is understood in the sense of Lebesgue.

The following theorem makes it possible to extend many statements about continuous function to measurable functions.

0.17. Lusin's theorem. *Let $\varphi\colon [a,b] \to \mathbb{R}$ be a measurable function. Then for any $\varepsilon > 0$, there is a continuous function $\psi_\varepsilon\colon [a,b] \to \mathbb{R}$ that coincides with φ outside of a set of measure at most ε. Moreover, if φ is bounded above and/or below by some constants, then we may assume that so is ψ_ε.*

Uniform continuity and convergence

Let $f\colon \mathcal{X} \to \mathcal{Y}$ be a map between metric spaces. If for any $\varepsilon > 0$ there is $\delta > 0$ such that

$$|x_1 - x_2|_{\mathcal{X}} < \delta \quad \Longrightarrow \quad |f(x_1) - f(x_2)|_{\mathcal{Y}} < \varepsilon,$$

then f is called *uniformly continuous.*

Evidently every uniformly continuous function is continuous; the converse does not hold. For example, the function $f(x) = x^2$ is continuous, but not uniformly continuous. However the following statement holds true:

0.18. Heine–Cantor theorem. *Any continuous function defined on a compact metric space is uniformly continuous.*

If the condition above holds for any function f_n in a sequence and δ depends solely on ε, then the sequence (f_n) is called *uniformly equicontinuous.* More precisely, a sequence of functions $f_n : \mathcal{X} \to \mathcal{Y}$ is called *uniformly equicontinuous* if for any $\varepsilon > 0$ there is $\delta > 0$ such that

$$|x_1 - x_2|_{\mathcal{X}} < \delta \quad \Longrightarrow \quad |f_n(x_1) - f_n(x_2)|_{\mathcal{Y}} < \varepsilon$$

for any n.

We say that a sequence of functions $f_i : \mathcal{X} \to \mathcal{Y}$ *converges uniformly* to a function $f_\infty : \mathcal{X} \to \mathcal{Y}$ if for any $\varepsilon > 0$, there is a natural number N such that for all $n \geqslant N$, we have $|f_\infty(x) - f_n(x)| < \varepsilon$ for all $x \in \mathcal{X}$.

0.19. Arzelá–Ascoli Theorem. *Suppose $\mathcal{X}$ and $\mathcal{Y}$ are compact metric spaces. Then any uniformly equicontinuous sequence of function $f_n\colon \mathcal{X} \to \mathcal{Y}$ has a subsequence that converges uniformly to a continuous function $f_\infty\colon \mathcal{X} \to \mathcal{Y}$.*

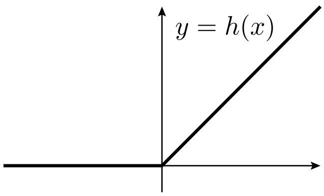
Cutoffs and mollifiers

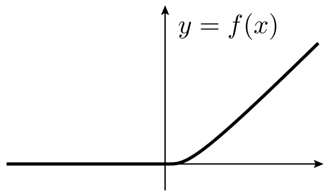
Here we construct examples of smooth functions that mimic behavior of some model functions. These functions are used to smooth model objects keeping its shape nearly unchanged.

For example, consider the following functions

$$h(t) = \begin{cases} 0 & \text{if } t \leqslant 0, \\ t & \text{if } t > 0. \end{cases} \qquad f(t) = \begin{cases} 0 & \text{if } t \leqslant 0, \\ \frac{t}{e^{1/t}} & \text{if } t > 0. \end{cases}$$

Note that h and f behave alike — both vanish at $t \leqslant 0$ and grow to



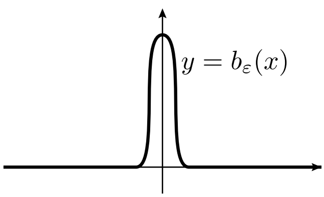


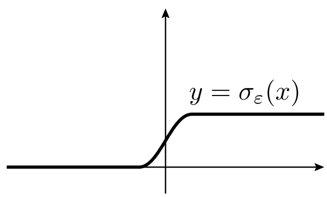
infinity for positive t. The function h is not smooth — its derivative at 0 is undefined. Unlike h, the function f is smooth. Indeed, the existence of all derivatives $f^{(n)}(x)$ at $x \neq 0$ is evident and direct calculations show that $f^{(n)}(0) = 0$ for all n.

Other useful examples of that type are the so-called *bell function* — a smooth function that is positive in an ε-neighborhood of zero and vanishing outside this neighborhood. An example of such function can be constructed based using the function f constructed above, say

$$b_\varepsilon(t) = c\cdot f(\varepsilon^2 - t^2);$$

typically one chooses the constant c so that $\int b_\varepsilon = 1$.





Another useful example is a sigmoid — nondecreasing function that vanish for $t \leqslant -\varepsilon$ and takes value 1 for any $t \geqslant \varepsilon$. For example the following function

$$\sigma_\varepsilon(t) = \int\limits_{-\infty}^{t} b_\varepsilon(x)\cdot dx.$$

F Multivariable calculus

The following material is discussed in any course of multivariable calculus, the classical book of Walter Rudin [71] is one of our favorites.

Regular value

A map $\boldsymbol{f}\colon \mathbb{R}^m \to \mathbb{R}^n$ can be thaut of as an array of functions

$$f_1, \dots, f_n\colon \mathbb{R}^m \to \mathbb{R}.$$

The map $\boldsymbol{f}$ is called *smooth* if each function f_i is smooth; that is, all partial derivatives of f_i are defined in the domain of definition of $\boldsymbol{f}$.

The Jacobian matrix of $\boldsymbol{f}$ at $\boldsymbol{x} \in \mathbb{R}^m$ is defined as

$$\mathrm{Jac}_{\boldsymbol{x}}\, \boldsymbol{f} = \begin{pmatrix} \dfrac{\partial f_1}{\partial x_1} & \cdots & \dfrac{\partial f_1}{\partial x_m} \\ \vdots & \ddots & \vdots \\ \dfrac{\partial f_n}{\partial x_1} & \cdots & \dfrac{\partial f_n}{\partial x_m} \end{pmatrix};$$

we assume that the right hand side is evaluated at $\boldsymbol{x} = (x_1, \dots, x_m)$.

If the Jacobian matrix defines a surjective linear map $\mathbb{R}^m \to \mathbb{R}^n$ (that is, if $\mathrm{rank}(\mathrm{Jac}_{\boldsymbol{x}}\, \boldsymbol{f}) = n$) then we say that $\boldsymbol{x}$ is a *regular point* of $\boldsymbol{f}$.

If for some $\boldsymbol{y} \in \mathbb{R}^n$ each point $\boldsymbol{x}$ such that $\boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{y}$ is regular, then we say that $\boldsymbol{y}$ is a *regular value* of $\boldsymbol{f}$. The following lemma states that *most* values of a smooth map are regular.

0.20. Sard's lemma. *Given a smooth map $\boldsymbol{f}\colon \Omega \to \mathbb{R}^n$ defined on an open set $\Omega \subset \mathbb{R}^m$, almost all values in $\mathbb{R}^n$ are regular.*

The words *almost all* mean all values, with the possible exceptions belonging to a set with vanishing *Lebesgue measure*. In particular if one chooses a random value equidistributed in an arbitrarily small ball $B \subset \subset \mathbb{R}^n$, then it is a regular value of $\boldsymbol{f}$ with probability 1.

Note that if $m < n$, then any point $\boldsymbol{y} = \boldsymbol{f}(\boldsymbol{x})$ is not a regular value of $\boldsymbol{f}$. Therefore the only regular value of $\boldsymbol{f}$ are the points in the complement of the image $\Im \boldsymbol{f}$. In this case, the theorem states that almost all points in $\mathbb{R}^n$, do *not* belong to $\Im \boldsymbol{f}$.

Inverse function theorem

The *inverse function theorem* gives a sufficient condition for a smooth map $\boldsymbol{f}$ to be invertible in a neighborhood of a given point $\boldsymbol{x}$. The condition is formulated in terms of the Jacobian matrix of $\boldsymbol{f}$ at $\boldsymbol{x}$.

The *implicit function theorem* is a close relative to the inverse function theorem; in fact it can be obtained as its corollary. It is used when we need to pass from parametric to implicit description of curves and surfaces.

Both theorems reduce the existence of a map satisfying a certain equation to a question in linear algebra. We use these two theorems only for $n \leqslant 3$.

0.21. Inverse function theorem. *Let $\boldsymbol{f} = (f_1, \dots, f_n)\colon \Omega \to \mathbb{R}^n$ be a smooth map defined on an open set $\Omega \subset \mathbb{R}^n$. Assume that the Jacobian matrix $\operatorname{Jac}_{\boldsymbol{x}} \boldsymbol{f}$ is invertible at some point $\boldsymbol{x} \in \Omega$. Then there is a smooth map $\boldsymbol{h}\colon \Phi \to \mathbb{R}^n$ defined in an open neighborhood Φ of $\boldsymbol{y} = \boldsymbol{f}(\boldsymbol{x})$ that is a* local inverse of $\boldsymbol{f}$ at $\boldsymbol{x}$*; that is, there is a neighborhood $\Psi \ni \boldsymbol{x}$ such that $\boldsymbol{f}$ defines a bijection $\Psi \leftrightarrow \Phi$ and $\boldsymbol{h} \circ \boldsymbol{f}$ is an identity map on Ψ.*

Moreover if an Ω contains an ε-neighborhood of $\boldsymbol{x}$, and the first and second partial derivatives $\frac{\partial f_i}{\partial x_j}$, $\frac{\partial^2 f_i}{\partial x_j \partial x_k}$ are bounded by a constant C for all i, j, and k, then we can assume that Φ is a δ-neighborhood of $\boldsymbol{y}$, for some $\delta > 0$ that depends only on ε and C.

0.22. Implicit function theorem. *Let $\boldsymbol{f} = (f_1, \dots, f_n)\colon \Omega \to \mathbb{R}^n$ be a smooth map, defined on a open subset $\Omega \subset \mathbb{R}^{n+m}$, where $m, n \geqslant 1$. Let us consider $\mathbb{R}^{n+m}$ as a product space $\mathbb{R}^n \times \mathbb{R}^m$ with coordinates $x_1, \dots, x_n, y_1, \dots, y_m$. Consider the following matrix*

$$M = \begin{pmatrix} \dfrac{\partial f_1}{\partial x_1} & \dots & \dfrac{\partial f_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \dfrac{\partial f_n}{\partial x_1} & \dots & \dfrac{\partial f_n}{\partial x_n} \end{pmatrix}$$

formed by the first n columns of the Jacobian matrix. Assume M is invertible at some point $\boldsymbol{x} = (x_1, \dots, x_n, y_1, \dots y_m)$ in the domain of definition of $\boldsymbol{f}$ and $\boldsymbol{f}(\boldsymbol{x}) = 0$. Then there is a neighborhood $\Psi \ni \boldsymbol{x}$ and a smooth function $\boldsymbol{h}\colon \mathbb{R}^m \to \mathbb{R}^n$ defined in a neighborhood $\Phi \ni 0$ such that for any $(x_1, \dots, x_n, y_1, \dots y_m) \in \Omega$, the equality

$$\boldsymbol{f}(x_1, \dots, x_n, y_1, \dots y_m) = 0$$

holds if and only if

$$(x_1, \dots x_n) = \boldsymbol{h}(y_1, \dots y_m).$$

Multiple integral

Let $\boldsymbol{f}\colon \mathbb{R}^n \to \mathbb{R}^n$ is a smooth map (maybe partially defined).

$$\operatorname{jac}_{\boldsymbol{x}} \boldsymbol{f} := |\det[\operatorname{Jac}_{\boldsymbol{x}} \boldsymbol{f}]|;$$

that is, $\operatorname{jac}_{\boldsymbol{x}} \boldsymbol{f}$ is the absolute value of the determinant of the Jacobian matrix of $\boldsymbol{f}$ at $\boldsymbol{x}$.

The following theorem plays the role of a substitution rule for multiple variables.

Borel subsets are defined as the class of subsets that are generated from open sets by applying the following operations recursively: countable union, countable intersection, and complement. Since complement of a closed set is open and the other way around these sets can be also generated from all closed sets. This class of sets includes virtually all sets that naturally appear in geometry but does not include pathological examples that create problems with integration.

0.23. Theorem. *Let $h\colon K \to \mathbb{R}$ be a continuous function on a Borel subset $K \subset \mathbb{R}^n$. Assume $\boldsymbol{f}\colon K \to \mathbb{R}^n$ is an injective smooth map. Then*

$$\int\limits_{\boldsymbol{x}\in K}\!\!\cdots\!\int h(\boldsymbol{x})\cdot \operatorname{jac}_{\boldsymbol{x}} \boldsymbol{f} = \int\limits_{\boldsymbol{y}\in \boldsymbol{f}(K)}\!\!\cdots\!\int h\circ \boldsymbol{f}^{-1}(\boldsymbol{y}).$$

Convex functions

The following statements will be used only for $n \leqslant 3$.

Let $f\colon \mathbb{R}^n \to \mathbb{R}$ be a smooth function (maybe partially defined). Choose a vector $\mathrm{w} \in \mathbb{R}^n$. Given a point $p \in \mathbb{R}^n$ consider the function $\varphi(t) = f(p + t\cdot\mathrm{w})$. Then the *directional derivative* $(D_{\mathrm{w}} f)(p)$ of f at p with respect to vector w is defined by

$$(D_{\mathrm{w}} f)(p) = \varphi'(0).$$

Recall that a function f is called *convex* if its epigraph $z \geqslant f(\boldsymbol{x})$ is a convex set in $\mathbb{R}^n \times \mathbb{R}$.

0.24. Theorem. *A smooth function $f\colon K \to \mathbb{R}$ defined on a convex subset $K \subset \mathbb{R}^n$ is convex if and only if one of the following equivalent condition holds:*

(a) The second directional derivative of f at any point in the direction of any vector is nonnegative; that is,

$$(D^2_{\mathrm{w}} f)(p) \geqslant 0$$

for any $p \in K$ and $\mathrm{w} \in \mathbb{R}^n$.

(b) The so-called Jensen's inequality

$$f\left((1-t)\cdot x_0 + t\cdot x_1\right) \leqslant (1-t)\cdot f(x_0) + t\cdot f(x_1)$$

holds for any $x_0, x_1 \in K$ and $t \in [0,1]$.

(c) For any $x_0, x_1 \in K$, we have

$$f\left(\frac{x_0 + x_1}{2}\right) \leqslant \frac{f(x_0) + f(x_1)}{2}.$$

G Ordinary differential equations

The following material is discussed at the very beginning of any course of ordinary differential equations; the classical book of Vladimir Arnold [7] is one of our favorites.

Systems of first order

The following theorem guarantees existence and uniqueness of solutions of an initial value problem for a system of ordinary first order differential equations

$$\begin{cases} x_1' &= f_1(x_1, \dots, x_n, t), \\ &\vdots \\ x_n' &= f_n(x_1, \dots, x_n, t), \end{cases}$$

where each $t \mapsto x_i = x_i(t)$ is a real valued function defined on a real interval $\mathbb{I}$ and each f_i is a smooth function defined on an open subset $\Omega \subset \mathbb{R}^n \times \mathbb{R}$.

The array of functions $(f_1, \dots, f_n)$ can be packed into one vector-valued function $\boldsymbol{f}\colon \Omega \to \mathbb{R}^n$; the same way the array $(x_1, \dots, x_n)$ can be packed into a vector $\boldsymbol{x} \in \mathbb{R}^n$. Therefore the system can be rewritten as one vector equation

$$\boldsymbol{x}' = \boldsymbol{f}(\boldsymbol{x}, t).$$

0.25. Theorem. *Suppose $\mathbb{I}$ is a real interval and $\boldsymbol{f}\colon \Omega \to \mathbb{R}^n$ is a smooth function defined on an open subset $\Omega \subset \mathbb{R}^n \times \mathbb{R}$. Then for any initial data $\boldsymbol{x}(t_0) = \boldsymbol{u}$ such that $(\boldsymbol{u}, t) \in \Omega$ the differential equation*

$$\boldsymbol{x}' = \boldsymbol{f}(\boldsymbol{x}, t)$$

has a unique solution $t \mapsto \boldsymbol{x}(t)$ defined at a maximal interval $\mathbb{J}$ that contains t_0. Moreover

- *(a) if $\mathbb{J} \neq \mathbb{R}$ (that is, if an end a of $\mathbb{J}$ is finite) then $\boldsymbol{x}(t)$ does not have a limit point in Ω as $t \to a$;*
- *(b) the function $(\boldsymbol{u}, t_0, t) \mapsto \boldsymbol{x}(t)$ has open domain of definition in $\Omega \times \mathbb{R}$ and it is smooth in this domain.*

Higher order

Suppose we have an ordinary differential equation of order k

$$\boldsymbol{x}^{(k)} = \boldsymbol{f}(\boldsymbol{x}, \dots, \boldsymbol{x}^{(k-1)}, t),$$

where $\boldsymbol{x} = \boldsymbol{x}(t)$ is a function from a real interval to $\mathbb{R}^n$.

This equation can be rewritten as k first order equations as follows with $k-1$ new variables $\boldsymbol{y}_i = \boldsymbol{x}^{(i)}$:

$$\begin{cases} \boldsymbol{x}' & = \boldsymbol{y}_1 \\ \boldsymbol{y}_1' & = \boldsymbol{y}_2 \\ & \vdots \\ \boldsymbol{y}_{k-1}'(t) & = \boldsymbol{f}(\boldsymbol{x}, \boldsymbol{y}_1, \dots, \boldsymbol{y}_{k-1}, t), \end{cases}$$

Using this trick one can reduce a higher order ordinary differential equation to a first order equation. In particular we get local existence and uniqueness for solutions of higher order equations as in Theorem 0.25.

H Topology

The following material is covered in any introductory text to topology; one of our favorites is a textbook of Czes Kosniowski [47].

Compact sets

A subset K of a metric space is called *compact* if any sequence of points (x_n) in K has a subsequence that converges to a point x_∞ in K.

The following properties follow directly from the definition:

◇ A closed subset of a compact space is compact.

◇ A continuous image of a compact space is compact.

0.26. Heine–Borel theorem. *A subset of Euclidean space is compact if and only if it is closed and bounded.*

Homeomorphisms and embedding

A bijection $f\colon \mathcal{X} \to \mathcal{Y}$ between metric spaces is called *homeomorphism* if f and its inverse f^{-1} are continuous. If $f\colon \mathcal{X} \to \mathcal{Y}$ is a homeomorphism to its image $f(\mathcal{X}) \subset \mathcal{Y}$, then it is called an *embedding*.

The following theorem characterizes homeomorphisms between compact spaces.

0.27. Theorem. *A continuous bijection f between compact metric spaces has a continuous inverse. In particular, we have the following:*

(a) Any continuous bijection between compact metric spaces is a homeomorphism.

(b) Any continuous injection from compact metric spaces to another metric space is an embedding.

Jordan's theorem

The first part of the following theorem was proved by Camille Jordan, the second part is due to Arthur Schoenflies:

0.28. Theorem. *The complement of any closed simple plane curve γ has exactly two connected components.*

Moreover, there is a homeomorphism $h\colon \mathbb{R}^2 \to \mathbb{R}^2$ that maps the unit circle to γ. In particular γ bounds a topological disc.

This theorem is known for its simple formulation and quite hard proof. By now many proofs of this theorem are known. For the first statement, a very short proof based on a somewhat developed technique is given by Patrick Doyle [25], among elementary proofs, one of our favorites is the proof given by Aleksei Filippov [29].

We use mostly the smooth case of this theorem which is much simpler. An amusing proof of this case was given by Gregory Chambers and Yevgeny Liokumovich [19].

Connectedness

Recall that a continuous map α from the unit interval $[0,1]$ to a Euclidean space is called a *path*. If $p=\alpha(0)$ and $q=\alpha(1)$, then we say that α connects p to q.

A set X in the Euclidean space is called *path-connected* if any two points $x,y\in X$ can be connected by a path lying in X.

A set X in the Euclidean space is called *connected* if one cannot cover X with two disjoint open sets V and W such that both intersections $X\cap V$ and $X\cap W$ are nonempty.

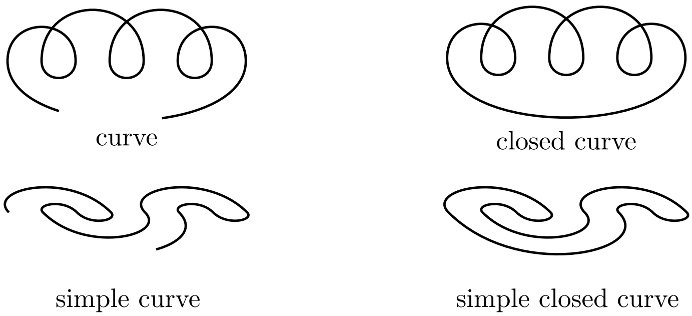
0.29. Proposition. *Any path-connected set is connected.*

Moreover, any open connected set in the Euclidean space or plane is path-connected.

Given a point $x\in X$, the maximal connected subset of X containing x is called the *connected component* of x in X.

Chapter 1

Definitions



A Simple curves

In the following definition we use the notion of *metric space* which is discussed in Section 0A. The Euclidean plane and space are the main examples of metric spaces that one should keep in mind.

Recall that a *real interval* is a connected subset of the real numbers.

Recall that a bijective continuous map $f\colon X \to Y$ between subsets of some metric spaces is called a *homeomorphism* if its inverse $f^{-1}\colon Y \to X$ is continuous.

1.1. Definition. *A connected subset γ in a metric space is called a* simple curve *if it is* locally *homeomorphic to a real interval.*

It turns out that any simple curve γ can be *parametrized* by a real interval or the circle. That is, there is a homeomorphism $\mathbb{G} \to \gamma$ where $\mathbb{G}$ is a real interval (open, closed or semi-open) or the circle

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

A complete proof of the latter statement is given by David Gale [33]. The proof is not hard, but it would take us away from the main subject. A finickity reader may add this property in the definition of curve.

If $\mathbb{G}$ is an open interval or a circle, we say that γ is a *curve without endpoints*, otherwise it is called a *curve with endpoints*. In the case when $\mathbb{G}$ is a circle we say that the curve is *closed*. When $\mathbb{G}$ is a closed interval, γ is called a *simple arc*.

A parametrization $\mathbb{G} \to \gamma$ describes a curve completely. We will denote a curve and its parametrization by the same letter; for example, we may say a plane curve γ is given with a parametrization $\gamma\colon (a, b) \to \mathbb{R}^2$. Note, however, that any simple curve admits many different parametrizations.

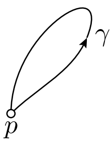
1.2. Exercise.

(a) *Show that the image of any continuous injective map* $\gamma\colon [0, 1] \to \mathbb{R}^2$ *is a simple arc.*

(b) *Find a continuous injective map* $\gamma\colon (0, 1) \to \mathbb{R}^2$ *such that its image is not a simple curve.*

B Parametrized curves

A *parametrized curve* is defined as a continuous map $\gamma\colon \mathbb{G} \to \mathcal{X}$ from a circle or a real interval (open, closed or semi-open) $\mathbb{G}$ to a metric space $\mathcal{X}$. For a parametrized curve we do not assume that γ is injective; in other words, a parametrized curve might have *self-intersections*.



If we say *curve* it means we do not want to specify whether it is a parametrized curve or a simple curve.

If the domain of a parametrized curve is a closed interval $[a, b]$, then the curve is called an *arc*. Further, if it is the unit interval $[0, 1]$, then it is also called a *path*. If in addition $p = \gamma(0) = \gamma(1)$, then γ is called a *loop*; in this case the point p is called the *base* of the loop.

Suppose $\mathbb{G}_1$ and $\mathbb{G}_2$ are either real intervals or circles. A continuous onto map $\tau\colon \mathbb{G}_1 \to \mathbb{G}_2$ is called *monotone* if for any $t \in \mathbb{G}_2$ the set $\tau^{-1}\{t\}$ is connected. Note that if $\mathbb{G}_1$ and $\mathbb{G}_2$ are intervals, then, by the intermediate value theorem, a monotone map is either nondecreaing or nonincreasing; that is, our definition agrees with the standard one when $\mathbb{G}_1$ and $\mathbb{G}_2$ are intervals.

Suppose that $\gamma_1\colon \mathbb{G}_1 \to \mathcal{X}$ and $\gamma_2\colon \mathbb{G}_2 \to \mathcal{X}$ are two parametrized curves such that $\gamma_1 = \gamma_2 \circ \tau$ for a monotone map $\tau\colon \mathbb{G}_1 \to \mathbb{G}_2$. Then we say that γ_2 is *reparametrization*[1] of γ_1 by τ.

[1] Note that in general, γ_1 is *not* a reparametrization of γ_2. In other words, according to our definition, the described relation *being a reparametrization* is not symmetric; in particular it is not an equivalence relation. Usually it is fixed by extending it to the

1.3. Advanced exercise. *Let X be a subset of the plane. Suppose that two distinct points $p, q \in X$ can be connected by a path in X. Show that there is a simple arc in X connecting p to q.*

C Smooth curves

Curves in the Euclidean space or plane are called *space curves* or, respectively, *plane curves.*

A parametrized space curve can be described by its coordinate functions

$$\gamma(t) = (x(t), y(t), z(t)).$$

Plane curves can be considered as a special case of space curves with $z(t) \equiv 0$.

Recall that a real-to-real function is called *smooth* if its derivatives of all orders are defined everywhere in the domain of definition. If each of the coordinate functions $x(t), y(t)$ and $z(t)$ of a space curve γ are smooth, then the parametrized curve is called *smooth.*

If the *velocity vector*

$$\gamma'(t) = (x'(t), y'(t), z'(t))$$

does not vanish at any point, then the parametrized curve γ is called *regular.*

A simple space curve is called *smooth* (respectively, *regular*) if it admits a smooth (respectively, regular) parametrization. Regular smooth curves are among the main objects in differential geometry; colloquially, the term *smooth curve* is often used as a shortcut for *smooth regular curves.*

1.4. Exercise. *Recall that (see Section 0E) that the following function is smooth:*

$$f(t) = \begin{cases} 0 & \textit{if } t \leqslant 0, \\ \frac{t}{e^{1/t}} & \textit{if } t > 0. \end{cases}$$

Show that $\alpha(t) = (f(t), f(-t))$ gives a smooth parametrization of a simple curve formed by the union of two half-axis in the plane.

Show that any smooth parametrization of this curve has a vanishing velocity vector at the origin. Conclude that this curve is not regular; that is, it does not admit a regular smooth parametrization.

1.5. Exercise. *Describe the set of real numbers ℓ such that the plane curve $\gamma_\ell(t) = (t + \ell\cdot \sin t, \ell\cdot \cos t)$, $t \in \mathbb{R}$ is*

(a) smooth; *(b) regular;* *(c) simple.*

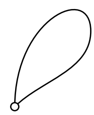
---

minimal equivalence relation that includes ours [13, 2.5.1]. But we will stick to our version.

D Periodic parametrizations

Any smooth regular closed curve can be described by a *periodic* parametrized curve $\gamma\colon \mathbb{R} \to \mathcal{X}$; that is, a curve such that $\gamma(t+\ell) = \gamma(t)$ for a fixed period $\ell > 0$ and all t. For example, the unit circle in the plane can be described by the $2{\cdot}\pi$-periodic parametrization $\gamma(t) = (\cos t, \sin t)$.

Any smooth regular closed curve can be also described by a smooth regular loop. But in general the closed curve described by a smooth regular loop might fail to be regular at its base; an example is shown on the diagram.



E Implicitly defined curves

Suppose $f\colon \mathbb{R}^2 \to \mathbb{R}$ is a smooth function; that is, all its partial derivatives are defined everywhere. Let $\gamma \subset \mathbb{R}^2$ be the set of solutions of the equation $f(x,y) = 0$.

Assume γ is connected. According to the implicit function theorem (0.22), the set γ is a smooth regular simple curve if 0 is a *regular!value* of f; that is, the gradient $\nabla_p f$ does not vanish at any point $p \in \gamma$. In other words, if $f(p) = 0$, then $f_x(p) \neq 0$ or $f_y(p) \neq 0$.[2]

The described condition is sufficient but *not necessary.* For example, zero is not a regular value the function $f(x,y) = y^2$, but the equation $f(x,y) = 0$ describes a smooth regular curve — the x-axis.

Similarly, assume f, h is a pair of smooth functions defined in $\mathbb{R}^3$. The system of equations

$$\begin{cases} f(x,y,z) = 0, \\ h(x,y,z) = 0. \end{cases}$$

defines a regular smooth space curve if the set γ of solutions is connected and 0 is a regular value of the map $F\colon \mathbb{R}^3 \to \mathbb{R}^2$ defined as

$$F\colon (x,y,z) \mapsto (f(x,y,z), h(x,y,z)).$$

It means that the gradients ∇f and ∇h are linearly independent at any point $p \in \gamma$. In other words, the Jacobian matrix

$$\operatorname{Jac}_p F = \begin{pmatrix} f_x & f_y & f_z \\ h_x & h_y & h_z \end{pmatrix}$$

for the map $F\colon \mathbb{R}^3 \to \mathbb{R}^2$ has rank 2 at any point $p \in \gamma$.

If a curve γ is described in such a way, then we say that it is *implicitly defined.* If a curve is defined by its parametrization, we say that it is *explicitly defined.*

[2] Here f_x is a shortcut notation for the partial derivative $\frac{\partial f}{\partial x}$.

The implicit function theorem guarantees the existence of regular smooth parametrizations for any implicitly defined curve. However, when it comes to calculations, it is usually easier to work directly with implicit representations.

1.6. Exercise. *Consider the set in the plane described by the equation* $y^2 = x^3$*. Is it a simple curve? Is it a smooth regular curve?*

1.7. Exercise. *Describe the set of real numbers* ℓ *such that the system of equations*

$$\begin{cases} x^2 + y^2 + z^2 &= 1 \\ x^2 + \ell{\cdot}x + y^2 &= 0 \end{cases}$$

describes a smooth regular curve.

F Proper curves

A parametrized curve γ in a metric space $\mathcal{X}$ is called *proper* if for any compact set $K \subset \mathcal{X}$, the inverse image $\gamma^{-1}(K)$ is compact.

For example, the curve $\gamma(t) = (e^t, 0, 0)$ defined on the real line is not proper. Indeed, the half-line $(-\infty, 0]$ is not compact, but it is the inverse image of the closed unit ball around the origin.

1.8. Exercise. *Suppose* $\gamma\colon \mathbb{R} \to \mathbb{R}^2$ *is a proper curve. Show that* $|\gamma(t)| \to \to \infty$ *as* $t \to \pm\infty$.

Recall that a closed interval is compact (0.26) and closed subsets of a compact set are compact; see Section 0H. It follows that closed curves and arcs are automatically proper; indeed, their parameter sets are compact.

A simple curve is called proper if it admits a proper parametrization.

1.9. Exercise. *Show that a simple space curve is proper if and only if its image is a closed set.*

A proper simple plane curve without endpoints is called *open*. The terms *open curve* and *closed curve* have nothing to do with open and closed sets.

1.10. Exercise. *Use the Jordan's theorem (0.28) to show that any simple open plane curve divides the plane in two connected components.*

Chapter 2

Length

A Definitions

Recall that a sequence

$$a = t_0 < t_1 < \cdots < t_k = b.$$

is called a *partition* of the interval $[a, b]$.

2.1. Definition. *Let $\gamma\colon [a, b] \to \mathcal{X}$ be a curve in a metric space. The* length *of γ is defined as*

$$\operatorname{length}\gamma = \sup\left\{\, |\gamma(t_0) - \gamma(t_1)|_{\mathcal{X}} + \cdots + |\gamma(t_{k-1}) - \gamma(t_k)|_{\mathcal{X}} \,\right\},$$

where the least upper bound is taken over all partitions (t_i) of $[a, b]$.

The length of γ is a nonnegative real number or infinity; the curve γ is called rectifiable *if its length is finite.*

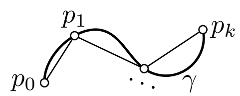
The length of a closed curve is defined as the length of the corresponding loop. If a curve is parameterized by an open or semi-open interval, then its length is defined as the least upper bound of the lengths of all its restrictions to closed intervals.

2.2. Exercise. *Suppose that curve $\gamma_1\colon [a_1, b_1] \to \mathbb{R}^3$ is a reparametrization of $\gamma_2\colon [a_2, b_2] \to \mathbb{R}^3$.*

$$\operatorname{length}\gamma_1 = \operatorname{length}\gamma_2.$$

2.3. Exercise. *Let $\alpha\colon [0, 1] \to \mathbb{R}^3$ be a simple path. Suppose a path $\beta\colon [0, 1] \to \mathbb{R}^3$ has the same image as α; that is, $\beta([0, 1]) = \alpha([0, 1])$. Show that*

$$\operatorname{length}\beta \geqslant \operatorname{length}\alpha.$$



Suppose $\gamma\colon [a,b] \to \mathbb{R}^3$ is a parametrized space curve. For a partition $a = t_0 < t_1 < \cdots < t_k = b$, set $p_i = \gamma(t_i)$. Then the polygonal line $p_0 \dots p_k$ is said to be *inscribed* in γ. If γ is closed, then $p_0 = p_k$, so the inscribed polygonal line is also closed.

Note that the length of a space curve γ can be defined as the least upper bound of the lengths of polygonal lines $p_0 \dots p_k$ inscribed in γ.

2.4. Exercise. *Assume $\gamma\colon [a,b] \to \mathbb{R}^3$ is a smooth curve. Show that*

(a) $\operatorname{length}\gamma \geqslant \int_a^b |\gamma'(t)|\cdot dt$; *(b)* $\operatorname{length}\gamma \leqslant \int_a^b |\gamma'(t)|\cdot dt$.

Conclude that

❶ $$\operatorname{length}\gamma = \int\limits_a^b |\gamma'(t)|\cdot dt.$$

2.5. Advanced exercises.

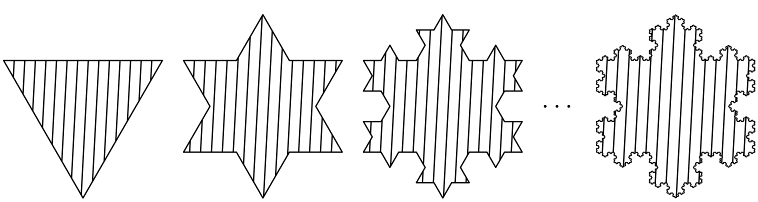
(a) *Show that the formula ❶ holds for any Lipschitz curve $\gamma\colon [a,b] \to \to \mathbb{R}^3$.*

(b) *Construct a non-constant curve $\gamma\colon [a,b] \to \mathbb{R}^3$ such that $\gamma'(t) = 0$ almost everywhere. (In particular, the formula ❶ does not hold for γ despite that both sides are defined.).*

B Nonrectifiable curves

Let us describe the so-called *Koch snowflake* — a classical example of a nonrectifiable curve.

Start with an equilateral triangle. For each side, divide it into three segments of equal length and then add an equilateral triangle with the middle segment as its base. Repeat this construction recursively with the obtained polygons. The Koch snowflake is the boundary of the union of



all the polygons. Three iterations and the resulting Koch snowflake are shown on the diagram.

2.6. Exercise.

(a) Show that the Koch snowflake is a simple closed curve; that is, it can be parametrized by a circle.
(b) Show that the Koch snowflake is not rectifiable.

C Arc-length parametrization

We say that a curve γ has an *arc-length parametrization* (also called *natural parametrization*) if

$$t_2 - t_1 = \operatorname{length} \gamma|_{[t_1, t_2]}$$

for any two parameter values $t_1 < t_2$; that is, the arc of γ from t_1 to t_2 has length $t_2 - t_1$.

Note that a smooth space curve $\gamma(t) = (x(t), y(t), z(t))$ is an arc-length parametrization if and only if it has unit velocity vector at all times; that is,

$$|\gamma'(t)| = \sqrt{x'(t)^2 + y'(t)^2 + z'(t)^2} = 1$$

for all t; by that reason smooth curves equipped with an arc-length parametrization are also called *unit-speed curves*. Observe that smooth unit-speed curves are automatically regular.

Note that any rectifiable curve γ can be reparametrized by its arc-length. Indeed, for a fixed value t_0,

$$s(t) = \begin{cases} \operatorname{length} \gamma|_{[t_0, t]} & \text{if } t \geqslant t_0, \\ -\operatorname{length} \gamma|_{[t, t_0]} & \text{if } t \leqslant t_0 \end{cases}$$

defines an arc-length parameter of the curve γ.

2.7. Proposition. *If $t \mapsto \gamma(t)$ is a smooth regular curve, then its arc-length parameterizaton is also smooth and regular. Moreover, the arc-length parameter s of γ can be written as an integral*

❶ $$s(t) = \int\limits_{t_0}^{t} |\gamma'(\tau)| \cdot d\tau.$$

Proof. The function $t \mapsto s(t)$ defined by ❶ is a smooth increasing function. Furthermore, by the fundamental theorem of calculus, $s'(t) = |\gamma'(t)|$. Since γ is regular, $s'(t) \neq 0$ for any parameter value t.

By the inverse function theorem (0.21) the inverse function $s^{-1}(t)$ is also smooth and $|(\gamma \circ s^{-1})'| \equiv 1$. Therefore $\gamma \circ s^{-1}$ is a unit-speed reparametrization of γ by s. By construction, $\gamma \circ s^{-1}$ is smooth, and since $|(\gamma \circ s^{-1})'| \equiv 1$, it is regular. □

Most of the time we use s for an arc-length parameter of a curve.

2.8. Exercise. *Reparametrize the helix*

$$\gamma_{a,b}(t) = (a\cdot \cos t, a\cdot \sin t, b\cdot t)$$

by its arc-length.

We will be interested in the properties of curves that are invariant under a reparametrization. Therefore we can always assume that any given smooth regular curve comes with an arc-length parametrization. A nice property of arc-length parametrizations is that they are almost canonical — these parametrizations differ only by a sign and an additive constant. By that reason, it is easier to express parametrization-independent quantities using arc-length parametrizations. This observation will be used in the definition of curvature and torsion.

On the other hand, it is usually impossible to find an explicit arc-length parametrization, which makes it hard to perform calculations; therefore it is often convenient to use the original parametrization.

D Convex curves

A simple plane curve is called *convex* if it bounds a convex region. Since the boundary of any region is closed, any convex curve is either closed or open (see Section 1F).

2.9. Proposition. *Assume a convex closed curve α lies inside the domain bounded by a simple closed plane curve β. Then*

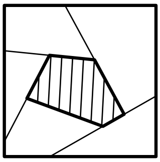
$$\operatorname{length}\alpha \leqslant \operatorname{length}\beta.$$

Note that in order to prove Proposition 2.9 it is sufficient to show that for any polygon P inscribed in α there is a polygon Q inscribed in β with equal or larger perimeter. Since any polygon P inscribed in α is convex, it is sufficient to prove the following lemma.

2.10. Lemma. *Let P and Q be polygons. Assume P is convex and $Q \supset P$. Then*

$$\operatorname{perim} P \leqslant \operatorname{perim} Q,$$

where $\operatorname{perim} P$ *denotes the perimeter of P.*



Proof. Note that by the triangle inequality, the inequality

$$\operatorname{perim} P \leqslant \operatorname{perim} Q$$

holds true if P can be obtained from Q by cutting it along a chord; that is, a line segment in Q that runs from boundary to boundary.

Observe that there is an increasing sequence of polygons

$$P = P_0 \subset P_1 \subset \cdots \subset P_n = Q$$

such that P_{i-1} obtained from P_i by cutting along a chord. Therefore

$$\begin{aligned}\operatorname{perim} P = \operatorname{perim} P_0 \leqslant \operatorname{perim} P_1 \leqslant \dots\\ \dots \leqslant \operatorname{perim} P_n = \operatorname{perim} Q\end{aligned}$$

and the lemma follows. □

Comment. Another proof of 2.10 can be obtained using closest point projection; see Section 11C.

2.11. Corollary. *Any convex closed plane curve is rectifiable.*

Proof. Any closed curve is bounded. Indeed, the curve can be described as an image of a loop $\alpha\colon [0,1] \to \mathbb{R}^2$, $\alpha(t) = (x(t), y(t))$. The coordinate functions $x(t)$ and $y(t)$ are continuous functions defined on $[0,1]$. This implies that the absolute values of both functions are bounded by some constant C. Therefore, α lies in the square defined by the inequalities $|x| \leqslant C$ and $|y| \leqslant C$.

By Proposition 2.9, the length of the curve cannot exceed $8{\cdot}C$ — the perimeter of the square. Hence the result. □

Recall that the convex hull of a set X is the smallest convex set that contains X; equivalently, the convex hull of X is the intersection of all convex sets containing X.

2.12. Exercise. *Let α be a simple closed plane curve. Denote by K the convex hull of α; let β be the boundary curve of K. Show that*

$$\operatorname{length} \alpha \geqslant \operatorname{length} \beta.$$

Try to show that the statement holds for arbitrary closed plane curves α, assuming only that K has nonempty interior.

E Crofton's formulas

For a function $f : \mathbb{S}^1 \to \mathbb{R}$, we will denote its average value as $\overline{f(\mathrm{u})}$. For a vector w, and a unit vector u, we will denote by w_{u} the orthogonal projection of w to the line in the direction of u; that is,

$$\mathrm{w}_{\mathrm{u}} = \langle \mathrm{u}, \mathrm{w} \rangle \cdot \mathrm{u}.$$

2.13. Theorem. *For any plane curve γ we have*

$$\operatorname{length} \gamma = \tfrac{\pi}{2} \cdot \overline{\operatorname{length} \gamma_{\mathrm{u}}},$$

where γ_{u} is the curve defined as $\gamma_{\mathrm{u}}(t) := (\gamma(t))_{\mathrm{u}}$.

Proof. Note that the magnitude of any plane vector w is proportional to the average magnitude of its projections; that is,

$$|\mathrm{w}| = k \cdot \overline{|\mathrm{w}_{\mathrm{u}}|}$$

for some $k \in \mathbb{R}$. The exact value of k can be found by integration[1], but we will find it in a different way. Note that for a smooth plane curve $\gamma \colon [a, b] \to \mathbb{R}^2$,

$$\gamma_{\mathrm{u}}'(t) = (\gamma'(t))_{\mathrm{u}} \text{ and } |\gamma_{\mathrm{u}}'(t)| = |\langle \mathrm{u}, \gamma'(t) \rangle|$$

for any $t \in [a, b]$. Then, according to Exercise 2.4,

$$\begin{aligned} \operatorname{length} \gamma &= \int_a^b |\gamma'(t)| \cdot dt = \\ &= \int_a^b k \cdot \overline{|\gamma_{\mathrm{u}}'(t)|} \cdot dt = \\ &= k \cdot \overline{\operatorname{length} \gamma_{\mathrm{u}}}. \end{aligned}$$

Since k is a universal constant, we can compute it by taking γ to be the unit circle. In this case

$$\operatorname{length} \gamma = 2 \cdot \pi.$$

Note that for any unit plane vector u, the curve γ_{u} runs back and forth along an interval of length 2. Then $\operatorname{length} \gamma_{\mathrm{u}} = 4$ for any u, and

$$\overline{\operatorname{length} \gamma_{\mathrm{u}}} = 4.$$

[1] It is the average value of $|\cos|$.

It follows that $2\cdot\pi = k\cdot 4$. Therefore Crofton's formula holds for arbitrary smooth curves.

Applying the same argument together with 2.5, we get that Crofton's formula holds for arbitrary rectifiable curves.

It remains to consider the nonrectifiable case; we have to show that

$$\operatorname{length}\gamma = \infty \quad\Longrightarrow\quad \overline{\operatorname{length}\gamma_{\mathrm{u}}} = \infty.$$

Observe that from the definition of length, we get

$$\operatorname{length}\gamma_{\mathrm{u}} + \operatorname{length}\gamma_{\mathrm{v}} \geqslant \operatorname{length}\gamma$$

for any plane curve γ and any pair (u,v) of orthonormal vectors in $\mathbb{R}^2$. Therefore, if γ has infinite length, then the average of lengths of γ_{u} is infinite as well. □

2.14. Exercise. *Show that any closed plane curve γ has length at least $\pi\cdot s$, where s is the average length of the projections of γ to lines. Moreover, equality holds if and only if γ is convex.*

Use this statement to give another solution to Exercise 2.12.

The following exercise gives analogous formulas in the Euclidean space.

As before, we denote by w_{u} the orthogonal projection of w to the line passing thru the origin with direction u. Further let us denote by $\mathrm{w}_{\mathrm{u}}^{\perp}$ the projection of w to the plane orthogonal to u; that is,

$$\mathrm{w}_{\mathrm{u}}^{\perp} = \mathrm{w} - \mathrm{w}_{\mathrm{u}}.$$

We will use notation $\overline{f(\mathrm{u})}$ for the average value of a function f defined on $\mathbb{S}^2$.

2.15. Advanced exercise. *Show that the length of a space curve is proportional to*

(a) the average length of its projections to all lines; that is,

$$\operatorname{length}\gamma = k\cdot\overline{\operatorname{length}\gamma_{\mathrm{u}}}$$

for some $k\in\mathbb{R}$.

(b) the average length of its projections to all planes; that is,

$$\operatorname{length}\gamma = k\cdot\overline{\operatorname{length}\gamma_{\mathrm{u}}^{\perp}}$$

for some $k\in\mathbb{R}$.

Find the value of k in each case.

F Semicontinuity of length

Recall that the lower limit of a sequence of real numbers (x_n) is denoted by

$$\varliminf_{n\to\infty} x_n.$$

It is defined as the lowest partial limit; that is, the lowest possible limit of a subsequence of (x_n). The lower limit is defined for any sequence of real numbers and it lies in the extended real line $[-\infty,\infty]$

2.16. Theorem. *Length is lower semicontinuous with respect to pointwise convergence of curves.*

More precisely, assume that a sequence of curves $\gamma_n\colon[a,b]\to\mathcal{X}$ in a metric space $\mathcal{X}$ converges pointwise to a curve $\gamma_\infty\colon[a,b]\to\mathcal{X}$; that is, for any fixed $t\in[a,b]$, we have $\gamma_n(t)\to\gamma_\infty(t)$ as $n\to\infty$. Then

❶ $$\varliminf_{n\to\infty}\operatorname{length}\gamma_n\geqslant\operatorname{length}\gamma_\infty.$$

Proof. Fix a partition $a=t_0<t_1<\dots<t_k=b$. Set

$$\begin{aligned}\Sigma_n&:=|\gamma_n(t_0)-\gamma_n(t_1)|+\dots+|\gamma_n(t_{k-1})-\gamma_n(t_k)|.\\ \Sigma_\infty&:=|\gamma_\infty(t_0)-\gamma_\infty(t_1)|+\dots+|\gamma_\infty(t_{k-1})-\gamma_\infty(t_k)|.\end{aligned}$$

Note that for each i we have

$$|\gamma_n(t_{i-1})-\gamma_n(t_i)|\to|\gamma_\infty(t_{i-1})-\gamma_\infty(t_i)|$$

and therefore

$$\Sigma_n\to\Sigma_\infty$$

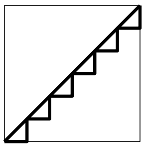
as $n\to\infty$. Note that

$$\Sigma_n\leqslant\operatorname{length}\gamma_n$$

for each n. Hence

$$\varliminf_{n\to\infty}\operatorname{length}\gamma_n\geqslant\Sigma_\infty.$$

Since the partition was arbitrary, by the definition of length, the inequality ❶ is obtained. □



The inequality ❶ might be strict. For example, the diagonal γ_∞ of the unit square can be approximated by stairs-like polygonal curves γ_n with sides parallel to the sides of the square (γ_6 is in the picture). In this case

$$\operatorname{length}\gamma_\infty=\sqrt2\quad\text{and}\quad\operatorname{length}\gamma_n=2$$

for any n.

G Length metric

Let $\mathcal{X}$ be a metric space. Given two points x, y in $\mathcal{X}$, denote by $\ell(x, y)$ the greatest lower bound of lengths of all paths connecting x to y; if there is no such path, then $\ell(x, y) = \infty$.

It is straightforward to see that the function ℓ satisfies all the axioms of a metric except it might take infinite values. Therefore if any two points in $\mathcal{X}$ can be connected by a rectifiable curve, then ℓ defines a new metric on $\mathcal{X}$; in this case ℓ is called the *induced length-metric*.

Evidently $\ell(x, y) \geqslant |x - y|$ for any pair of points $x, y \in \mathcal{X}$. If the equality holds for all pairs, then the metric $|{*} - {*}|$ is said to be a *length-metric* and the corresponding metric space is called a *length-metric space*.

Most of the time we consider length-metric spaces. In particular the Euclidean space is a length-metric space. A subspace A of a length-metric space $\mathcal{X}$ is not necessarily a length-metric space; the induced length distance between points x and y in the subspace A will be denoted as $|x - y|_A$; that is, $|x - y|_A$ is the greatest lower bound of the lengths of paths in A from x to y.

2.17. Exercise. *Let $A \subset \mathbb{R}^3$ be a closed subset. Show that A is convex if and only if*

$$|x - y|_A = |x - y|_{\mathbb{R}^3}$$

for any $x, y \in A$

H Spherical curves

Let us denote by $\mathbb{S}^2$ the unit sphere in the space; that is,

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

A space curve γ is called *spherical* if it runs in the unit sphere; that is, $|\gamma(t)| = 1$, or equivalently, $\gamma(t) \in \mathbb{S}^2$ for any t.

Recall that $\measuredangle(u, v)$ denotes the angle between two vectors u and v.

2.18. Observation. *For any $u, v \in \mathbb{S}^2$, we have*

$$|u - v|_{\mathbb{S}^2} = \measuredangle(u, v)$$

Proof. The short arc γ of a great circle from u to v in $\mathbb{S}^2$ has length $\measuredangle(u, v)$. Therefore

$$|u - v|_{\mathbb{S}^2} \leqslant \measuredangle(u, v).$$

It remains to prove the opposite inequality. In other words, we need to show that given a polygonal line $\beta = p_0 \dots p_n$ inscribed in γ there is a polygonal line $\beta_1 = q_0 \dots q_n$ inscribed in any given spherical path γ_1 connecting u to v such that

$$\text{❶} \qquad \operatorname{length}\beta_1 \geqslant \operatorname{length}\beta.$$

Define q_i as the first point on γ_1 such that $|u - p_i| = |u - q_i|$, but set $q_n = v$. Clearly β_1 is inscribed in γ_1 and according to the triangle inequality for angles (0.12), we have that

$$\measuredangle(q_{i-1}, q_i) \geqslant \measuredangle(p_{i-1}, p_i).$$

Therefore

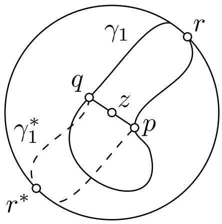
$$|q_{i-1} - q_i| \geqslant |p_{i-1} - p_i|$$

and ❶ follows. □

2.19. Hemisphere lemma. *Any closed spherical curve of length less than $2\cdot\pi$ lies in an open hemisphere.*

This lemma is a keystone in the proof of Fenchel's theorem that will be proven in the next chapter; see 3.10. The lemma is not as simple as you might think — try to prove it yourself before reading the proof. The following proof is due to Stephanie Alexander.

Proof. Let γ be a closed curve in $\mathbb{S}^2$ of length $2\cdot\ell$. Suppose $\ell < \pi$.



Let us divide γ into two arcs γ_1 and γ_2 of length ℓ; denote their endpoints by p and q. Note that

$$\begin{aligned}\measuredangle(p,q) &\leqslant \operatorname{length}\gamma_1 = \\ &= \ell < \\ &< \pi.\end{aligned}$$

Denote by z be the midpoint between p and q in $\mathbb{S}^2$; that is, z is the midpoint of the short arc of a great circle from p to q in $\mathbb{S}^2$. We claim that γ lies in the open north hemisphere with north pole at z. If not, γ intersects the equator in a point r. Without loss of generality we may assume that r lies on γ_1.

Rotate the arc γ_1 by the angle π around the line thru z and the center of the sphere. The obtained arc γ_1^* together with γ_1 forms a closed curve of length $2\cdot\ell$ passing thru r and its antipodal point r^*. Therefore

$$\tfrac{1}{2}\cdot\operatorname{length}\gamma = \ell \geqslant \measuredangle(r, r^*) = \pi,$$

a contradiction. □

2.20. Exercise. *Describe a simple closed spherical curve that does not pass thru a pair of antipodal points and does not lie in any open hemisphere.*

2.21. Exercise. *Suppose that a closed simple spherical curve γ divides $\mathbb{S}^2$ into two regions of equal area. Show that*

$$\operatorname{length}\gamma \geqslant 2\cdot\pi.$$

2.22. Exercise. *Find a flaw in the solution of the following problem. Come up with a correct argument.*

Problem. Suppose that a closed plane curve γ has length at most 4. Show that γ lies in a unit disc.

Wrong solution. Note that it is sufficient to show that diameter of γ is at most 2; that is,

$$|p-q| \leqslant 2 \tag{❷}$$

for any two points p and q on γ.

The length of γ cannot be smaller than the closed inscribed polygonal line which goes from p to q and back to p. Therefore

$$2\cdot|p-q| \leqslant \operatorname{length}\gamma \leqslant 4;$$

whence ❷ follows. □

2.23. Advanced exercises. *Given unit vectors $\mathrm{u}, \mathrm{w} \in \mathbb{S}^2$, denote by w_u the closest point to w on the equator with pole at u; in other words, if $\mathrm{w}^\perp$ is the projection of w to the plane perpendicular to u, then w_u is the unit vector in the direction of $\mathrm{w}^\perp$. The vector w_u is defined if $\mathrm{w} \neq \pm\mathrm{u}$.*

(a) *Show that for any spherical curve γ we have*

$$\operatorname{length}\gamma = \overline{\operatorname{length}\gamma_\mathrm{u}},$$

where $\overline{\operatorname{length}\gamma_\mathrm{u}}$ denotes the average length of γ_u with u varying in $\mathbb{S}^2$. (This is a spherical analog of Crofton's formula.)

(b) *Use (a) to give another proof of the hemisphere lemma (2.19).*

Chapter 3

Curvature

A Acceleration of a unit-speed curve

Recall that any regular smooth curve can be reparametrized by its arc-length (2.7). The obtained parametrized curve, say γ, remains to be smooth and it has unit speed; that is, $|\gamma'(s)| = 1$ for all s. The following proposition states that in this case the acceleration vector stays perpendicular to the velocity vector.

3.1. Proposition. *Assume γ is a smooth unit-speed space curve. Then $\gamma'(s) \perp \gamma''(s)$ for any s.*

The scalar product (also known as dot product) of two vectors v and w will be denoted by $\langle \mathrm{v}, \mathrm{w} \rangle$. Recall that the derivative of a scalar product satisfies the product rule; that is, if $\mathrm{v} = \mathrm{v}(t)$ and $\mathrm{w} = \mathrm{w}(t)$ are smooth vector-valued functions of a real parameter t, then

$$\langle \mathrm{v}, \mathrm{w} \rangle' = \langle \mathrm{v}', \mathrm{w} \rangle + \langle \mathrm{v}, \mathrm{w}' \rangle.$$

Proof. The identity $|\gamma'| = 1$ can be rewritten as $\langle \gamma', \gamma' \rangle = 1$. Differentiating both sides, we get $2 \cdot \langle \gamma'', \gamma' \rangle = \langle \gamma', \gamma' \rangle' = 0$; whence $\gamma'' \perp \gamma'$. $\square$

B Curvature

For a unit-speed smooth space curve γ the magnitude $|\gamma''(s)|$ of its acceleration is called its *curvature* at time s. If γ is simple, then we can say that $|\gamma''(s)|$ is the curvature at the point $p = \gamma(s)$ without ambiguity. The curvature is usually denoted by $\kappa(s)$ or $\kappa(s)_\gamma$ and in the case of simple curves it might be also denoted by $\kappa(p)$ or $\kappa(p)_\gamma$.

The curvature measures how fast the curve turns; if you drive along a plane curve, then the curvature describes the position of your steering wheel at the given point.

In general, the term *curvature* is used for anything that measures how much a *geometric object* deviates from being *straight*; for curves, it measures how fast it deviates from a straight line.

3.2. Exercise. *Show that a regular simple space curve has 0 curvature at each point if and only if it is a segment of a straight line.*

3.3. Exercise. *Let γ be a smooth simple space curve and let γ_λ be a scaled copy of γ with factor $\lambda > 0$; that is, $\gamma_\lambda(t) = \lambda{\cdot}\gamma(t)$ for any t. Show that*

$$\kappa(\lambda{\cdot}p)_{\gamma_\lambda} = \frac{\kappa(p)_\gamma}{\lambda}$$

for any $p \in \gamma$.

3.4. Exercise. *Show that any regular smooth unit-speed spherical curve has curvature at least 1 at each time.*

C Tangent indicatrix

Let γ be a regular smooth space curve. The curve

❶ $$\mathrm{T}(t) = \tfrac{\gamma'(t)}{|\gamma'(t)|};$$

it is called a *tangent indicatrix* of γ. Note that $|\mathrm{T}(t)| = 1$ for any t; that is, T is a spherical curve.

If $s \mapsto \gamma(s)$ is a unit-speed parametrization, then $\mathrm{T}(s) = \gamma'(s)$. In this case we have the following expression for curvature:

$$\kappa(s) = |\mathrm{T}'(s)| = |\gamma''(s)|.$$

For general parametrization $t \mapsto \gamma(t)$, we have instead

❷ $$\kappa(t) = \frac{|\mathrm{T}'(t)|}{|\gamma'(t)|}.$$

Indeed, for an arc-length reparametrization by $s(t)$, we have $s'(t) = |\gamma'(t)|$. Therefore

$$\begin{aligned}\kappa &= \left|\frac{d\mathrm{T}}{ds}\right| = \\ &= |\tfrac{d\mathrm{T}}{dt}|/|\tfrac{ds}{dt}| = \\ &= \frac{|\mathrm{T}'|}{|\gamma'|}.\end{aligned}$$

It follows that the indicatrix of a smooth regular curve γ is regular if the curvature of γ does not vanish.

3.5. Exercise. *Use the formulas ❶ and ❷ to show that for any smooth regular space curve γ we have the following expressions for its curvature:*

(a)
$$\kappa = \frac{|\mathrm{w}|}{|\gamma'|^2},$$
where $\mathrm{w} = \mathrm{w}(t)$ denotes the projection of $\gamma''(t)$ to the plane normal to $\gamma'(t)$;

(b)
$$\kappa = \frac{|\gamma'' \times \gamma'|}{|\gamma'|^3},$$
where $\times$ denotes the vector product *(also known as* cross product*).*

3.6. Exercise. *Apply the formulas in the previous exercise to show that if f is a smooth real function, then its graph $y = f(x)$ has curvature*
$$\kappa(p) = \frac{|f''(x)|}{(1 + f'(x)^2)^{\frac{3}{2}}}$$
at the point $p = (x, f(x))$.

3.7. Exercise. *Show that any smooth regular curve $\gamma\colon \mathbb{I} \to \mathbb{R}^3$ with curvature at most 1 can be approximated by smooth curves with constant curvature 1.*

In other words, construct a sequence $\gamma_n\colon \mathbb{I} \to \mathbb{R}^3$ of smooth regular curves with constant curvature 1 such that $\gamma_n(t) \to \gamma(t)$ for any t as $n \to \infty$.

D Tangent curves

Let γ be a smooth regular space curve and T its tangent indicatrix. The line thru $\gamma(t)$ in the direction of $\mathrm{T}(t)$ is called the *tangent line* of γ at t.

The tangent line could be also defined as a unique line that has that has *first order of contact* with γ at s; that is, $\rho(\ell) = o(\ell)$, where $\rho(\ell)$ denotes the distance from $\gamma(s + \ell)$ to the line.

We say that a smooth regular curve γ_1 at s_1 is *tangent* to a smooth regular curve γ_2 at s_2 if $\gamma_1(s_1) = \gamma_2(s_2)$ and the tangent line of γ_1 at s_1 coincides with the tangent line of γ_2 at s_2; if both curves are simple we can also say that they are tangent at the point $p = \gamma_1(s_1) = \gamma_2(s_2)$ without ambiguity.

E Total curvature

Let $\gamma\colon \mathbb{I}\to\mathbb{R}^3$ be a smooth unit-speed curve and T its tangent indicatrix. The integral

$$\Phi(\gamma):=\int\limits_{\mathbb{I}}\kappa(s)\cdot ds$$

is called the *total curvature of* γ.

Rewriting the above integral using a change of variables produce a formula for a general parametrization $t\mapsto\gamma(t)$:

❶
$$\Phi(\gamma):=\int\limits_{\mathbb{I}}\kappa(t)\cdot|\gamma'(t)|\cdot dt.$$

3.8. Exercise. *Find the curvature of the helix*

$$\gamma_{a,b}(t)=(a\cdot\cos t,a\cdot\sin t,b\cdot t),$$

its tangent indicatrix and the total curvature of its arc $\gamma_{a,b}|_{[0,2\cdot\pi]}$.

Note that for a unit-speed smooth curve, the speed of its tangent indicatrix equals its curvature. Therefore 2.4 implies the following:

3.9. Observation. *The total curvature of a smooth regular curve is the length of its tangent indicatrix.*

3.10. Fenchel's theorem. *The total curvature of any closed regular space curve is at least* $2\cdot\pi$.

Proof. Fix a closed regular space curve γ. We can assume that γ is described by a unit-speed loop $\gamma\colon[a,b]\to\mathbb{R}^3$; in this case $\gamma(a)=\gamma(b)$ and $\gamma'(a)=\gamma'(b)$.

Consider its tangent indicatrix $\mathrm{T}=\gamma'$. Recall that $|\mathrm{T}(s)|=1$ for any s; that is, T is a closed spherical curve.

Let us show that T cannot lie in a hemisphere. Assume the contrary; without loss of generality we can assume that it lies in the north hemisphere defined by the inequality $z>0$ in (x,y,z)-coordinates. In other words, if $\gamma(t)=(x(t),y(t),z(t))$, then $z'(t)>0$ for any t. Therefore

$$z(b)-z(a)=\int\limits_a^b z'(s)\cdot ds>0.$$

In particular, $\gamma(a)\neq\gamma(b)$, a contradiction.

Applying the observation (3.9) and the hemisphere lemma (2.19), we get

$$\Phi(\gamma) = \operatorname{length} \mathrm{T} \geqslant 2\cdot\pi. \qquad \square$$

3.11. Exercise. *Show that a closed space curve γ with curvature at most 1 cannot be shorter than the unit circle; that is,*

$$\operatorname{length} \gamma \geqslant 2\cdot\pi.$$

3.12. Advanced exercise. *Suppose that γ is a smooth regular space curve that does not pass thru the origin. Consider the spherical curve defined as $\sigma(t) = \frac{\gamma(t)}{|\gamma(t)|}$ for any t. Show that*

$$\operatorname{length} \sigma < \Phi(\gamma) + \pi.$$

Moreover, if γ is closed, then

$$\operatorname{length} \sigma \leqslant \Phi(\gamma).$$

Note that the last inequality gives an alternative proof of Fenchel's theorem. Indeed, without loss of generality we can assume that the origin lies on a chord of γ. In this case the closed spherical curve σ goes from a point to its antipode and comes back; it takes length π each way, whence

$$\operatorname{length} \sigma \geqslant 2\cdot\pi.$$

Recall that the curvature of a spherical curve is at least 1 (see 3.4). In particular, the length of a spherical curve cannot exceed its total curvature. The following theorem shows that the same inequality holds for *closed* curves in a unit ball.

3.13. Theorem. *Let γ be a smooth regular closed curve that lies in a unit ball. Then*

$$\Phi(\gamma) \geqslant \operatorname{length} \gamma.$$

The 2-dimensional case of this theorem was proved by István Fáry [28]. It was generalized by Don Chakerian [17] to higher dimensions. This theorem has many very interesting and very different proofs; a number of them are collected by Serge Tabachnikov [75]. The following exercise will guide you thru another proof by Don Chakerian [18]:

3.14. Exercise. *Let $\gamma\colon [0,\ell] \to \mathbb{R}^3$ be a smooth unit-speed closed curve that lies in the unit ball; that is, $|\gamma| \leqslant 1$.*

(a) Show that

$$\langle\gamma''(s),\gamma(s)\rangle \geqslant -\kappa(s)$$

for any s.

(b) Use part (a) to show that

$$\int_0^\ell \langle\gamma(s),\gamma'(s)\rangle'\cdot ds \geqslant \operatorname{length}\gamma - \Phi(\gamma).$$

(c) Suppose that $\gamma(0)=\gamma(\ell)$ and $\gamma'(0)=\gamma'(\ell)$. Show that

$$\int_0^\ell \langle\gamma(s),\gamma'(s)\rangle'\cdot ds = 0.$$

Use this equality together with part (b) to prove 3.13.

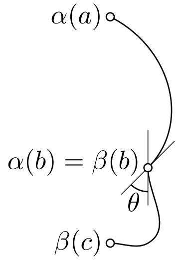
F Piecewise smooth curves

Assume $\alpha\colon [a,b]\to\mathbb{R}^3$ and $\beta\colon [b,c]\to\mathbb{R}^3$ are two curves such that $\alpha(b) = \beta(b)$. Note that these two curves can be combined into one $\gamma\colon [a,c]\to \mathbb{R}^3$ by the rule

$$\gamma(t)=\begin{cases}\alpha(t) & \text{if} \quad t\leqslant b,\\ \beta(t) & \text{if} \quad t\geqslant b.\end{cases}$$

The obtained curve γ is called the *concatenation* of α and β. (The condition $\alpha(b)=\beta(b)$ ensures that the map $t\mapsto\gamma(t)$ is continuous.)

The same definition of concatenation can be applied if α and/or β are defined on semiopen intervals $(a,b]$ and/or $[b,c)$.



The assumption that the intervals of definition of α and β fit together is not essential — one can concatenate any of the curves as long as the endpoint of α coincides with the starting point of β. If this is the case, then the time intervals of both curves can be shifted so that they fit together.

If in addition $\beta(c)=\alpha(a)$, then we can do cyclic concatenation of these curves; this way we obtain a closed curve.

If $\alpha'(b)$ and $\beta'(b)$ are defined, then the angle $\theta=\measuredangle(\alpha'(b),\beta'(b))$ is called the *external angle* of γ at time b. If $\theta=\pi$, then we say that γ has a *cusp* at time b.

A space curve γ is called *piecewise smooth and regular* if it can be presented as an iterated concatenation of a finite number of smooth regular curves; if γ is closed, then the concatenation is assumed to be cyclic.

If γ is a concatenation of smooth regular arcs $\gamma_1, \dots, \gamma_n$, then the total curvature of γ is defined as a sum of the total curvatures of γ_i and the external angles; that is,

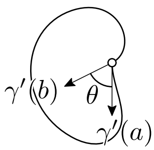
$$\Phi(\gamma) = \Phi(\gamma_1) + \dots + \Phi(\gamma_n) + \theta_1 + \dots + \theta_{n-1}$$

where θ_i is the external angle at the joint between γ_i and γ_{i+1}; if γ is closed, then

$$\Phi(\gamma) = \Phi(\gamma_1) + \dots + \Phi(\gamma_n) + \theta_1 + \dots + \theta_n,$$

where θ_n is the external angle at the joint between γ_n and γ_1.

In particular, for a smooth regular loop $\gamma\colon [a,b] \to \to \mathbb{R}^3$, the total curvature of the corresponding closed curve $\hat\gamma$ is defined as



$$\Phi(\hat\gamma) := \Phi(\gamma) + \theta,$$

where $\theta = \measuredangle(\gamma'(a), \gamma'(b))$.

3.15. Generalized Fenchel's theorem. *Let γ be a closed piecewise smooth regular space curve. Then*

$$\Phi(\gamma) \geqslant 2{\cdot}\pi.$$

Proof. Suppose γ is a cyclic concatenation of n smooth regular arcs $\gamma_1, \dots, \gamma_n$. Denote by $\theta_1, \dots, \theta_n$ its external angles. We need to show that

➊ $$\Phi(\gamma_1) + \dots + \Phi(\gamma_n) + \theta_1 + \dots + \theta_n \geqslant 2{\cdot}\pi.$$

Consider the tangent indicatrix T_i for each arc γ_i; these are smooth spherical arcs.

The same argument as in the proof of Fenchel's theorem, shows that the curves $\mathrm{T}_1, \dots, \mathrm{T}_n$ cannot lie in an open hemisphere.

Note that the spherical distance from the endpoint of T_i to the starting point of T_{i+1} is equal to the external angle θ_i (we enumerate the arcs modulo n, so $\gamma_{n+1} = \gamma_1$). Let us connect the endpoint of T_i to the starting point of T_{i+1} by a short arc of a great circle in the sphere. This way we get a closed spherical curve that is $\theta_1 + \dots + \theta_n$ longer then the total length of $\mathrm{T}_1, \dots, \mathrm{T}_n$.

Applying the hemisphere lemma (2.19) to the obtained closed curve, we get that

$$\operatorname{length} \mathrm{T}_1 + \dots + \operatorname{length} \mathrm{T}_n + \theta_1 + \dots + \theta_n \geqslant 2{\cdot}\pi.$$

By 3.9, the statement follows. □

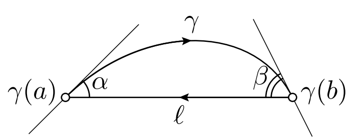
3.16. Chord lemma. *Let $\gamma\colon [a,b] \to \mathbb{R}^3$ be a smooth regular arc, and ℓ be its chord. Assume γ meets ℓ at angles α and β at $\gamma(a)$ and $\gamma(b)$, respectively; that is,*

$$\alpha = \measuredangle(\mathrm{w}, \gamma'(a)) \quad \textit{and} \quad \beta = \measuredangle(\mathrm{w}, \gamma'(b)),$$

where $\mathrm{w} = \gamma(b) - \gamma(a)$. Then

❷ $$\Phi(\gamma) \geqslant \alpha + \beta.$$

Proof. Let us parameterize the chord ℓ from $\gamma(b)$ to $\gamma(a)$ and consider the cyclic concatenation $\bar\gamma$ of γ and ℓ. The closed curve $\bar\gamma$ has two external angles $\pi - \alpha$ and $\pi - \beta$.



Since the curvature of ℓ vanishes, we get

$$\Phi(\bar\gamma) = \Phi(\gamma) + (\pi - \alpha) + (\pi - \beta).$$

According to the generalized Fenechel's theorem (3.15), $\Phi(\bar\gamma) \geqslant 2{\cdot}\pi$; hence ❷ follows. □

3.17. Exercise. *Show that the estimate in the chord lemma is optimal. That is, given two points p, q and two unit vectors u, v in $\mathbb{R}^3$, show that there is a smooth regular curve γ that starts at p in the direction u and ends at q in the direction v such that $\Phi(\gamma)$ is arbitrarily close to $\measuredangle(\mathrm{w}, \mathrm{u}) + \measuredangle(\mathrm{w}, \mathrm{v})$, where $\mathrm{w} = q - p$.*

G Polygonal lines

Polygonal lines are a particular case of piecewise smooth regular curves; each arc in its concatenation is a line segment. Since the curvature of a line segment vanishes, the total curvature of a polygonal line is the sum of its external angles.

3.18. Exercise. *Let a, b, c, d and x be distinct points in $\mathbb{R}^3$. Show that the total curvature of the polygonal line $abcd$ cannot exceed the total curvature of $abxcd$; that is,*

$$\Phi(abcd) \leqslant \Phi(abxcd).$$

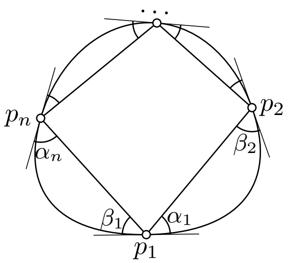
Use this statement to show that any closed polygonal line has curvature at least $2{\cdot}\pi$.

3.19. Proposition. *Assume a polygonal line $\beta = p_0 \dots p_n$ is inscribed in a smooth regular curve γ. Then*

$$\Phi(\gamma) \geqslant \Phi(\beta).$$

Moreover if γ is closed we allow the inscribed polygonal line β to be closed.

Proof. Since the curvature of line segments vanishes, the total curvature of the polygonal line is the sum of external angles $\theta_i = \pi - \measuredangle[p_i\,{}^{p_{i-1}}_{p_{i+1}}]$. Assume $p_i = \gamma(t_i)$. Set



$$\begin{aligned} \mathrm{w}_i &= p_{i+1} - p_i, & \mathrm{v}_i &= \gamma'(t_i), \\ \alpha_i &= \measuredangle(\mathrm{w}_i, \mathrm{v}_i), & \beta_i &= \measuredangle(\mathrm{w}_{i-1}, \mathrm{v}_i). \end{aligned}$$

In the case of a closed curve we use indexes modulo n; so in this case we have $p_{n+1} = p_1$.

Note that $\theta_i = \measuredangle(\mathrm{w}_{i-1}, \mathrm{w}_i)$. By triangle inequality for angles 0.12, we get that

$$\theta_i \leqslant \alpha_i + \beta_i.$$

By the chord lemma, the total curvature of the arc of γ from p_i to p_{i+1} is at least $\alpha_i + \beta_{i+1}$.

Therefore if γ is a closed curve, we have

$$\begin{aligned} \Phi(\beta) &= \theta_1 + \dots + \theta_n \leqslant \\ &\leqslant \beta_1 + \alpha_1 + \dots + \beta_n + \alpha_n = \\ &= (\alpha_1 + \beta_2) + \dots + (\alpha_n + \beta_1) \leqslant \\ &\leqslant \Phi(\gamma). \end{aligned}$$

If γ is an arc, the argument is analogous:

$$\begin{aligned} \Phi(\beta) &= \theta_1 + \dots + \theta_{n-1} \leqslant \\ &\leqslant \beta_1 + \alpha_1 + \dots + \beta_{n-1} + \alpha_{n-1} \leqslant \\ &\leqslant (\alpha_0 + \beta_1) + \dots + (\alpha_{n-1} + \beta_n) \leqslant \\ &\leqslant \Phi(\gamma). \end{aligned}$$

□

The following exercise states that the inequality in 3.19 is optimal.

3.20. Exercise. *Show that for any regular smooth space curve γ we have that*

$$\Phi(\gamma) = \sup\{\Phi(\beta)\},$$

where the least upper bound is taken over all polygonal lines β inscribed in γ (if γ is closed we assume that so is β).

This exercise can be used to generalize the notion of total curvature of arbitrary curve γ. Namely it can be defined as *the least upper bound on the total curvatures of inscribed nondegenerate polygonal lines inscribed in* γ.

It is possible to generalize most of the statements in this chapter to the (nonsmooth) curves of finite total curvature. This theory was developed by Alexandr Alexandrov and Yuri Reshetnyak [3]; a good survey on the subject is written by John Sullivan [74].

3.21. Exercise.

(a) *Draw a smooth regular plane curve γ that has a self-intersection and such that $\Phi(\gamma) < 2\cdot\pi$.*

(b) *Show that if a smooth regular curve $\gamma\colon [a,b] \to \mathbb{R}^3$ has a self-intersection, then $\Phi(\gamma) > \pi$.*

3.22. Proposition. *The equality case in Fenchel's theorem holds only for convex plane curves; that is, if the total curvature of a smooth regular space curve γ equals $2\cdot\pi$, then γ is a convex plane curve.*

The proof is an application of Proposition 3.19.

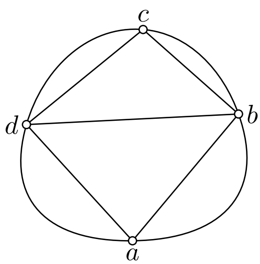
Proof. Consider an inscribed quadrangle $abcd$ in γ. By the definition of total curvature, we have that

$$\begin{aligned}\Phi(abcd) &= (\pi - \measuredangle[a\,{}^{d}_{b}]) + (\pi - \measuredangle[b\,{}^{a}_{c}]) + (\pi - \measuredangle[c\,{}^{b}_{d}]) + (\pi - \measuredangle[d\,{}^{c}_{a}]) = \\ &= 4\cdot\pi - (\measuredangle[a\,{}^{d}_{b}] + \measuredangle[b\,{}^{a}_{c}] + \measuredangle[c\,{}^{b}_{d}] + \measuredangle[d\,{}^{c}_{a}])\end{aligned}$$

Note that

❶ $$\measuredangle[b\,{}^{a}_{c}] \leqslant \measuredangle[b\,{}^{a}_{d}] + \measuredangle[b\,{}^{d}_{c}] \quad\text{and}\quad \measuredangle[d\,{}^{c}_{a}] \leqslant \measuredangle[d\,{}^{c}_{b}] + \measuredangle[d\,{}^{b}_{a}].$$

The sum of angles in any triangle is π, so combining these inequalities, we get that



$$\begin{aligned}\Phi(abcd) \geqslant 4\cdot\pi &- (\measuredangle[a\,{}^{d}_{b}] + \measuredangle[b\,{}^{a}_{d}] + \measuredangle[d\,{}^{b}_{a}]) - \\ &- (\measuredangle[c\,{}^{b}_{d}] + \measuredangle[d\,{}^{c}_{b}] + \measuredangle[b\,{}^{d}_{c}]) = \\ = 2\cdot\pi.\end{aligned}$$

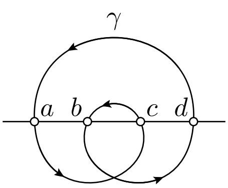
By 3.19,

$$\Phi(abcd) \leqslant \Phi(\gamma) \leqslant 2\cdot\pi.$$

Therefore we have equalities in ❶. It means that the point d lies in the angle abc and the point b lies in the angle cda. That is, $abcd$ is a convex plane quadrangle.

It follows that any quadrangle inscribed in γ is a convex plane quadrangle. Therefore all points of γ lie in one plane defined by three points on γ. Further, since any quadrangle inscribed in γ is a convex, we get that γ is a convex plane curve. □

3.23. Exercise. *Suppose that a closed curve γ crosses a line at four points a, b, c and d. Assume that these points appear on the line in the order a, b, c, d and they appear on the curve γ in the order a, c, b, d. Show that*



$$\Phi(\gamma) \geqslant 4{\cdot}\pi.$$

Lines crossing a curve at four points as in the exercise are called *alternating quadrisecants.* It turns out that any *nontrivial knot* admits an alternating quadrisecant [24]; according to the exercise the latter implies the so-called *Fáry–Milnor theorem* — the total curvature of any knot exceeds $4{\cdot}\pi$.

H Bow lemma

The following lemma was proved by Axel Schur [72]; it is a differential-geometric analog of the so called *arm lemma* of Augustin-Louis Cauchy.

3.24. Lemma. *Let $\gamma_1\colon [a,b] \to \mathbb{R}^2$ and $\gamma_2\colon [a,b] \to \mathbb{R}^3$ be two smooth unit-speed curves. Suppose that $\kappa(s)_{\gamma_1} \geqslant \kappa(s)_{\gamma_2}$ for any s and the curve γ_1 is an arc of a convex curve; that is, it runs in the boundary of a convex plane figure. Then the distance between the ends of γ_1 cannot exceed the distance between the ends of γ_2; that is,*

$$|\gamma_1(b) - \gamma_1(a)| \leqslant |\gamma_2(b) - \gamma_2(a)|.$$

The following exercise states that the condition that γ_1 is a convex arc is necessary. It is instructive to do this exercise before reading the proof of the lemma.

3.25. Exercise. *Construct two simple smooth unit-speed plane curves $\gamma_1, \gamma_2\colon [a,b] \to \mathbb{R}^2$ such that that $\kappa(s)_{\gamma_1} > \kappa(s)_{\gamma_2} > 0$ for any s and*

$$|\gamma_1(b) - \gamma_1(a)| > |\gamma_2(b) - \gamma_2(a)|.$$

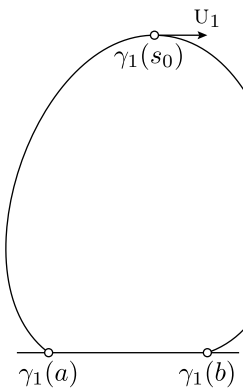
Proof. Denote by T_1 and T_2 the tangent indicatrixes of γ_1 and γ_2, respectively.

Let $\gamma_1(s_0)$ be the point on γ_1 furthest to the line thru $\gamma(a)$ and $\gamma(b)$. Consider two unit vectors

$$\mathrm{U}_1 = \mathrm{T}_1(s_0) = \gamma_1'(s_0) \quad \text{and} \quad \mathrm{U}_2 = \mathrm{T}_2(s_0) = \gamma_2'(s_0).$$

Since γ_1 is an arc of a convex curve, its indicatrix T_1 runs in one direction along the unit circle. Suppose $s \leqslant s_0$, then

$$\begin{aligned}
\measuredangle(\gamma_1'(s), \mathrm{U}_1) &= \measuredangle(\mathrm{T}_1(s), \mathrm{T}_1(s_0)) = \\
&= \operatorname{length}(\mathrm{T}_1|_{[s,s_0]}) = \\
&= \int_s^{s_0} |\mathrm{T}_1'(t)|\cdot dt = \\
&= \int_s^{s_0} \kappa_1(t)\cdot dt \geqslant \\
&\geqslant \int_s^{s_0} \kappa_2(t)\cdot dt = \\
&= \int_s^{s_0} |\mathrm{T}_2'(t)|\cdot dt = \\
&= \operatorname{length}(\mathrm{T}_1|_{[s,s_0]}) \geqslant \\
&\geqslant \measuredangle(\mathrm{T}_2(s), \mathrm{T}_2(s_0)) = \\
&= \measuredangle(\gamma_2'(s), \mathrm{U}_2).
\end{aligned}$$



That is,

❶ $$\measuredangle(\gamma_1'(s), \mathrm{U}_1) \geqslant \measuredangle(\gamma_2'(s), \mathrm{U}_2)$$

if $s \geqslant s_0$. The same argument shows that ❶ holds true for $s \geqslant s_0$. Therefore the inequality ❶ holds for any s.

Since

$$1 = |\gamma_1'(s)| = |\gamma_2'(s)| = |\mathrm{U}_1| = |\mathrm{U}_2|,$$

the inequality ❶ implies that

❷ $$\langle \gamma_1'(s), \mathrm{U}_1\rangle \leqslant \langle \gamma_2'(s), \mathrm{U}_2\rangle$$

for any s.

Further, since U_1 is a unit vector parallel to $\gamma_1(b) - \gamma_1(a)$, we have that

$$|\gamma_1(b) - \gamma_1(a)| = \langle \mathrm{U}_1, \gamma_1(b) - \gamma_1(a)\rangle,$$

and since u_2 is a unit vector, we have that

$$|\gamma_2(b)-\gamma_2(a)| \geqslant \langle \mathrm{u}_2, \gamma_2(b)-\gamma_2(a)\rangle.$$

Integrating ❷, we get

$$\begin{aligned}|\gamma_1(b)-\gamma_1(a)| = \langle \mathrm{u}_1, \gamma_1(b)-\gamma_1(a)\rangle &= \\ &= \int_a^b \langle \mathrm{u}_1, \gamma_1'(s)\rangle\cdot ds \leqslant \\ &\leqslant \int_a^b \langle \mathrm{u}_2, \gamma_2'(s)\rangle\cdot ds = \\ &= \langle \mathrm{u}_2, \gamma_2(b)-\gamma_2(a)\rangle \leqslant \\ &\leqslant |\gamma_2(b)-\gamma_2(a)|.\end{aligned}$$

□

3.26. Exercise. *Let $\gamma\colon [a,b]\to\mathbb{R}^3$ be a smooth regular curve and $0 < \theta \leqslant \frac{\pi}{2}$. Assume*

$$\Phi(\gamma)\leqslant 2\cdot\theta.$$

(a) Show that

$$|\gamma(b)-\gamma(a)| > \cos\theta\cdot \operatorname{length}\gamma.$$

(b) Use part (a) to give another solution of 3.21b.

(c) Show that the inequality in (a) is optimal; that is, given θ there is a smooth regular curve γ such that $\frac{|\gamma(b)-\gamma(a)|}{\operatorname{length}\gamma}$ is arbitrarily close to $\cos\theta$.

The statement in the following exercise is generally attributed to Hermann Schwarz [72].

3.27. Exercise. *Let p and q be points in a unit circle dividing it in two arcs with lengths $\ell_1 < \ell_2$. Suppose a space curve γ connects p to q and has curvature at most 1. Show that either*

$$\operatorname{length}\gamma \leqslant \ell_1 \quad \textit{or} \quad \operatorname{length}\gamma \geqslant \ell_2.$$

The following exercise generalizes 3.11.

3.28. Exercise. *Suppose $\gamma\colon [a,b]\to\mathbb{R}^3$ is a smooth regular loop with curvature at most 1. Show that*

$$\operatorname{length}\gamma \geqslant 2\cdot\pi.$$

Chapter 4

Torsion

This chapter provides mostly a practice in computations. Except for the definitions in Section 4A, it will not be used in the sequel.

A Frenet frame

Let γ be a smooth regular space curve. Without loss of generality, we may assume that γ has an arc-length parametrization, so the velocity vector $\mathrm{T}(s) = \gamma'(s)$ is unit.

Assume its curvature does not vanish at some time s; in other words, $\gamma''(s) \neq 0$. Then we can define the so-called *normal vector* at s as

$$\mathrm{N}(s) = \frac{\gamma''(s)}{|\gamma''(s)|}.$$

Note that

$$\mathrm{T}'(s) = \gamma''(s) = \kappa(s)\cdot\mathrm{N}(s).$$

According to 3.1, $\mathrm{N}(s) \perp \mathrm{T}(s)$. Therefore the vector product

$$\mathrm{B}(s) = \mathrm{T}(s) \times \mathrm{N}(s)$$

is a unit vector; moreover the triple $\mathrm{T}(s), \mathrm{N}(s), \mathrm{B}(s)$ an oriented orthonormal basis in $\mathbb{R}^3$; in particular, we have that

❶
$$\begin{aligned} \langle \mathrm{T}, \mathrm{T}\rangle &= 1, & \langle \mathrm{N}, \mathrm{N}\rangle &= 1, & \langle \mathrm{B}, \mathrm{B}\rangle &= 1, \\ \langle \mathrm{T}, \mathrm{N}\rangle &= 0, & \langle \mathrm{N}, \mathrm{B}\rangle &= 0, & \langle \mathrm{B}, \mathrm{T}\rangle &= 0. \end{aligned}$$

The orthonormal basis $\mathrm{T}(s), \mathrm{N}(s), \mathrm{B}(s)$ is called the *Frenet frame* at s; the vectors in the frame are called *tangent*, *normal* and *binormal* respectively. Note that the frame $\mathrm{T}(s), \mathrm{N}(s), \mathrm{B}(s)$ is defined only if $\kappa(s) \neq 0$.

The plane Π_s thru $\gamma(s)$ spanned by vectors $\mathrm{T}(s)$ and $\mathrm{N}(s)$ is called the *osculating plane* at s; equivalently it can be defined as a plane thru $\gamma(s)$ that is perpendicular to the binormal vector $\mathrm{B}(s)$. This is the unique plane that has a *second order of contact* with γ at s; that is, $\rho(\ell) = o(\ell^2)$, where $\rho(\ell)$ denotes the distance from $\gamma(s+\ell)$ to Π_s.

B Torsion

Let γ be a smooth unit-speed space curve and $\mathrm{T}(s), \mathrm{N}(s), \mathrm{B}(s)$ be its Frenet frame. The value

$$\tau(s) = \langle \mathrm{N}'(s), \mathrm{B}(s)\rangle$$

is called the *torsion* of γ at s.

Note that the torsion $\tau(s_0)$ is defined if $\kappa(s_0) \neq 0$. Indeed, since the function $s \mapsto \kappa(s)$ is continuous, $\kappa(s_0) \neq 0$ implies that $\kappa(s) \neq 0$ for all s near s_0. Therefore the Frenet frame is also defined in an open interval containing s_0. Clearly $\mathrm{T}(s)$, $\mathrm{N}(s)$ and $\mathrm{B}(s)$ depend smoothly on s in their domains of definition. Therefore $\mathrm{N}'(s_0)$ is defined and so is the torsion $\tau(s_0) = \langle \mathrm{N}'(s_0), \mathrm{B}(s_0)\rangle$.

4.1. Exercise. *Given real numbers a and b, calculate the curvature and the torsion of the helix*

$$\gamma_{a,b}(t) = (a\cdot \cos t, a\cdot \sin t, b\cdot t).$$

Conclude that for any $\kappa > 0$ and τ there is a helix with constant curvature κ and torsion τ.

C Frenet formulas

Assume that the Frenet frame $\mathrm{T}(s), \mathrm{N}(s), \mathrm{B}(s)$ of a curve γ is defined at s. Recall that

❶ $$\mathrm{T}' = \kappa\cdot\mathrm{N}.$$

Let us find the remaining derivatives N' and B' in the frame $\mathrm{T}, \mathrm{N}, \mathrm{B}$.

First let us show that

❷ $$\mathrm{N}' = -\kappa\cdot\mathrm{T} + \tau\cdot\mathrm{B}.$$

Since the frame $\mathrm{T}, \mathrm{N}, \mathrm{B}$ is orthonormal, the above formula is equivalent to the following three identities:

❸ $$\langle \mathrm{N}', \mathrm{T}\rangle = -\kappa, \quad \langle \mathrm{N}', \mathrm{N}\rangle = 0, \quad \langle \mathrm{N}', \mathrm{B}\rangle = \tau,$$

The last identity follows from the definition of torsion. The second one is a consequence of the identity $\langle \mathrm{N}, \mathrm{N} \rangle = 1$ in ❶. By differentiating the identity $\langle \mathrm{T}, \mathrm{N} \rangle = 0$ in ❶ we get

$$\langle \mathrm{T}', \mathrm{N} \rangle + \langle \mathrm{T}, \mathrm{N}' \rangle = 0.$$

Applying ❶, we get the first equation in ❸.

Differentiating the third identity in ❶, we get that $\mathrm{B}' \perp \mathrm{B}$. Taking further derivatives of the other identities with B in ❶, we get that

$$\begin{aligned} \langle \mathrm{B}', \mathrm{T} \rangle &= -\langle \mathrm{B}, \mathrm{T}' \rangle = -\kappa \cdot \langle \mathrm{B}, \mathrm{N} \rangle = 0, \\ \langle \mathrm{B}', \mathrm{N} \rangle &= -\langle \mathrm{B}, \mathrm{N}' \rangle = \tau. \end{aligned}$$

Since the frame $\mathrm{T}, \mathrm{N}, \mathrm{B}$ is orthonormal, it follows that

❹ $$\mathrm{B}' = -\tau \cdot \mathrm{N}.$$

The equations ❶, ❷ and ❹ are called *Frenet formulas.* All three can be written as one matrix identity:

$$\begin{pmatrix} \mathrm{T}' \\ \mathrm{N}' \\ \mathrm{B}' \end{pmatrix} = \begin{pmatrix} 0 & \kappa & 0 \\ -\kappa & 0 & \tau \\ 0 & -\tau & 0 \end{pmatrix} \cdot \begin{pmatrix} \mathrm{T} \\ \mathrm{N} \\ \mathrm{B} \end{pmatrix}.$$

Since B is the normal vector to the osculating plane, Equation ❹ shows that the torsion measures how fast the osculating plane rotates when one travels along γ.

4.2. Exercise. *Deduce the formula ❹ from ❶ and ❷ by differentiating the identity* $\mathrm{B} = \mathrm{T} \times \mathrm{N}$.

4.3. Exercise. *Let γ be a regular space curve with nonvanishing curvature. Show that γ lies in a plane if and only if its torsion vanishes.*

4.4. Exercise. *Let γ be a smooth regular space curve, κ and τ its curvature and torsion, and $\mathrm{T}, \mathrm{N}, \mathrm{B}$ its Frenet frame. Show that*

$$\mathrm{B} = \frac{\gamma' \times \gamma''}{|\gamma' \times \gamma''|} \quad \textit{and} \quad \tau = \frac{\langle \gamma' \times \gamma'', \gamma''' \rangle}{|\gamma' \times \gamma''|^2}.$$

D Curves of constant slope

We say that a smooth regular space curve γ has *constant slope* if its velocity vector makes a constant angle with a fixed direction. The following theorem was proved by Michel Ange Lancret [52] more than two centuries ago.

4.5. Theorem. *Let γ be a smooth regular curve; denote by κ and τ its curvature and torsion. Suppose $\kappa(s) > 0$ for all s. Then γ has constant slope if and only if the ratio $\frac{\tau}{\kappa}$ is constant.*

The following exercise will guide you thru the proof of the theorem.

4.6. Exercise. *Let γ be a smooth regular space curve with nonvanishing curvature, $\mathrm{T}, \mathrm{N}, \mathrm{B}$ its Frenet frame and κ, τ its curvature and torsion.*

(a) Assume that $\langle \mathrm{W}, \mathrm{T}\rangle$ is constant for a fixed nonzero vector W. Show that

$$\langle \mathrm{W}, \mathrm{N}\rangle = 0.$$

Use it to show that

$$\langle \mathrm{W}, -\kappa\cdot\mathrm{T} + \tau\cdot\mathrm{B}\rangle = 0.$$

Use these two identities to show that $\frac{\tau}{\kappa}$ is constant; this proves the "only if" part of the theorem.

(b) Assume $\frac{\tau}{\kappa}$ is constant, show that the vector $\mathrm{W} = \frac{\tau}{\kappa}\cdot\mathrm{T}+\mathrm{B}$ is constant. Conclude that γ has constant slope; this proves the "if" part of the theorem.

Let γ be a smooth unit-speed curve and s_0 a fixed real number. Then the curve

$$\alpha(s) = \gamma(s) + (s_0 - s)\cdot\gamma'(s)$$

is called the *evolvent* of γ. Note that if $\ell(s)$ denotes the tangent line to γ at s, then $\alpha(s) \in \ell(s)$ and $\alpha'(s) \perp \ell$ for all s.

4.7. Exercise. *Show that the evolvent of a constant slope curve is a plane curve.*

E Spherical curves

4.8. Theorem. *Suppose that γ is a smooth regular space curve with nonvanishing torsion τ and (therefore) curvature κ. Then γ lies in a unit sphere if and only if the following identity holds true:*

$$\left|\frac{\kappa'}{\tau}\right| = \kappa\cdot\sqrt{\kappa^2 - 1}.$$

The proof is another application of the Frenet formulas; we present it in the form of a guided exercise:

4.9. Exercise. *Suppose γ is a smooth unit-speed space curve. Denote by $\mathrm{T}, \mathrm{N}, \mathrm{B}$ its Frenet frame and by κ, τ its curvature and torsion.*

Assume that γ is spherical; that is, $|\gamma(s)| = 1$ for any s. Show that

(a) *$\langle \mathrm{T}, \gamma\rangle = 0$; conclude that $\langle \mathrm{N}, \gamma\rangle^2 + \langle \mathrm{B}, \gamma\rangle^2 = 1$.*
(b) *$\langle \mathrm{N}, \gamma\rangle = -\frac{1}{\kappa}$;*
(c) *$\langle \mathrm{B}, \gamma\rangle' = \frac{\tau}{\kappa}$.*
(d) *Use (c) to show that if γ is closed, then $\tau(s) = 0$ for some s.*
(e) *Assume that the torsion of γ does not vanish. Use (a)–(c) to show that*

$$\left|\frac{\kappa'}{\tau}\right| = \kappa\cdot\sqrt{\kappa^2 - 1}.$$

(It proves the "only if" part of the theorem.)

Now assume that γ is a space curve that satisfies the identity in (e).

(f) *Show that $p = \gamma + \frac{1}{\kappa}\cdot\mathrm{N} + \frac{\kappa'}{\kappa^2\cdot\tau}\cdot\mathrm{B}$ is constant; conclude that γ lies in the unit sphere centered at p. (It proves the "if" part of the theorem.)*

For a unit-speed curve γ with nonzero curvature and torsion at s, the sphere Σ_s with center

$$p(s) = \gamma(s) + \tfrac{1}{\kappa(s)}\cdot\mathrm{N}(s) + \tfrac{\kappa'(s)}{\kappa^2(s)\cdot\tau(s)}\cdot\mathrm{B}(s)$$

and passing thru $\gamma(s)$ is called the *osculatingsphere* of γ at s. This is the unique sphere that has *third order of contact* with γ at s; that is, $\rho(\ell) = o(\ell^3)$, where $\rho(\ell)$ denotes the distance from $\gamma(s+\ell)$ to Σ_s.

F Fundamental theorem of space curves

4.10. Theorem. *Let $\kappa(s)$ and $\tau(s)$ be two smooth real valued functions defined on a real interval $\mathbb{I}$. Suppose $\kappa(s) > 0$ for all s. Then there is a smooth unit-speed curve $\gamma\colon \mathbb{I} \to \mathbb{R}^3$ with curvature $\kappa(s)$ and torsion $\tau(s)$ for every s. Moreover γ is uniquely defined up to a rigid motion of the space.*

The proof is an application of the theorem on existence and uniqueness of solutions of ordinary differential equations (0.25).

Proof. Fix a parameter value s_0, a point $\gamma(s_0)$ and an oriented orthonormal frame $\mathrm{T}(s_0)$, $\mathrm{N}(s_0)$, $\mathrm{B}(s_0)$.

Consider the following system of differential equations

❶
$$\begin{cases}\gamma' = \mathrm{T},\\ \mathrm{T}' = \kappa\cdot\mathrm{N},\\ \mathrm{N}' = -\kappa\cdot\mathrm{T} + \tau\cdot\mathrm{B},\\ \mathrm{B}' = -\tau\cdot\mathrm{N}.\end{cases}$$

with the initial condition $\gamma(s_0)$ and an oriented orthonormal frame $\mathrm{T}(s_0)$, $\mathrm{N}(s_0)$, $\mathrm{B}(s_0)$. (The system of equations has four vector equations, so it can be rewritten as a system of 12 scalar equations.)

By 0.25, this system has a unique solution which is defined in a maximal subinterval $\mathbb{J} \subset \mathbb{I}$ containing s_0. Let us show that actually $\mathbb{J} = \mathbb{I}$.

Observe that

$$\langle \mathrm{T}, \mathrm{T} \rangle' = \langle \mathrm{N}, \mathrm{N} \rangle' = \langle \mathrm{B}, \mathrm{B} \rangle' = \langle \mathrm{T}, \mathrm{N} \rangle' = \langle \mathrm{T}, \mathrm{B} \rangle' = \langle \mathrm{B}, \mathrm{T} \rangle' = 0.$$

Indeed,

$$\begin{aligned}
\langle \mathrm{T}, \mathrm{T} \rangle' &= 2\cdot\langle \mathrm{T}, \mathrm{T}' \rangle = 2\cdot\kappa\cdot\langle \mathrm{T}, \mathrm{N} \rangle = 0,\\
\langle \mathrm{N}, \mathrm{N} \rangle' &= 2\cdot\langle \mathrm{N}, \mathrm{N}' \rangle = -2\cdot\kappa\cdot\langle \mathrm{N}, \mathrm{T} \rangle + 2\cdot\tau\cdot\langle \mathrm{N}, \mathrm{B} \rangle = 0,\\
\langle \mathrm{B}, \mathrm{B} \rangle' &= 2\cdot\langle \mathrm{B}, \mathrm{B}' \rangle = -2\cdot\tau\langle \mathrm{B}, \mathrm{N} \rangle = 0,\\
\langle \mathrm{T}, \mathrm{N} \rangle' &= \langle \mathrm{T}', \mathrm{N} \rangle + \langle \mathrm{T}, \mathrm{N}' \rangle = \kappa\cdot\langle \mathrm{N}, \mathrm{N} \rangle - \kappa\cdot\langle \mathrm{T}, \mathrm{T} \rangle + \tau\cdot\langle \mathrm{T}, \mathrm{B} \rangle = 0,\\
\langle \mathrm{N}, \mathrm{B} \rangle' &= \langle \mathrm{N}', \mathrm{B} \rangle + \langle \mathrm{N}, \mathrm{B}' \rangle = 0,\\
\langle \mathrm{B}, \mathrm{T} \rangle' &= \langle \mathrm{B}', \mathrm{T} \rangle + \langle \mathrm{B}, \mathrm{T}' \rangle = -\tau\cdot\langle \mathrm{N}, \mathrm{T} \rangle + \kappa\cdot\langle \mathrm{B}, \mathrm{N} \rangle = 0.
\end{aligned}$$

It follows that the values $\langle \mathrm{T}, \mathrm{T} \rangle$, $\langle \mathrm{N}, \mathrm{N} \rangle$, $\langle \mathrm{B}, \mathrm{B} \rangle$, $\langle \mathrm{T}, \mathrm{N} \rangle$, $\langle \mathrm{T}, \mathrm{N} \rangle$, $\langle \mathrm{B}, \mathrm{T} \rangle$ are constant functions of s. Since we choose $\mathrm{T}(s_0)$, $\mathrm{N}(s_0)$, $\mathrm{B}(s_0)$ to be an oriented orthonormal frame, we have that the triple $\mathrm{T}(s)$, $\mathrm{N}(s)$, $\mathrm{B}(s)$ is an oriented orthonormal frame for any s. In particular, $|\gamma'(s)| = 1$ for all s.

Assume $\mathbb{J} \subsetneq \mathbb{I}$. Then an end of $\mathbb{J}$, say a, lies in the interior of $\mathbb{I}$. By Theorem 0.25, at least one of the values $\gamma(s)$, $\mathrm{T}(s)$, $\mathrm{N}(s)$, $\mathrm{B}(s)$ escapes to infinity as $s \to a$. But this is impossible since the vectors $\mathrm{T}(s)$, $\mathrm{N}(s)$, $\mathrm{B}(s)$ remain unit and $|\gamma'(s)| = |\mathrm{T}(s)| = 1$ — a contradiction. Hence $\mathbb{J} = \mathbb{I}$.

Now assume there are two curves γ_1 and γ_2 with the given curvature and torsion functions. Applying a motion of the space we can assume that $\gamma_1(s_0) = \gamma_2(s_0)$ and the Frenet frames of the curves coincide at s_0. Then $\gamma_1 = \gamma_2$ by the uniqueness of solutions to the system (0.25). Hence the last statement follows. □

4.11. Exercise. *Assume a curve $\gamma\colon \mathbb{R} \to \mathbb{R}^3$ has constant curvature and torsion. Show that γ is a helix, possibly degenerate to a circle; that is, in a suitable coordinate system we have*

$$\gamma(t) = (a\cdot \cos t, a\cdot \sin t, b\cdot t)$$

for some constants a and b.

4.12. Advanced exercise. *Let γ be a smooth regular space curve such that the distance $|\gamma(t) - \gamma(t + \ell)|$ depends only on ℓ. Show that γ is a helix, possibly degenerate to a line or a circle.*

Chapter 5

Signed curvature

A Definitions

Suppose γ is a smooth unit-speed plane curve, so $\mathrm{T}(s) = \gamma'(s)$ is its unit tangent vector for any s.

Let us rotate $\mathrm{T}(s)$ by the angle $\frac{\pi}{2}$ counterclockwise; denote the obtained vector by $\mathrm{N}(s)$. The pair $\mathrm{T}(s), \mathrm{N}(s)$ is an oriented orthonormal frame in the plane which is analogous to the Frenet frame defined in Section 4A; we will keep the name *Frenet frame* for it.

Recall that $\gamma''(s) \perp \gamma'(s)$ (see 3.1). Therefore

❶
$$\mathrm{T}'(s) = k(s)\cdot\mathrm{N}(s).$$

for some real number $k(s)$; the value $k(s)$ is called *signed curvature* of γ at s. We may use notation $k(s)_\gamma$ if we need to specify the curve γ.

Note that

$$\kappa(s) = |k(s)|;$$

that is, up to sign, the signed curvature $k(s)$ equals the curvature $\kappa(s)$ of γ at s defined in Section 3B; the sign tells us in which direction γ turns — if γ is turning left at time s, then $k(s) > 0$. If we want to emphasise that we are working with the *non-signed* curvature of the curve, we call it *absolute curvature.*

Note that if we reverse the parametrization of γ or change the orientation of the plane, then the signed curvature changes its sign.

Since $\mathrm{T}(s), \mathrm{N}(s)$ is an orthonormal frame, we have

$$\langle \mathrm{T}, \mathrm{T}\rangle = 1, \qquad \langle \mathrm{N}, \mathrm{N}\rangle = 1, \qquad \langle \mathrm{T}, \mathrm{N}\rangle = 0,$$

Differentiating these identities we get

$$\langle \mathrm{T}', \mathrm{T}\rangle = 0, \qquad \langle \mathrm{N}', \mathrm{N}\rangle = 0, \qquad \langle \mathrm{T}', \mathrm{N}\rangle + \langle \mathrm{T}, \mathrm{N}'\rangle = 0,$$

By ❶, $\langle \mathrm{T}', \mathrm{N}\rangle = k$ and therefore $\langle \mathrm{T}, \mathrm{N}'\rangle = -k$. Hence we get

❷ $$\mathrm{N}'(s) = -k(s)\cdot \mathrm{T}(s).$$

The equations ❶ and ❷ are the Frenet formulas for plane curves. They can be written in matrix form as:

$$\begin{pmatrix} \mathrm{T}' \\ \mathrm{N}' \end{pmatrix} = \begin{pmatrix} 0 & k \\ -k & 0 \end{pmatrix} \cdot \begin{pmatrix} \mathrm{T} \\ \mathrm{N} \end{pmatrix}.$$

5.1. Exercise. *Let $\gamma_0\colon [a,b] \to \mathbb{R}^2$ be a smooth regular curve and* T *its tangent indicatrix. Consider another curve $\gamma_1\colon [a,b] \to \mathbb{R}^2$ defined by $\gamma_1(t) = \gamma_0(t) + \mathrm{T}(t)$. Show that*

$$\operatorname{length}\gamma_0 \leqslant \operatorname{length}\gamma_1.$$

The curves γ_0 and γ_1 in the exercise above describe the tracks of an idealized bicycle with distance 1 from the rear to the front wheel. Thus by the exercise, the front wheel must have a longer track. For more on the geometry of bicycle tracks, see the survey of Robert Foote, Mark Levi, and Serge Tabachnikov [31] and the references therein.

B Fundamental theorem of plane curves

5.2. Theorem. *Let $k(s)$ be a smooth real valued function defined on a real interval $\mathbb{I}$. Then there is a smooth unit-speed curve $\gamma\colon \mathbb{I} \to \mathbb{R}^2$ with signed curvature $k(s)$. Moreover, γ is uniquely defined up to a rigid motion of the plane.*

This is the fundamental theorem of plane curves; it is a direct analog of 4.10 and it can be proved along the same lines. We present a slightly simpler proof.

Proof. Fix $s_0 \in \mathbb{I}$. Consider the function

$$\theta(s) = \int\limits_{s_0}^{s} k(t)\cdot dt.$$

Note that by the fundamental theorem of calculus, we have $\theta'(s) = k(s)$ for all s.

Set

$$\mathrm{T}(s) = (\cos[\theta(s)], \sin[\theta(s)])$$

and let $\mathrm{N}(s)$ be its counterclockwise rotation by angle $\frac{\pi}{2}$; so

$$\mathrm{N}(s) = (-\sin[\theta(s)], \cos[\theta(s)]).$$

Consider the curve

$$\gamma(s) = \int\limits_{s_0}^{s} \mathrm{T}(s)\cdot ds.$$

Since $|\gamma'| = |\mathrm{T}| = 1$, the curve γ is unit-speed and T, N is its Frenet frame.

Note that

$$\begin{aligned}\gamma''(s) &= \mathrm{T}'(s) = \\ &= (\cos[\theta(s)]', \sin[\theta(s)]') = \\ &= \theta'(s)\cdot(-\sin[\theta(s)], \cos[\theta(s)]) = \\ &= k(s)\cdot\mathrm{N}(s).\end{aligned}$$

So $k(s)$ is the signed curvature of γ at s.

This proves the existence; it remains to prove the uniqueness.

Assume γ_1 and γ_2 are two curves that satisfy the assumptions of the theorem. Applying a rigid motion, we can assume that $\gamma_1(s_0) = \gamma_2(s_0) = 0$ and the Frenet frame of both curves at s_0 is formed by the coordinate frame $(1,0), (0,1)$. Let us denote by $\mathrm{T}_1, \mathrm{N}_1$ and $\mathrm{T}_2, \mathrm{N}_2$ the Frenet frames of γ_1 and γ_2 respectively. Both triples $\gamma_i, \mathrm{T}_i, \mathrm{N}_i$ satisfy the following system of ordinary differential equations

$$\begin{cases}\gamma_i' = \mathrm{T}_i, \\ \mathrm{T}_i' = k\cdot\mathrm{N}_i, \\ \mathrm{N}_i' = -k\cdot\mathrm{T}_i.\end{cases}$$

Moreover, they have the same initial values at s_0. By the uniqueness of solutions of ordinary differential equations (0.25), we have $\gamma_1 = \gamma_2$. □

Note that from the proof of the theorem we obtain the following corollary:

5.3. Corollary. *Let $\gamma\colon \mathbb{I} \to \mathbb{R}^2$ be a smooth unit-speed curve and $s_0 \in \mathbb{I}$. Denote by k the signed curvature of γ. Assume an oriented (x,y)-coordinate system is chosen in such a way that $\gamma(s_0)$ is the origin and $\gamma'(s_0)$ points in the direction of the x-axis. Then*

$$\gamma'(s) = (\cos[\theta(s)], \sin[\theta(s)]),$$

for all s, where

$$\theta(s) = \int\limits_{s_0}^{s} k(t)\cdot dt.$$

C Total signed curvature

Let $\gamma\colon \mathbb{I}\to\mathbb{R}^2$ be a smooth unit-speed plane curve. The *total signed curvature* of γ, denoted by $\Psi(\gamma)$, is defined as the integral

❶ $$\Psi(\gamma)=\int\limits_{\mathbb{I}} k(s)\cdot ds,$$

where k denotes the signed curvature of γ.

Note that if $\mathbb{I}=[a,b]$, then

❷ $$\Psi(\gamma)=\theta(b)-\theta(a),$$

where θ is as in 5.3.

If γ is a piecewise smooth and regular plane curve, then we define its total signed curvature as the sum of the total signed curvatures of its arcs plus the sum of the *signed* external angles at its joints; they are positive where γ turns left, negative where γ turns right, and 0 where γ goes straight. It is undefined if it turns exactly backward; that is, if the curve has a cusp. That is, if γ is a concatenation of smooth and regular arcs $\gamma_1,\dots,\gamma_n$, then

$$\Psi(\gamma)=\Psi(\gamma_1)+\dots+\Psi(\gamma_n)+\theta_1+\dots+\theta_{n-1}$$

where θ_i is the signed external angle at the joint between γ_i and γ_{i+1}. If γ is closed, then the concatenation is cyclic and

$$\Psi(\gamma)=\Psi(\gamma_1)+\dots+\Psi(\gamma_n)+\theta_1+\dots+\theta_n,$$

where θ_n is the signed external angle at the joint between γ_n and γ_1.

Since $\left|\int k(s)\cdot ds\right|\leqslant\int|k(s)|\cdot ds$, we have

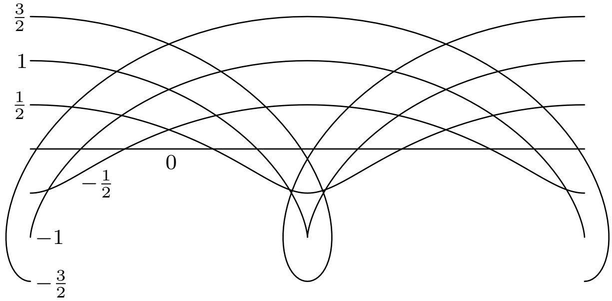
❸ $$|\Psi(\gamma)|\leqslant\Phi(\gamma)$$

for any smooth regular plane curve γ; that is, the total signed curvature Ψ cannot exceed the total curvature Φ in absolute value. Note that the equality holds if and only if the signed curvature does not change sign.

5.4. Exercise. *A trochoid is a curve traced out by a point fixed to a wheel as it rolls along a straight line. A family of* trochoids $\gamma_a\colon[0,2\cdot\pi]\to\mathbb{R}^2$ *(see the picture) can be parametrized as*

$$\gamma_a(t)=(t+a\cdot\sin t, a\cdot\cos t).$$

(a) Given $a\in\mathbb{R}$, find $\Psi(\gamma_a)$ if it is defined.
(b) Given $a\in\mathbb{R}$, find $\Phi(\gamma_a)$.



5.5. Proposition. *Any closed simple smooth regular plane curve γ has total signed curvature $\pm 2\cdot\pi$; it is $+2\cdot\pi$ if the region bounded by γ lies on the left from it and $-2\cdot\pi$ otherwise.*

Moreover the same statement holds for any closed piecewise simple smooth regular plane curve γ if its total signed curvature is defined.

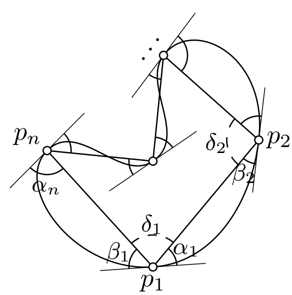
This proposition is called sometimes *Umlaufsatz*; it is a differential-geometric analog of the theorem about the sum of the internal angles of a polygon (0.9) which we use in the proof. A more conceptual proof was given by Heinz Hopf [41], [42, p. 42].

Proof. Without loss of generality we may assume that γ is oriented in such a way that the region bounded by γ lies on the left from it. We can also assume that it is parametrized by arc-length.

Consider a closed polygonal line $\beta = p_1 \dots p_n$ inscribed in γ. We can assume that the arcs between the vertices are sufficiently small so that the polygonal line is simple and each arc γ_i from p_i to p_{i+1} has small total absolute curvature, say $\Phi(\gamma_i) < \pi$ for each i.

As usual we use indexes modulo n, in particular $p_{n+1} = p_1$. Assume $p_i = \gamma(t_i)$. Set

$$\mathrm{w}_i = p_{i+1} - p_i, \qquad \mathrm{v}_i = \gamma'(t_i),$$
$$\alpha_i = \measuredangle(\mathrm{v}_i, \mathrm{w}_i), \qquad \beta_i = \measuredangle(\mathrm{w}_{i-1}, \mathrm{v}_i),$$



where $\alpha_i, \beta_i \in (-\pi, \pi]$ are signed angles — α_i is positive if w_i points to the left from v_i.

By ❷, the value

❹ $$\Psi(\gamma_i) - \alpha_i - \beta_{i+1}$$

is a multiple of $2\cdot\pi$. Since $\Phi(\gamma_i) < \pi$, by the chord lemma (3.16), we also have that $|\alpha_i| + |\beta_i| < \pi$. By ❸, we have that $|\Psi(\gamma_i)| \leqslant \Phi(\gamma_i)$; therefore the value in ❹ vanishes. In other word,for each i we have

$$\Psi(\gamma_i) = \alpha_i + \beta_{i+1}.$$

Note that

❺ $$\delta_i = \pi - \alpha_i - \beta_i$$

is the internal angle of β at p_i; $\delta_i \in (0, 2\cdot\pi)$ for each i. Recall that the sum of the internal angles of an n-gon is $(n-2)\cdot\pi$ (see 0.9); that is,

$$\delta_1 + \dots + \delta_n = (n-2)\cdot\pi.$$

Therefore

❻ $$\begin{aligned}\Psi(\gamma) &= \Psi(\gamma)_1 + \dots + \Psi(\gamma)_n = \\ &= (\alpha_1 + \beta_2) + \dots + (\alpha_n + \beta_1) = \\ &= (\beta_1 + \alpha_1) + \dots + (\beta_n + \alpha_n) = \\ &= (\pi - \delta_1) + \dots + (\pi - \delta_n) = \\ &= n\cdot\pi - (n-2)\cdot\pi = \\ &= 2\cdot\pi.\end{aligned}$$

The case of piecewise smooth and regular curves is done the same way; we need to subdivide the arcs in the cyclic concatenation further to meet the requirement above and instead of equation ❺ we have

$$\delta_i = \pi - \alpha_i - \beta_i - \theta_i,$$

where θ_i is the signed external angle of γ at p_i; it vanishes if the curve γ is smooth at p_i. Therefore instead of equation ❻, we have

$$\begin{aligned}\Psi(\gamma) &= \Psi(\gamma_1) + \dots + \Psi(\gamma_n) + \theta_1 + \dots + \theta_n = \\ &= (\alpha_1 + \beta_2) + \dots + (\alpha_n + \beta_1) = \\ &= (\beta_1 + \alpha_1 + \theta_1) + \dots + (\beta_n + \alpha_n + \theta_n) = \\ &= (\pi - \delta_1) + \dots + (\pi - \delta_n) = \\ &= n\cdot\pi - (n-2)\cdot\pi = \\ &= 2\cdot\pi.\end{aligned}$$

□

5.6. Exercise. *Draw a smooth regular closed plane curve γ such that*
(a) $\Psi(\gamma) = 0$;
(b) $\Psi(\gamma) = \Phi(\gamma) = 10\cdot\pi$;
(c) $\Psi(\gamma) = 2\cdot\pi$ *and* $\Phi(\gamma) = 4\cdot\pi$.

5.7. Exercise. *Let $\gamma\colon [a,b] \to \mathbb{R}$ be a smooth regular plane curve with Frenet frame T, N. Given a real parameter ℓ, consider the curve $\gamma_\ell(t) = \gamma(t) + \ell\cdot\mathrm{N}(t)$; it is called a* parallel curve of γ at signed distance ℓ.

(a) Show that γ_ℓ *is a regular curve if* $\ell\cdot k(t) \neq 1$ *for all* t*, where* $k(t)$ *denotes the signed curvature of* γ*.*

(b) Set $L(\ell) = \mathrm{length}\,\gamma_\ell$*. Show that*

$$L(\ell) = L(0) - \ell\cdot\Psi(\gamma) \qquad ❼$$

for all ℓ *sufficiently close to* 0*.*

(c) Describe an example showing that formula ❼ does not hold for all ℓ*.*

D Osculating circline

5.8. Proposition. *Given a point* p*, a unit vector* T *and a real number* k*, there is a unique smooth unit-speed curve* $\sigma\colon\mathbb{R}\to\mathbb{R}^2$ *that starts at* p *in the direction of* T *and has constant signed curvature* k*.*

Moreover, if $k=0$*, then it is a line* $\sigma(s) = p + s\cdot\mathrm{T}$*; if* $k\neq 0$*, then* σ *runs around a circle of radius* $\frac{1}{|k|}$ *with center at* $p+\frac{1}{k}\cdot\mathrm{N}$*, where* T,N *is an oriented orthonormal frame.*

Further we will use the term *circline* for *a circle or a line*; these are the only plane curves with constant signed curvature.

Proof. The proof is done by a calculation based on 5.2 and 5.3.

Suppose $s_0 = 0$, choose a coordinate system such that p is its origin and T points in the direction of the x-axis. Therefore N points in the direction of the y-axis. Then

$$\begin{aligned}\theta(s) &= \int_0^s k\cdot dt = \\ &= k\cdot s.\end{aligned}$$

Therefore

$$\sigma'(s) = (\cos[k\cdot s], \sin[k\cdot s]).$$

It remains to integrate the last identity. If $k=0$, we get

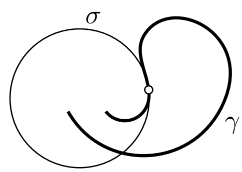
$$\sigma(s) = (s, 0)$$

which describes the line $\sigma(s) = p + s\cdot\mathrm{T}$.

If $k\neq 0$, we get

$$\sigma(s) = (\tfrac{1}{k}\cdot\sin[k\cdot s], \tfrac{1}{k}\cdot(1-\cos[k\cdot s])).$$

which is the circle of radius $r = \frac{1}{|k|}$ centered at $(0,\frac{1}{k}) = p + \frac{1}{k}\cdot\mathrm{N}$. $\square$



5.9. Definition. *Let γ be a smooth unit-speed plane curve; denote by $k(s)$ the signed curvature of γ at s.*

The unit-speed curve σ of constant signed curvature $k(s)$ that starts at $\gamma(s)$ in the direction $\gamma'(s)$ is called the osculating circline *of γ at s.*

The center and radius of the osculating circle at a given point are called the center of curvature *and* radius of curvature *of the curve at that point.*

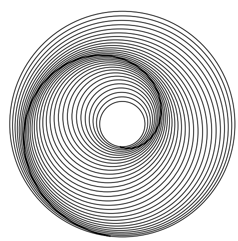
The osculating circle σ_s can be also defined as the unique circline that has *second order of contact* with γ at s; that is, $\rho(\ell) = o(\ell^2)$, where $\rho(\ell)$ denotes the distance from $\gamma(s+\ell)$ to σ_s.

The following exercise is recommended to the reader familiar with the notion of *inversion.*

5.10. Advanced exercise. *Suppose γ is a smooth regular plane curve that does not pass thru the origin. Let $\hat{\gamma}$ be the inversion of γ with respect to the unit circle centered at the origin. Show that the osculating circline of $\hat{\gamma}$ at s is the inversion of the osculating circline of γ at s.*

E Spiral lemma

The following lemma was proved by Peter Tait [76] and later rediscovered by Adolf Kneser [46].



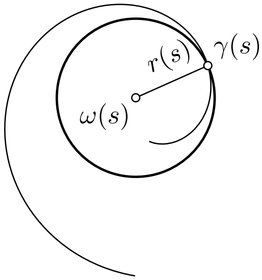
5.11. Lemma. *Assume that γ is a smooth regular plane curve with strictly decreasing positive signed curvature. Then the osculating circles of γ are nested; that is, if σ_s denoted the osculating circle of γ at s, then σ_{s_0} lies in the open disc bounded by σ_{s_1} for any $s_0 < s_1$.*

It turns out that the osculating circles of the curve γ give a peculiar foliation of an annulus by circles; it has the following property: if a smooth function is constant on each osculating circle, then it must be constant in the annulus [see 32, Lecture 10]. Also note that the curve γ is tangent to a circle of the foliation at each of its points. However, it does not run along any of those circles.

Proof. Let $\mathrm{T}(s), \mathrm{N}(s)$ be the Frenet frame, $\omega(s)$, $r(s)$ the center and radius of curvature of γ. By 5.8,

$$\omega(s) = \gamma(s) + r(s)\cdot\mathrm{N}(s).$$

Since $k > 0$, we have that $r(s)\cdot k(s) = 1$. Therefore applying Frenet formula ❷, we get that



$$\begin{aligned}\omega'(s) &= \gamma'(s) + r'(s)\cdot\mathrm{N}(s) + r(s)\cdot\mathrm{N}'(s) = \\ &= \mathrm{T}(s) + r'(s)\cdot\mathrm{N}(s) - r(s)\cdot k(s)\cdot\mathrm{T}(s) = \\ &= r'(s)\cdot\mathrm{N}(s).\end{aligned}$$

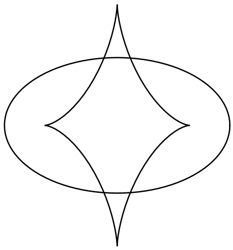
Since $k(s)$ is decreasing, $r(s)$ is increasing; therefore $r' \geqslant 0$. It follows that $|\omega'(s)| = r'(s)$ and $\omega'(s)$ points in the direction of $\mathrm{N}(s)$.

Since $\mathrm{N}'(s) = -k(s)\cdot\mathrm{T}(s)$, the direction of $\omega'(s)$ cannot have constant direction on a nontrivial interval; that is, the curve $s \mapsto \omega(s)$ contains no line segments. Therefore

$$\begin{aligned}|\omega(s_1) - \omega(s_0)| &< \mathrm{length}(\omega|_{[s_0,s_1]}) = \\ &= \int_{s_0}^{s_1} |\omega'(s)|\cdot ds = \\ &= \int_{s_0}^{s_1} r'(s)\cdot ds = \\ &= r(s_1) - r(s_0).\end{aligned}$$

In other words, the distance between the centers of σ_{s_1} and σ_{s_0} is strictly less than the difference between their radii. Therefore the osculating circle at s_0 lies inside the osculating circle at s_1 without touching it. □

The curve $s \mapsto \omega(s)$ is called the *evolute* of γ; it traces the centers of curvature of the curve. The evolute of γ can be written as



$$\omega(t) = \gamma(t) + \tfrac{1}{k(t)}\cdot\mathrm{N}(t)$$

and in the proof we showed that $(\tfrac{1}{k})'\cdot\mathrm{N}$ is its velocity vector.

5.12. Exercise. *Show that the stretched astroid*

$$\omega(t) = (\tfrac{a^2-b^2}{a}\cdot\cos^3 t, \tfrac{b^2-a^2}{b}\cdot\sin^3 t)$$

is an evolute of the ellipse defined by

$$\gamma(t) = (a\cdot\cos t, b\cdot\sin t).$$

The following theorem states formally that *if you drive on the plane and turn the steering wheel to the left all the time, then you will not be able to come back to the place you started.*

5.13. Theorem. *Assume γ is a smooth regular plane curve with positive and strictly monotonic signed curvature. Then γ is simple.*

The same statement holds true without assuming positivity of curvature; the proof requires only minor modifications.

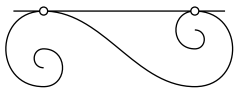
Proof of 5.13. Note that $\gamma(s)$ lies on the osculating circle σ_s of γ at s. If $s_1 \neq s_0$, then by lemma 5.11, σ_{s_0} does not intersect σ_{s_1}. Therefore $\gamma(s_1) \neq \gamma(s_0)$, hence the result. □

5.14. Exercise. *Show that a 3-dimensional analog of the theorem does not hold. That is, there are self-intersecting smooth regular space curves with strictly monotonic curvature.*

5.15. Exercise. *Assume that γ is a smooth regular plane curve with positive strictly monotonic signed curvature.*

(a) Show that no line can be tangent to γ at two distinct points.

(b) Show that no circle can be tangent to γ at three distinct points.



Note that part *(a)* does not hold if we allow the curvature to be negative; an example is shown on the diagram.

Chapter 6

Supporting curves

A Cooriented tangent curves

Suppose γ_1 and γ_2 are smooth regular plane curves. Recall that the curves γ_1 and γ_2 are tangent at the time parameters t_1 and t_2 if $\gamma_1(t_1) = \gamma_2(t_2)$ and they share the tangent line at these time parameters. In this case the point $p = \gamma_1(t_1) = \gamma_2(t_2)$ is called a *point of tangency* of the curves. If both curves are simple, then without ambiguity we may say that γ_1 and γ_2 are tangent at the point p.

Note that if γ_1 and γ_2 are tangent at the time parameters t_1 and t_2, then the velocity vectors $\gamma_1'(t_1)$ and $\gamma_2'(t_2)$ are parallel. If $\gamma_1'(t_1)$ and $\gamma_2'(t_2)$

have the same direction, we say that the curves are *cooriented*, if their directions are opposite, the curves are called *counteroriented*.

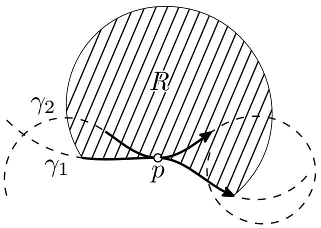
Note that reverting the parametrization of one of the curves, cooriented curves become counteroriented and vice versa; so we can always assume the curves are cooriented at any given point of tangency.

B Supporting curves

Let γ_1 and γ_2 be two smooth regular plane curves that share a point

$$p = \gamma_1(t_1) = \gamma_2(t_2)$$

which is not an endpoint of any of the curves. Suppose that there is $\varepsilon > 0$ such that the arc $\gamma_2|_{[t_2-\varepsilon,t_2+\varepsilon]}$ lies in a closed plane region R with the arc $\gamma_1|_{[t_1-\varepsilon,t_1+\varepsilon]}$ in its boundary, then we say that γ_1 *locally supports* γ_2 at the time parameters t_1 and t_2. If both curves are simple, then without ambiguity we also could say that γ_1 *locally supports* γ_2 at the point p.



If the curves on the diagram are oriented according to the arrows, then γ_1 supports γ_2 from the right at p (as well as γ_2 supports γ_1 from the left at p).

Suppose γ_1 is a simple proper curve, so it divides the plane into two closed regions that lie on the left and right from γ_1. Then we say that γ_1 *globally supports* γ_2 at t_2 if γ_2 runs in one of these closed regions and $\gamma_2(t_2)$ lies on γ.

Further, suppose γ_2 is closed, so it divides the plane into two regions. We say that a point lies *inside* (respectively, *outside*) of γ_2 if it lies in the bounded region (respectively, unbounded) region. In this case we say that γ_1 supports γ_2 *from inside* (*from outside*) if γ_1 supports γ_2 and lies in the region inside it (respectively outside it).

Note that if γ_1 and γ_2 share a point $p = \gamma_1(t_1) = \gamma_2(t_2)$ and are not tangent at t_1 and t_2, then at time t_2 the curve γ_2 crosses γ_1 moving from one of its sides to the other. It follows that γ_1 cannot locally support γ_2 at the time parameters t_1 and t_2. Whence we get the following:

6.1. Definition-Observation. *Let γ_1 and γ_2 be two smooth regular plane curves. Suppose γ_1 locally supports γ_2 at time parameters t_1 and t_2. Then γ_1 is tangent to γ_2 at t_1 and t_2.*

In such a case, if the curves are cooriented and the region R in the definition of supporting curves lies on the right (left) from the arc of γ_1, then we say that γ_1 supports γ_2 from the left (respectively right).

We say that a smooth regular plane curve γ has a *vertex* at s if the signed curvature function is critical at s; that is, if $k'(s)_\gamma = 0$. If in addition to that, γ is simple, we could say that the point $p = \gamma(s)$ is a vertex of γ.

6.2. Exercise. *Assume that the osculating circle σ_s of a smooth regular simple plane curve γ locally supports γ at $p = \gamma(s)$. Show that p is a vertex of γ.*

C Supporting test

The following proposition resembles the second derivative test.

6.3. Proposition. *Let γ_1 and γ_2 be two smooth regular plane curves.*

Suppose γ_1 locally supports γ_2 from the left (right) at the time parameters t_1 and t_2. Then

$$k_1(t_1) \leqslant k_2(t_2) \quad (\textit{respectively} \quad k_1(t_1) \geqslant k_2(t_2)).$$

where k_1 and k_2 denote the signed curvature of γ_1 and γ_2 respectively.

A partial converse also holds. Namely, if γ_1 and γ_2 tangent and cooriented at the time parameters t_1 and t_2 then γ_1 locally supports γ_2 from the left (right) at the time parameters t_1 and t_2 if

$$k_1(t_1) < k_2(t_2) \quad (\textit{respectively} \quad k_1(t_1) > k_2(t_2)).$$

Proof. Without loss of generality, we can assume that $t_1 = t_2 = 0$, the shared point $\gamma_1(0) = \gamma_2(0)$ is the origin and the velocity vectors $\gamma_1'(0)$, $\gamma_2'(0)$ point in the direction of x-axis.

Note that small arcs of $\gamma_1|_{[-\varepsilon,+\varepsilon]}$ and $\gamma_2|_{[-\varepsilon,+\varepsilon]}$ can be described as a graph $y = f_1(x)$ and $y = f_2(x)$ for smooth functions f_1 and f_2 such that $f_i(0) = 0$ and $f_i'(0) = 0$. Note that $f_1''(0) = k_1(0)$ and $f_2''(0) = k_2(0)$ (see 3.6)

Clearly, γ_1 supports γ_2 from the left (right) if

$$f_1(x) \leqslant f_2(x) \quad (\text{respectively} \quad f_1(x) \geqslant f_2(x))$$

for all sufficiently small values of x. Applying the second derivative test, we get the result. □

6.4. Advanced exercise. *Let γ_0 and γ_1 be two smooth unit-speed simple plane curves that are tangent and cooriented at the point $p = \gamma_0(0) = \gamma_1(0)$. Assume $k_0(s) \leqslant k_1(s)$ for any s. Show that γ_0 locally supports γ_1 from the right at p.*

Give an example of two proper curves γ_0 and γ_1 satisfying the above condition such that γ_0 does not support γ_1 at p globally.

Note that according to 3.13 for any closed smooth regular curve that runs in a unit disc, the average of its absolute curvature is at least 1; in particular, there is a point where the absolute curvature is at least 1. The following exercise says that this statement also holds for loops.

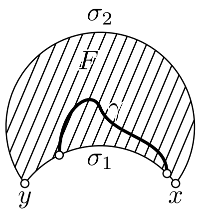
6.5. Exercise. *Assume a closed smooth regular plane loop γ runs in a unit disc. Show that there is a point on γ with absolute curvature at least 1.*

6.6. Exercise. *Assume a closed smooth regular plane curve γ runs between two parallel lines at distance 2 from each other. Show that there is a point on γ with absolute curvature at least 1.*

Try to prove the same for a smooth regular plane loop.

6.7. Exercise. *Assume a closed smooth regular plane curve γ runs inside of a triangle $\triangle$ with inradius 1; that is, the circle tangent to all three sides of $\triangle$ has radius 1. Show that there is a point on γ with absolute curvature at least* 1.

The three exercises above are baby cases of 6.15; try to find a direct solution, without using 6.14.



6.8. Exercise. *Let F be a plane figure bounded by two circle arcs σ_1 and σ_2 of signed curvature 1 that run from x to y. Suppose σ_1 is shorter than σ_2. Assume a smooth regular arc γ runs in F and has both endpoints on σ_1. Show that the absolute curvature of γ is at least 1 at some parameter value.*

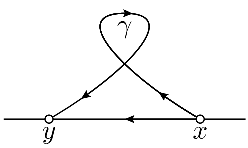
D Convex curves

Recall that a plane curve is convex if it bounds a convex region.

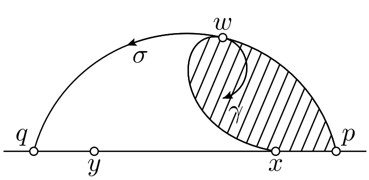
6.9. Proposition. *Suppose that a closed simple plane curve γ bounds a figure F. Then F is convex if and only if the signed curvature of γ does not change sign.*

6.10. Lens lemma. *Let γ be a smooth regular simple plane curve that runs from x to y. Assume that γ runs on the right side (left side) of the oriented line xy and only its endpoints x and y lie on the line. Then γ has a point with positive (respectively negative) signed curvature.*

Note that the lemma fails for curves with self-intersections. For example, the curve γ on the diagram always turns right, so it has negative curvature everywhere, but it lies on the right side of the line xy.

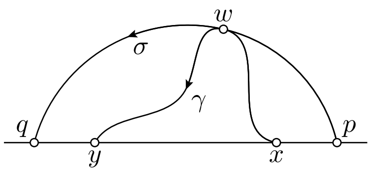


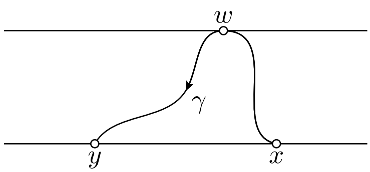
Proof. Choose points p and q on the line xy so that the points p, x, y, q appear in that order. We can assume that p and q lie sufficiently far from x and y, so that the half-disc with diameter pq contains γ.



Consider the smallest disc segment with chord $[p,q]$ that contains γ. Note that its arc σ supports γ at some point $w = \gamma(t_0)$.

Let us parameterize σ from p to q. Note that the γ and σ are tangent and cooriented at w. If not, then the arc of γ from w to y would be trapped in the curvilinear triangle xwp bounded by the line segment $[p,x]$ and the arcs of σ, γ. But this is impossible since y does not belong to this triangle.





It follows that σ supports γ at t_0 from the right. By 6.3,

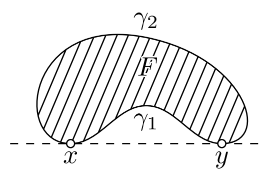
$$k(w)_\gamma \geqslant k_\sigma > 0.$$

□

Remark. Instead of taking the minimal disc segment, one can take a point w on γ that maximizes the distance to the line xy. The same argument shows that the curvature at w is nonnegative, which is slightly weaker than the required positive curvature.

Proof of 6.9; "only-if" part. If F is convex, then every tangent line of γ supports γ. If a point moves along γ, the figure F has to stay on one side from its tangent line; that is, we can assume that each tangent line supports γ on one side, say on the right. Since a line has vanishing curvature, the supporting test (6.3) implies that $k \geqslant 0$ at each point.

"If" part. Denote by K the convex hull of F. If F is not convex, then F is a proper subset of K. Therefore ∂K contains a line segment that is not a part of ∂F. In other words, there is a line that supports γ at two points, say x and y that divide γ in two arcs γ_1 and γ_2, both distinct from the line segment $[x,y]$.



Note that one of the arcs γ_1 or γ_2 is parametrized from x to y and the other from y to x. Passing to a smaller arc if necessary we can ensure that only its endpoints lie on the line. Applying the lens lemma, we get that the arcs γ_1 and γ_2 contain points with signed curvatures of opposite signs. □

6.11. Exercise. *Suppose γ is a smooth regular simple closed convex plane curve of diameter larger than 2. Show that γ has a point with absolute curvature less than 1.*

6.12. Exercise. *Suppose γ is a simple smooth regular plane curve with positive signed curvature. Assume γ crosses a line ℓ at the points $p_1, p_2, \dots$ $\dots, p_n$ and these points appear on γ in the same order.*

(a) Show that p_2 cannot lie between p_1 and p_3 on ℓ.

(b) Show that if p_3 lies between p_1 and p_2 on ℓ, then the points appear on ℓ in the following order:

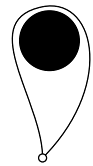
$$p_1, p_3, \dots, p_4, p_2.$$

(c) Describe all possible orders of the points p_i on ℓ.

E Moon in a puddle

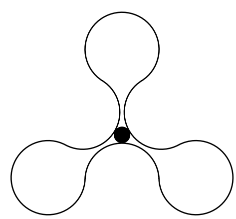
The following theorem is a slight generalization of the theorem proved by Vladimir Ionin and German Pestov in [43]. For convex curves, this result was known earlier [11, §24].

6.13. Theorem. *Assume γ is a simple closed smooth regular plane loop with absolute curvature bounded by 1. Then it surrounds a unit disc.*



This theorem gives a simple but nontrivial example of the so-called *local to global theorems* — based on some local data (in this case the curvature of a curve) we conclude a global property (in this case existence of a large disc surrounded by the curve).

A straightforward approach would be to start with some disc in the region bounded by the curve and blow it up to maximize its radius. However, as one may see from the spinner-like example on the diagram it does not always lead to a solution — a closed plane curve of curvature at most 1 may surround a disc of radius smaller than 1 that cannot be enlarged continuously.

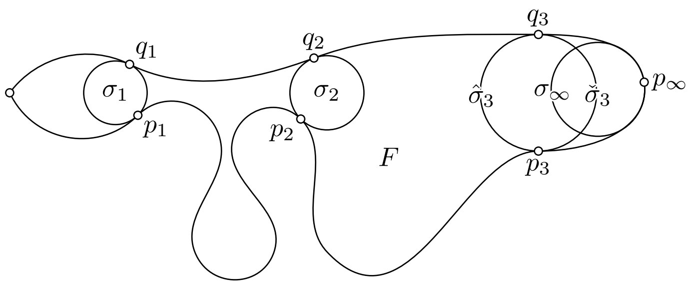


6.14. Key lemma. *Assume γ is a simple closed smooth regular plane loop. Then at one point of γ (distinct from its base) its osculating circle σ globally supports γ from the inside.*

First let us show that the theorem follows from the lemma.

Proof of 6.13 modulo 6.14. Since γ has absolute curvature at most 1, each osculating circle has radius at least 1. According to the key lemma one of the osculating circles σ globally supports γ from inside. In particular σ lies inside of γ, whence the result. $\square$

Proof of 6.14. Denote by F the closed region surrounded by γ. We can assume that F lies on the left from γ. Arguing by contradiction, assume that the osculating circle at each point $p \in \gamma$ does not lie in F.



Given a point $p \in \gamma$ let us consider the maximal circle σ that lies completely in F and tangent to γ at p. The circle σ will be called the *incircle* of F at p.

Note that the curvature k_σ of the incircle σ has to be larger than $k(p)_\gamma$. Indeed, since σ supports γ from the left, by 6.3 we have $k_\sigma \geqslant k(p)_\gamma$; in the case of equality, σ is the osculating circle at p. The latter is impossible by our assumption.

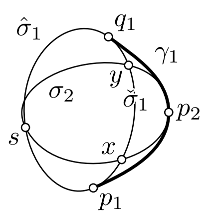
It follows that σ has to touch γ at another point. Otherwise we can increase σ slightly while keeping it inside F.

Indeed, since $k_\sigma > k(p)_\gamma$, by 6.3 we can choose a neighborhood U of p such that after a slight increase of σ, the intersection $U \cap \sigma$ is still in F. On the other hand, if σ does not touch γ at another point, then after some (maybe smaller) increase of σ the complement $\sigma \backslash U$ is still in F. That is, a slightly increased σ is still in F — a contradiction.

Choose a point p_1 on γ that is distinct from its base point. Let σ_1 be the incircle at p_1. Denote by γ_1 an arc of γ from p_1 to a first point q_1 on σ_1. Denote by $\hat\sigma_1$ and $\check\sigma_1$ two arcs of σ_1 from p_1 to q_1 such that the cyclic concatenation of $\hat\sigma_1$ and γ_1 surrounds $\check\sigma_1$.

Let p_2 be the midpoint of γ_1. Denote by σ_2 the incircle at p_2.

Note that σ_2 cannot intersect $\hat\sigma_1$. Otherwise, if σ_2 intersects $\hat\sigma_1$ at some point s, then σ_2 has to have two more common points with $\check\sigma_1$, say x and y — one for each arc of σ_2 from p_2 to s. Therefore $\sigma_1 = \sigma_2$ since these two circles have three common points: s, x, and y. On the other hand, by construction, $p_2 \in \sigma_2$ and $p_2 \notin \sigma_1$ — a contradiction.



Two ovals pretend to be circles.

Recall that σ_2 has to touch γ at another point. From above it follows that it can only touch γ_1 and therefore we can choose an arc $\gamma_2 \subset \gamma_1$ that runs from p_2 to a first point q_2 on σ_2. Since p_2 is the midpoint of γ_1, we have that

$$\text{length}\,\gamma_2 < \tfrac{1}{2}\cdot \text{length}\,\gamma_1. \qquad ❶$$

Repeating this construction recursively, we get an infinite sequence of arcs $\gamma_1 \supset \gamma_2 \supset \dots$; by ❶, we also get that

$$\operatorname{length}\gamma_n \to 0 \quad \text{as} \quad n \to \infty.$$

Therefore the intersection

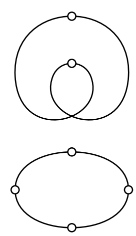
$$\bigcap_n \gamma_n$$

contains a single point; denote it by p_∞.

Let σ_∞ be the incircle at p_∞; it has to touch γ at another point, say q_∞. The same argument as above shows that $q_\infty \in \gamma_n$ for any n. It follows that $q_\infty = p_\infty$ — a contradiction. ☐

6.15. Exercise. *Assume that a closed smooth regular curve γ lies in a figure F bounded by a closed simple plane curve. Suppose that R is the maximal radius of discs that lie in F. Show that absolute curvature of γ is at least $\frac{1}{R}$ at some parameter value.*

F Four-vertex theorem



Recall that a vertex of a smooth regular curve is defined as a critical point of its signed curvature; in particular, any local minimum (or maximum) of the signed curvature is a vertex. For example, every point of a circle is a vertex.

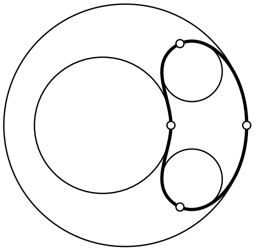
6.16. Four-vertex theorem. *Any smooth regular simple plane curve has at least four vertices.*

Evidently any closed smooth regular curve has at least two vertices — where the minimum and the maximum of the curvature are attained. On the diagram the vertices are marked; the first curve has one self-intersection and exactly two vertices; the second curve has exactly four vertices and no self-intersections.

The four-vertex theorem was first proved by Syamadas Mukhopadhyaya [61] for convex curves. Many different proofs and generalizations are known. One of our favorite proofs was given by Robert Osserman [63]. We give a proof of the following stronger statement based on the key lemma in the previous section.

6.17. Theorem. *Any smooth regular simple plane curve is globally supported by its osculating circle at least at 4 distinct points; two from inside and two from outside.*

Proof of 6.16 modulo 6.17. First note that if an osculating circline σ at a point p supports γ locally, then p is a vertex. Indeed, if not, then a small arc around p has monotonic curvature.



Applying the spiral lemma (5.11) we get that the osculating circles at this arc are nested. In particular the curve γ crosses σ at p and therefore σ does not locally support γ at p. $\square$

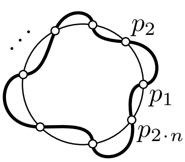
Proof of 6.17. According to the key lemma (6.14), there is a point $p \in \gamma$ such that its osculating circle supports γ from inside. The curve γ can be considered as a loop with the base at p. Therefore the key lemma implies the existence of another point $q \in \gamma$ with the same property.

It shows the existence of two osculating circles that support γ from inside; it remains to show existence of two osculating circles that support γ from outside.

In order to get the osculating circles supporting γ from outside, one can repeat the proof of key lemma taking instead of incircle the circline of maximal signed curvature that supports the curve from outside, assuming that γ is oriented so that the region on the left from it is bounded.

(Alternatively, if one applies to γ an inversion with respect to a circle whose center lies inside γ, then the obtained curve γ_1 also has two osculating circles that support γ_1 from inside. According to 5.10, these osculating circlines are inverses of the osculating circlines of γ. Note that the region lying inside of γ is mapped to the region outside of γ_1 and the other way around. Therefore these two circlines correspond to the osculating circlines supporting γ from outside.) $\square$

6.18. Advanced exercise. *Suppose γ is a closed simple smooth regular plane curve and σ is a circle. Assume γ crosses σ at the points $p_1, \dots, p_{2\cdot n}$ and these points appear in the same cycle order on γ and on σ. Show that γ has at least $2\cdot n$ vertices.*



Construct an example of a closed simple smooth regular plane curve γ with only 4 vertices that crosses a given circle at arbitrarily many points.

Chapter 7

Definitions

A Topological surfaces

We will be mostly interested in smooth regular surfaces defined in the following section. However we will sometimes use the following general definition.

A connected subset Σ in the Euclidean space $\mathbb{R}^3$ is called a *topological surface* (more precisely an *embedded surface without boundary*) if any point of $p \in \Sigma$ admits a neighborhood W in Σ that can be parametrized by an open subset in the Euclidean plane; that is, there is an injective continuous map $U \to W$ from an open set $U \subset \mathbb{R}^2$ such that its inverse $W \to U$ is also continuous.

B Smooth surfaces

Recall that a function f of two variables x and y is called *smooth* if all its partial derivatives $\frac{\partial^{m+n}}{\partial x^m \partial y^n} f$ are defined and are continuous in the domain of definition of f.

A connected set $\Sigma \subset \mathbb{R}^3$ is called a *smooth surface* (we use it as a shortcut for the more precise term *smooth regular embedded surface*) if it can be described locally as a graph of a smooth function in an appropriate coordinate system.

More precisely, for any point $p \in \Sigma$ one can choose a coordinate system (x, y, z) and a neighborhood $U \ni p$ such that the intersection $W = U \cap \Sigma$ is a graph $z = f(x, y)$ of a smooth function f defined in an open domain of the (x, y)-plane.

Examples. The simplest example of a smooth surface is the (x, y)-plane

$$\Pi = \left\{ (x, y, z) \in \mathbb{R}^3 \ : \ z = 0 \right\}.$$

The plane Π is a surface since it can be described as the graph of the function $f(x,y)=0$.

All other planes are smooth surfaces as well since one can choose a coordinate system so that it becomes the (x,y)-plane. We may also present a plane as a graph of a linear function $f(x,y)=a{\cdot}x+b{\cdot}y+c$ for some constants a, b and c (assuming the plane is not perpendicular to the (x,y)-plane, in which case a different coordinate system is required to write the plane as the graph of a function).

A more interesting example is the unit sphere

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

This set is not the graph of any function, but $\mathbb{S}^2$ is locally a graph; it can be covered by the following 6 graphs:

$$\begin{aligned}
z&=f_\pm(x,y)=\pm\sqrt{1-x^2-y^2},\\
y&=g_\pm(x,z)=\pm\sqrt{1-x^2-z^2},\\
x&=h_\pm(y,z)=\pm\sqrt{1-y^2-z^2},
\end{aligned}$$

where each function $f_\pm, g_\pm, h_\pm$ is defined in an open unit disc. Any point $p\in\mathbb{S}^2$ lies in one of these graphs therefore $\mathbb{S}^2$ is a surface. Since each function is smooth, so is the surface $\mathbb{S}^2$.

C Surfaces with boundary

A connected subset in a surface that is bounded by one or more curves is called a *surface with boundary*; such curves form the *boundary line* of the surface.

When we say *surface* we usually mean a *smooth regular surface without boundary*; we may use the term *surface without boundary* if we need to emphasize it; otherwise we may use the term *surface with possibly nonempty boundary.*

D Proper, closed and open surfaces

If the surface Σ is formed by a closed set, then it is called *proper*. For example, for any smooth function f defined on the whole plane, its graph $z=f(x,y)$ is a proper surface. The sphere $\mathbb{S}^2$ gives another example of proper surface.

On the other hand, the open disc

$$\left\{\,(x,y,z)\in\mathbb{R}^3\;:\;x^2+y^2<1,\;z=0\,\right\}$$

is not proper; this set is neither open nor closed.

A compact surface without boundary is called *closed* (this term is closely related to *closed curve* but has nothing to do with *closed set*).

A proper noncompact surface without boundary is called *open* (again the term *open set* is not relevant).

For example, the paraboloid $z = x^2 + y^2$ is an open surface; the sphere $\mathbb{S}^2$ is a closed surface.

Note that any proper surface without boundary is either closed or open.

The following claim is a three-dimensional analog of the plane separation theorem (1.10). Despite it might look obvious, its proof is not trivial at all; a standard proof uses the so-called *Alexander's duality* which is a classical technique in algebraic topology [see 39]. We omit its proof since it would take us far away from the main subject.

7.1. Claim. *The complement of any proper topological surface without boundary (or, equivalently any open or closed topological surface) has exactly two connected components.*

E Implicitly defined surfaces

7.2. Proposition. *Let $f\colon \mathbb{R}^3 \to \mathbb{R}$ be a smooth function. Suppose that 0 is a regular value of f; that is, $\nabla_p f \neq 0$ at any point p such that $f(p) = 0$. Then any connected component Σ of the set of solutions of the equation $f(x, y, z) = 0$ is a smooth surface.*

Proof. Fix $p \in \Sigma$. Since $\nabla_p f \neq 0$ we have

$$f_x(p) \neq 0, \quad f_y(p) \neq 0, \quad \text{or} \quad f_z(p) \neq 0.$$

We may assume that $f_z(p) \neq 0$; otherwise permute the coordinates x, y, z.

The implicit function theorem (0.22) implies that a neighborhood of p in Σ is the graph $z = h(x, y)$ of a smooth function h defined on an open domain in $\mathbb{R}^2$. It remains to apply the definition of smooth surface (Section 7B). $\square$

7.3. Exercise. *For which constants ℓ does the following equation*

$$x^2 + y^2 - z^2 = \ell$$

describe a smooth regular surface.

F Local parametrizations

Let U be an open domain in $\mathbb{R}^2$ and $s\colon U \to \mathbb{R}^3$ be a smooth map. We say that s is regular if its Jacobian has maximal rank; in this case it means that the vectors s_u and s_v are linearly independent at any $(u,v) \in U$; equivalently $s_u \times s_v \neq 0$, where $\times$ denotes the vector product.

7.4. Proposition. *If $s\colon U \to \mathbb{R}^3$ is a smooth regular embedding of an open connected set $U \subset \mathbb{R}^2$, then its image $\Sigma = s(U)$ is a smooth surface.*

Proof. Set

$$s(u,v) = (x(u,v), y(u,v), z(u,v)).$$

Since s is regular, its Jacobian matrix

$$\operatorname{Jac} s = \begin{pmatrix} x_u & x_v \\ y_u & y_v \\ z_u & z_v \end{pmatrix}$$

has rank two at any pint $(u,v) \in U$.

Choose a point $p \in \Sigma$; by shifting the (x,y,z) and (u,v) coordinate systems we may assume that p is the origin and $p = s(0,0)$. Permuting the coordinates x, y, z if necessary, we may assume that the matrix

$$\begin{pmatrix} x_u & x_v \\ y_u & y_v \end{pmatrix},$$

is invertible at the origin. Note that this is the Jacobian matrix of the map

$$(u,v) \mapsto (x(u,v), y(u,v)).$$

The inverse function theorem (0.21) implies that there is a smooth regular map $w\colon (x,y) \mapsto (u,v)$ defined on an open set $W \ni 0$ in the (x,y)-plane such that $w(0,0) = (0,0)$ and $s \circ w(x,y) = (x, y, f(x,y))$ for some smooth function f. That is, the graph $z = f(x,y)$ for $(x,y) \in W$ is a subset in Σ. By the inverse function theorem this graph is open in Σ.

Since p is arbitrary, we get that Σ is a surface. $\square$

If we have s and Σ as in the proposition, then we say that s is a *smooth parametrization* of the surface Σ.

Not all the smooth surfaces can be described by such a parametrization; for example the sphere $\mathbb{S}^2$ cannot. However, any smooth surface Σ admits a local parametrization; that is, any point $p \in \Sigma$ admits an open neighborhood $W \subset \Sigma$ with a smooth regular parametrization s. In this case any point in W can be described by two parameters, usually denoted by u and v, which are called *local coordinates* at p. The map s is called a *chart* of Σ.

If W is a graph $z = h(x,y)$ of a smooth function h, then the map

$$s\colon (u,v) \mapsto (u,v,h(u,v))$$

is a chart. Indeed, s has an inverse $(u,v,h(u,v)) \mapsto (u,v)$ which is continuous; that is, s is an embedding. Further, $s_u = (1,0,h_u)$ and $s_v = (0,1,h_v)$. Whence the partial derivatives s_u and s_v are linearly independent; that is, s is a regular map.

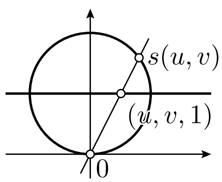
Note that from 7.4, we obtain the following corollary.

7.5. Corollary. *A connected set $\Sigma \subset \mathbb{R}^3$ is a smooth regular surface if and only if a neighborhood of any point in Σ can be covered by a chart.*

7.6. Exercise. *Consider the following map*

$$s(u,v) = \left(\tfrac{2\cdot u}{1+u^2+v^2}, \tfrac{2\cdot v}{1+u^2+v^2}, \tfrac{2}{1+u^2+v^2}\right).$$

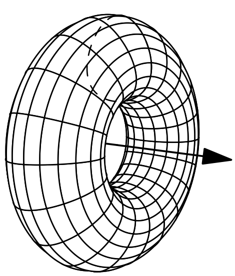
Show that s is a chart of the unit sphere centered at $(0,0,1)$; describe the image of s.



The map s in the exercise can be visualized using the following map

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

which is called *stereographic projection* from the plane $z = 1$ to the unit sphere with center at $(0,0,1)$. Note that the point $(u,v,1)$ and its image lie on the same half-line emerging from the origin.



Let $\gamma(t) = (x(t), y(t))$ be a plane curve. Recall that the *surface of revolution* of the curve γ around the x-axis can be described as the image of the map

$$(t,s) \mapsto (x(t), y(t)\cdot \cos s, y(t)\cdot \sin s).$$

For fixed t or s the obtained curves are called *meridians* or *parallels* of the surface, respectively; note that parallels are formed by circles in the plane perpendicular to the axis of rotation. The curve γ is called the *generatrix* of the surface.

7.7. Exercise. *Assume γ is a closed simple smooth regular plane curve that does not intersect the x-axis. Show that the surface of revolution around the x-axis with generatrix γ is a smooth regular surface.*

G Global parametrizations

A surface can be described by an embedding from a known surface to the space.

For example, consider the ellipsoid

$$\Sigma_{a,b,c} = \left\{ (x,y,z) \in \mathbb{R}^3 \ : \ \tfrac{x^2}{a^2} + \tfrac{y^2}{b^2} + \tfrac{z^2}{c^2} = 1 \right\}$$

for some positive numbers a, b, and c. Note that by 7.2, $\Sigma_{a,b,c}$ is a smooth regular surface. Indeed, set $f(x,y,z) = \tfrac{x^2}{a^2} + \tfrac{y^2}{b^2} + \tfrac{z^2}{c^2}$, then

$$\nabla f(x,y,z) = (\tfrac{2}{a^2}\cdot x, \tfrac{2}{b^2}\cdot y, \tfrac{2}{c^2}\cdot z).$$

Therefore $\nabla f \neq 0$ if $f = 1$; that is, 1 is a regular value of f.

Note that $\Sigma_{a,b,c}$ can be defined as the image of the map $s\colon \mathbb{S}^2 \to \mathbb{R}^3$, defined as the restriction of the following map to the unit sphere $\mathbb{S}^2$:

$$(x,y,z) \mapsto (a\cdot x, b\cdot y, c\cdot z).$$

For a surface Σ, a map $s\colon \Sigma \to \mathbb{R}^3$ is called a *smooth parametrized surface* if s is injective and for any chart $f\colon U \to \Sigma$ the composition $s \circ f$ is smooth and regular; that is, all partial derivatives $\frac{\partial^{m+n}}{\partial u^m \partial v^n}(s \circ f)$ exist and are continuous in the domain of definition and the two vectors $\frac{\partial}{\partial u}(s \circ f)$ and $\frac{\partial}{\partial v}(s \circ f)$ are linearly independent.

Note that in this case the image $\Sigma^* = s(\Sigma)$ is a smooth surface. The latter follows since for any chart $f\colon U \to \Sigma$ the composition $s \circ f\colon U \to \Sigma^*$ is a chart of Σ^*. The map s is called a *diffeomorphism* from Σ to Σ^*; the surfaces Σ and Σ^* are said to be *diffeomorphic* if there is a diffeomorphism $s\colon \Sigma \to \Sigma^*$.

7.8. Advanced exercise. *Show that the surfaces Σ and Θ are diffeomorphic if*

(a) Σ and Θ obtained from the plane by removing n points.
(b) Σ and Θ are open convex subsets of a plane bounded by smooth curves.
(c) Σ and Θ are open convex subsets of a plane.
(d) Σ and Θ are open star-shaped subsets of a plane.

Chapter 8

First order structure

A Tangent plane

8.1. Definition. *Let Σ be a smooth surface. A vector* w *is a* tangent vector *of Σ at p if and only if there is a curve γ that runs in Σ and has* w *as a velocity vector at p; that is, $p = \gamma(t)$ and $\mathrm{w} = \gamma'(t)$ for some t.*

8.2. Proposition-Definition. *Let Σ be a smooth surface and $p \in \Sigma$. Then the set of tangent vectors of Σ at p forms a plane; this plane is called the* tangent plane *of Σ at p.*

Moreover if $s\colon U \to \Sigma$ is a local chart and $p = s(u_p, v_p)$, then the tangent plane of Σ at p is spanned by vectors $s_u(u_p, v_p)$ and $s_v(u_p, v_p)$.

The tangent plane to Σ at p is usually denoted by T_p or $\mathrm{T}_p\Sigma$. This plane T_p might be considered as a linear subspace of $\mathbb{R}^3$ or as a parallel plane passing thru p; the latter is sometimes called the *affine tangent plane*. The affine tangent plane can be interpreted as the best approximation at p of the surface Σ by a plane. More precisely, it has *first order of contact* with Σ at p; that is, $\rho(q) = o(|p - q|)$, where $q \in \Sigma$ and $\rho(q)$ denotes the distance from q to T_p.

Proof. Fix a chart s at p. Assume γ is a smooth curve that starts at p. Without loss of generality, we can assume that γ is covered by the chart; in particular, there are smooth functions $u(t)$ and $v(t)$ such that

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

Applying chain rule, we get

$$\gamma' = s_u{\cdot}u' + s_v{\cdot}v';$$

that is, γ' is a linear combination of s_u and s_v.

Since the smooth functions $u(t)$ and $v(t)$ can be chosen arbitrarily, any linear combination of s_u and s_v is a tangent vector at p. $\square$

8.3. Exercise. *Let $f:\mathbb{R}^3\to\mathbb{R}$ be a smooth function with 0 as a regular value and Σ be a surface described as a connected component of the set of solutions $f(x,y,z)=0$. Show that the tangent plane $\mathrm{T}_p\Sigma$ is perpendicular to the gradient $\nabla_p f$ at any point $p\in\Sigma$.*

8.4. Exercise. *Let Σ be a smooth surface and $p\in\Sigma$. Choose (x,y,z)-coordinates. Show that a neighborhood of p in Σ is a graph $z=f(x,y)$ of a smooth function f defined on an open subset in the (x,y)-plane if and only if the tangent plane T_p is not* vertical*; that is, if T_p is not perpendicular to the (x,y)-plane.*

8.5. Exercise. *Show that if a smooth surface Σ meets a plane Π at a single point p, then Π is tangent to Σ at p.*

B Directional derivative

In this section we extend the definition of directional derivative to smooth functions defined on smooth surfaces.

First let us recall the standard definition of directional derivative.

Suppose f is a function defined at a point p in the space, and w a vector. Consider the function

$$h(t)=f(p+t\cdot\mathrm{w}).$$

Then the directional derivative of f at p along w is defined as

$$D_{\mathrm{w}}f(p):=h'(0).$$

8.6. Proposition-Definition. *Let Σ be a smooth regular surface and f a smooth function defined on Σ. Suppose γ is a smooth curve in Σ that starts at p with velocity vector $\mathrm{w}\in\mathrm{T}_p$; that is, $\gamma(0)=p$ and $\gamma'(0)=\mathrm{w}$. Then the derivative $(f\circ\gamma)'(0)$ depends only on f, p and w; it is called the* directional derivative of f along w at p *and denoted by*

$$D_{\mathrm{w}}f,\quad D_{\mathrm{w}}f(p),\quad \text{or}\quad D_{\mathrm{w}}f(p)_\Sigma$$

— we may omit p and Σ if it is clear from the context.

Moreover, if $(u,v)\mapsto s(u,v)$ is a local chart at p, then

$$D_{\mathrm{w}}f=a\cdot f_u+b\cdot f_v,$$

where $\mathrm{w} = a{\cdot}s_u + b{\cdot}s_v$.

Note that our definition agrees with the standard definition of directional derivative if Σ is a plane. Indeed, in this case $\gamma(t) = p + \mathrm{w}{\cdot}t$ is a curve in Σ that starts at p with velocity vector w. For a general surface the point $p + \mathrm{w}{\cdot}t$ might not lie on the surface; therefore the function f might be undefined at this point; therefore the standard definition does not work.

Proof. Without loss of generality, we may assume that γ is covered by the chart s; if not we can chop γ. In this case

$$\gamma(t) = s(u(t), v(t))$$

for some smooth functions u, v defined in a neighborhood of 0 such that $u(0) = u_p$ and $v(0) = v_p$.

Applying the chain rule, we get that

$$\gamma'(0) = u'(0){\cdot}s_u + v'(0){\cdot}s_v$$

at (u_p, v_p). Since $\mathrm{w} = \gamma'(0)$ and the vectors s_u, s_v are linearly independent, we get that $a = u'(0)$ and $b = v'(0)$.

Applying the chain rule again, we get that

❶ $$(f \circ \gamma)'(0) = a{\cdot}f_u + b{\cdot}f_v.$$

at (u_p, v_p).

Notice that the left hand side in ❶ does not depend on the choice of the chart s and the right hand side depends only on p, w, f, and s. It follows that $(f \circ \gamma)'(0)$ depends only on p, w and f.

The last statement follows from ❶. □

C Tangent vectors as functionals

In this section we introduce a more conceptual way to define tangent vectors. We will not use this approach in the sequel, but it is better to know about it.

A tangent vector $\mathrm{w} \in \mathrm{T}_p$ to a smooth surface Σ defines a linear functional[1] D_{w} that swallows a smooth function φ defined in a neighborhood of p in Σ and spits its directional derivative $D_{\mathrm{w}}\varphi$. It is straightforward to check that the functional D obeys the product rule:

❶ $$D_{\mathrm{w}}(\varphi{\cdot}\psi) = (D_{\mathrm{w}}\varphi){\cdot}\psi(p) + \varphi(p){\cdot}(D_{\mathrm{w}}\psi).$$

[1] Term *functional* is used for functions that take a function as an argument and return a number.

It is not hard to show that the tangent vector w is completely determined by the functional D_{w}. Moreover tangent vectors at p can be *defined* as linear functionals on the space of smooth functions that satisfy the product rule ❶.

This definition grabs the key algebraic property of tangent vectors. It might be a less intuitive way to think about tangent vectors, but it is often convenient to use in the proofs. For example 8.6 becomes a tautology.

D Differential of map

Any smooth map s from a surface Σ to $\mathbb{R}^3$ can be described by its coordinate functions $s(p) = (x(p), y(p), z(p))$. To take a directional derivative of the map we should take the directional derivative of each of its coordinate functions.

$$D_{\mathrm{w}}s := (D_{\mathrm{w}}x, D_{\mathrm{w}}y, D_{\mathrm{w}}z).$$

Assume s is a smooth map from one smooth surface Σ_0 to another Σ_1 and $p \in \Sigma_0$. Note that $D_{\mathrm{w}}s(p) \in \mathrm{T}_{s(p)}\Sigma_1$ for any $\mathrm{w} \in \mathrm{T}_p$. Indeed, choose a curve γ_0 in Σ_0 such that $\gamma_0(0) = p$ and $\gamma_0'(0) = \mathrm{w}$. Observe that $\gamma_1 = s \circ \gamma_0$ is a smooth curve in Σ_1 and by the definition of directional derivative, we have $D_{\mathrm{w}}s(p) = \gamma_1'(0)$. It remains to note that $\gamma_1(0) = s(p)$ and therefore its velocity $\gamma_1'(0)$ is in $\mathrm{T}_{s(p)}\Sigma_1$.

Recall that 8.6 implies that $d_p s\colon \mathrm{w} \mapsto D_{\mathrm{w}}s$ defines a linear map $d_p s\colon \mathrm{T}_p\Sigma_0 \to \mathrm{T}_{s(p)}\Sigma_1$; that is,

$$D_{c\cdot\mathrm{w}}s = c\cdot D_{\mathrm{w}}s(p) \quad \text{and} \quad D_{\mathrm{v}+\mathrm{w}}s = D_{\mathrm{v}}s(p) + D_{\mathrm{w}}s(p)$$

for any $c \in \mathbb{R}$ and $\mathrm{v}, \mathrm{w} \in \mathrm{T}_p$. The map $d_p s$ is called the *differential* of s at p.

The differential $d_p s$ can be described by a $2{\times}2$-matrix M in orthonormal bases of T_p and $\mathrm{T}_{s(p)}\Sigma_1$. Set $\operatorname{jac}_p s = |\det M|$; this value does not depend on the choice of orthonormal bases in T_p and $\mathrm{T}_{s(p)}\Sigma_1$.

Let $s_1 : \Sigma_1 \to \Sigma_2$ be another smooth map between smooth surfaces Σ_1 and Σ_2. Suppose that $p_1 = s(p) \in \Sigma_1$; observe that

$$d_p(s_1 \circ s) = d_{p_1}s_1 \circ d_p s.$$

It follows that

❶ $$\operatorname{jac}_p(s_1 \circ s) = \operatorname{jac}_{p_1} s_1 \cdot \operatorname{jac}_p s.$$

If Σ_0 is a domain in the (u, v)-plane, then the value $\operatorname{jac}_p s$ can be found using the following formulas

$$\begin{aligned}\operatorname{jac} s &= |s_v \times s_u| = \\ &= \sqrt{\langle s_u, s_u\rangle\cdot\langle s_v, s_v\rangle - \langle s_u, s_v\rangle^2} = \\ &= \sqrt{\det[\operatorname{Jac}^\top s\cdot \operatorname{Jac} s]}.\end{aligned}$$

where $\operatorname{Jac} s$ denotes the Jacobian matrix of s; it is a $2{\times}3$ matrix with column vectors s_u and s_v.

The value $\operatorname{jac}_p s$ has the following geometric meaning: if P_0 is a region in T_p and $P_1 = (d_p s)(P_0)$, then

$$\operatorname{area} P_1 = \operatorname{jac}_p s\cdot \operatorname{area} P_0.$$

This identity will become important in the definition of surface area.

E Surface integral and area

Let Σ be a smooth surface and $h\colon \Sigma \to \mathbb{R}$ be a smooth function. Let us define the integral $\iint_R h$ of the function h along a region $R \subset \Sigma$. The definition will be used mostly along surfaces with boundary, but the definition can be applied to any Borel set $R \subset \Sigma$.

Recall that $\operatorname{jac}_p s$ is defined in the previous section. Assume that there is a chart $(u, v) \mapsto s(u, v)$ of Σ defined on an open set $U \subset \mathbb{R}^2$ such that $R \subset s(U)$. In this case set

❶ $$\iint\limits_R h := \iint\limits_{s^{-1}(R)} h \circ s(u, v)\cdot \operatorname{jac}_{(u,v)} s\cdot du\cdot dv.$$

By the substitution rule (0.23), the right hand side in ❶ does not depend on the choice of s. That is, if $s_1 : U_1 \to \Sigma$ is another chart such that $s_1(U_1) \supset R$, then

$$\iint\limits_{s^{-1}(R)} h \circ s(u, v)\cdot \operatorname{jac}_{(u,v)} s\cdot du\cdot dv = \iint\limits_{s_1^{-1}(R)} h \circ s_1(u, v)\cdot \operatorname{jac}_{(u,v)} s_1\cdot du\cdot dv.$$

In other words, the defining identity ❶ makes sense.

A general region R can be subdivided into regions $R_1, R_2 \dots$ such that each R_i lies in the image of some chart. After that one could define the integral along R as the sum

$$\iint\limits_R h := \iint\limits_{R_1} h + \iint\limits_{R_2} h + \dots$$

It is straightforward to check that the value $\iint_R h$ does not depend on the choice of such subdivision.

The area of a region R in a smooth surface Σ is defined as the surface integral

$$\operatorname{area} R = \iint\limits_R 1.$$

The following proposition provides a substitution rule for surface integral.

8.7. Area formula. *Suppose $s\colon \Sigma_0 \to \Sigma_1$ is a smooth parametrization of a smooth surface Σ_1 by a smooth surface Σ_0. Then for any region $R \subset \Sigma_0$ and any smooth function $f\colon \Sigma_1 \to \mathbb{R}$ we have*

$$\iint\limits_R (f \circ s)\cdot \operatorname{jac} s = \int\limits_{s(R)} f.$$

In particular, if $f \equiv 1$, we have

$$\iint\limits_R \operatorname{jac} s = \operatorname{area}[s(R)].$$

Proof. Follows from ❶ and the definition of surface integral. □

Remark. The notion of area of a surface is closely related to the length of a curve. However, to define length we use a different idea — it was defined as the least upper bound on the lengths of inscribed polygonal lines. It turns out that an analogous definition does not work even for very simple surfaces. The latter is shown by a classical example — the so-called *Schwarz's boot.* This example and different approaches to the notion of area are discussed in a popular article of Vladimir Dubrovsky [26].

F Normal vector and orientation

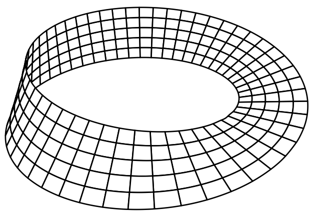
A unit vector that is normal to T_p is usually denoted by $\nu(p)$; it is uniquely defined up to sign.

A surface Σ is called *oriented* if it is equipped with a unit normal vector field ν; that is, a continuous map $p \mapsto \nu(p)$ such that $\nu(p) \perp \mathrm{T}_p$ and $|\nu(p)| = 1$ for any p. The choice of the field ν is called the *orientation* of Σ. A surface Σ is called *orientable* if it can be oriented. Note that each orientable surface admits two orientations: ν and $-\nu$.

Let Σ be a smooth oriented surface with unit normal field ν. The map $\nu\colon \Sigma \to \mathbb{S}^2$ defined by $p \mapsto \nu(p)$ is called the *spherical map* or *Gauss map*.

For surfaces, the spherical map plays essentially the same role as the tangent indicatrix for curves.

The Möbius strip shown on the diagram gives an example of a nonorientable surface — there is no choice of normal vector field that is continuous along the middle of the strip (it changes the sign if you try to go around).



Note that each surface is locally orientable. In fact each chart $s(u,v)$ admits an orientation

$$\nu = \frac{s_u \times s_v}{|s_u \times s_v|}.$$

Indeed, the vectors s_u and s_v are tangent vectors at p; since they are linearly independent, their vector product does not vanish and it is perpendicular to the tangent plane. Evidently $(u,v) \mapsto \nu(u,v)$ is a continuous map. Therefore ν is a unit normal field.

8.8. Exercise. *Let $h : \mathbb{R}^3 \to \mathbb{R}$ be a smooth function with 0 as a regular value and Σ a surface described as a connected component of the set of solutions $h(x,y,z) = 0$. Show that Σ is orientable.*

Recall that any proper surface without boundary in the Euclidean space divides it into two connected components (7.1). Therefore we can choose the unit normal field on any smooth proper surface that points into one of the components of the complement. Therefore we obtain the following observation.

8.9. Observation. *Any smooth proper surface in the Euclidean space is oriented.*

In particular it follows that the Möbius strip cannot be extended to a proper smooth surface without boundary.

G Sections

8.10. Advanced exercise. *Let Π be the (x,y)-plane and $A \subset \Pi$ be any closed subset. Show that there is an open smooth regular surface Σ with $\Sigma \cap \Pi = A$.*

The exercise above says that plane sections of a smooth regular surface might look complicated. The following lemma makes it possible to perturb the plane so that the section becomes nice.

8.11. Lemma. *Let Σ be a smooth regular surface. Suppose $f\colon \mathbb{R}^3 \to \mathbb{R}$ is a smooth function. Then for any constant r_0 there is an arbitrarily close value r such that each connected component of the intersection of the level set $L_r = f^{-1}\{r\}$ with Σ is a smooth regular curve.*

Proof. The surface Σ can be covered by a countable set of charts $s_i\colon U_i \to \to \Sigma$. Note that the composition $f \circ s_i$ is a smooth function for any i. By Sard's lemma (0.20), almost all real numbers r are regular values for each $f \circ s_i$.

Fix such a value r sufficiently close to r_0 and consider the level set L_r described by the equation $f(x, y, z) = r$. Any point in the intersection $\Sigma \cap L_r$ lies in the image of one of the charts. From above it admits a neighborhood which is a regular smooth curve; hence the result. $\square$

8.12. Corollary. *Let Σ be a smooth surface. Then for any plane Π there is a parallel plane Π^* that lies arbitrary close to Π and such that the intersection $\Sigma \cap \Pi^*$ is a union of disjoint smooth curves.*

Chapter 9

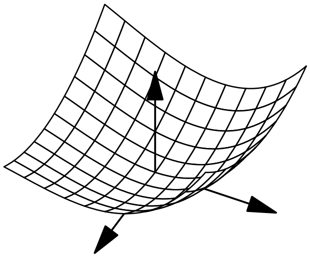
Curvatures

A Tangent-normal coordinates

Fix a point p in a smooth oriented surface Σ. Consider a coordinate system (x, y, z) with origin at p such that the (x, y)-plane coincides with T_p and the z-axis points in the direction of the normal vector $\nu(p)$. By 8.4, we can present Σ locally at p as a graph $z = f(x, y)$ of a smooth function. Note that

$$f(0,0) = 0, \qquad f_x(0,0) = 0, \qquad f_y(0,0) = 0.$$

The first equality holds since $p = (0, 0, 0)$ lies on the graph and the other two equalities mean that the tangent plane at p is horizontal.



Set

$$\begin{aligned} \ell &= f_{xx}(0,0), \\ m &= f_{xy}(0,0) = f_{yx}(0,0), \\ n &= f_{yy}(0,0). \end{aligned}$$

The *Taylor series* for f at $(0, 0)$ up to the second order term can be then written as

$$f(x, y) = \tfrac{1}{2}(\ell{\cdot}x^2 + 2{\cdot}m{\cdot}x{\cdot}y + n{\cdot}y^2) + o(x^2 + y^2).$$

Note that values ℓ, m, and n are completely determined by this equation. The so-called *osculating paraboloid*

$$z = \tfrac{1}{2}{\cdot}(\ell{\cdot}x^2 + 2{\cdot}m{\cdot}x{\cdot}y + n{\cdot}y^2)$$

has *second order of contact* with Σ at p.

Note that

$$\ell\cdot x^2 + 2\cdot m\cdot x\cdot y + n\cdot y^2 = \langle M_p\cdot\left(\begin{smallmatrix} x\\ y\end{smallmatrix}\right), \left(\begin{smallmatrix} x\\ y\end{smallmatrix}\right)\rangle,$$

where M_p is the so-called *Hessian matrix* of f at $(0,0)$,

❶ $$M_p = \begin{pmatrix} \ell & m \\ m & n \end{pmatrix}.$$

B Principal curvatures

Note that tangent-normal coordinates give an almost canonical coordinate system in a neighborhood of p; it is unique up to a rotation of the (x,y)-plane. Rotating the (x,y)-plane results in rewriting the matrix M_p in the new basis.

Since the Hessian matrix M_p is symmetric, by the spectral theorem (0.15) it is diagonalizable by orthogonal matrices. That is, by rotating the (x,y)-plane we can assume that $m = 0$ in ❶; see 0.15. In this case

$$M_p = \begin{pmatrix} k_1 & 0 \\ 0 & k_2 \end{pmatrix},$$

the diagonal components k_1 and k_2 of M_p are called the *principal curvatures* of Σ at p; they might also be denoted as $k_1(p)$ and $k_2(p)$, or $k_1(p)_\Sigma$ and $k_2(p)_\Sigma$; if we need to emphasize that we compute them at the point p for the surface Σ. We will always assume that $k_1 \leqslant k_2$.

Note that if $x = f(x,y)$ is a local graph representation of Σ in these coordinates, then

$$f(x,y) = \tfrac{1}{2}\cdot(k_1\cdot x^2 + k_2\cdot y^2) + o(x^2 + y^2).$$

The principal curvatures can be also defined as the eigenvalues of the Hessian matrix M_p. The eigendirections of M_p are called the *principal directions* of Σ at p. Note that if $k_1(p) \neq k_2(p)$, then p has exactly two principal directions, which are perpendicular to each other; if $k_1(p) = k_2(p)$ then all tangent directions at p are principal.

Note that if we revert the orientation of Σ, then the principal curvatures switch their signs and indexes.

A smooth regular curve on a surface Σ that always runs in the principal directions is called a *line of curvature* of Σ.

9.1. Exercise. *Assume that a smooth surface Σ is mirror symmetric with respect to a plane Π. Suppose that Σ and Π intersect along a smooth regular curve γ. Show that γ is a line of curvature of Σ.*

C More curvatures

Let p be a point on an oriented smooth surface Σ.

The product

$$K(p) = k_1(p)\cdot k_2(p)$$

is called the *Gauss curvature* at p. We may denote it by K, $K(p)$, or $K(p)_\Sigma$ if we need to specify the point p and the surface Σ.

Since the determinant is equal to the product of the eigenvalues, we get

$$K = \ell\cdot n - m^2,$$

where $M_p = \left(\begin{smallmatrix} \ell & m \\ m & n \end{smallmatrix}\right)$ is the Hessian matrix.

9.2. Exercise. *Show that any surface with positive Gauss curvature is orientable.*

The sum

$$H(p) = k_1(p) + k_2(p)$$

is called the *mean curvature*[1] at p. We may also denote it by $H(p)_\Sigma$. The mean curvature can be also interpreted as the trace of the Hessian matrix $M_p = \left(\begin{smallmatrix} \ell & m \\ m & n \end{smallmatrix}\right)$; that is,

$$H = \ell + n$$

A surface with vanishing mean curvature is called *minimal.*

Note that reverting the orientation of Σ does not change the Gauss curvature, but changes the sign of the mean curvature. In particular, the Gauss curvature is well defined for a nonoriented surfaces.

D Shape operator

In the following definitions we use the notion of directional derivative defined in 8.6.

Let p be a point on a smooth surface Σ with orientation defined by the unit normal field ν. Given $\mathrm{w} \in \mathrm{T}_p$, its *shape operator* is defined by

$$\mathrm{Shape}_p\, \mathrm{w} = -D_{\mathrm{w}}\nu.$$

Equivalently, the shape operator can be defined by

➊ $$\mathrm{Shape} = -d\nu,$$

[1]Some authors define the mean curvature as $\frac{1}{2}\cdot(k_1(p) + k_2(p))$ — the mean value of the principal curvatures. It suits the name better, but it is not as convenient when it comes to calculations.

where $d\nu$ denotes the differential of the spherical map $\nu\colon\Sigma\to\mathbb{S}^2$; that is, $d_p\nu(\mathrm{v})=(D_{\mathrm{v}}\nu)(p)$.

Recall that $d_p\nu$ is a linear map $\mathrm{T}_p\Sigma\to\mathrm{T}_{\nu(p)}\mathbb{S}^2$. Note that $\mathrm{T}_p\Sigma$ coincides with $\mathrm{T}_{\nu(p)}\mathbb{S}^2$ — both of them are normal subspaces to $\nu(p)$. Therefore Shape_p is indeed a linear operator $\mathrm{T}_p\to\mathrm{T}_p$ (the latter also follows from 9.3).

For a point $p\in\Sigma$ the shape operator of a tangent vector $\mathrm{w}\in\mathrm{T}_p$ will be denoted by $\mathrm{Shape}\,\mathrm{w}$ if it is clear from the context which base point p and which surface we are working with; otherwise we may use the notations

$$\mathrm{Shape}_p(\mathrm{w})\quad\text{or}\quad\mathrm{Shape}_p(\mathrm{w})_\Sigma$$

as the situation requires. [2]

9.3. Theorem. *Suppose that $(u,v)\mapsto s(u,v)$ is a smooth map to a smooth surface Σ with unit normal field ν. Then*

$$\begin{aligned}\langle\mathrm{Shape}(s_u),s_u\rangle&=\langle s_{uu},\nu\rangle, & \langle\mathrm{Shape}(s_v),s_u\rangle&=\langle s_{uv},\nu\rangle,\\ \langle\mathrm{Shape}(s_u),s_v\rangle&=\langle s_{uv},\nu\rangle, & \langle\mathrm{Shape}(s_v),s_v\rangle&=\langle s_{vv},\nu\rangle,\\ \langle\mathrm{Shape}(s_u),\nu\rangle&=0, & \langle\mathrm{Shape}(s_v),\nu\rangle&=0\end{aligned}$$

for any (u,v).

Proof. We will use the shortcut $\nu=\nu(u,v)$ for $\nu(s(u,v))$, so

❷ $$\mathrm{Shape}(s_u)=-D_{s_u}\nu=-\nu_u,\quad \mathrm{Shape}(s_v)=-D_{s_v}\nu=-\nu_v.$$

Note that ν is a unit vector orthogonal to s_u and s_v; therefore

$$\langle\nu,s_u\rangle\equiv0,\qquad\langle\nu,s_v\rangle\equiv0,\qquad\langle\nu,\nu\rangle\equiv1.$$

Taking partial derivatives of these two identities we get

$$\begin{aligned}\langle\nu_u,s_u\rangle+\langle\nu,s_{uu}\rangle&=0, & \langle\nu_v,s_u\rangle+\langle\nu,s_{uv}\rangle&=0,\\ \langle\nu_u,s_v\rangle+\langle\nu,s_{uv}\rangle&=0, & \langle\nu_v,s_v\rangle+\langle\nu,s_{vv}\rangle&=0,\\ 2\cdot\langle\nu_u,\nu\rangle&=0, & 2\cdot\langle\nu_v,\nu\rangle&=0.\end{aligned}$$

It remains to plug in the expressions from ❷. □

9.4. Exercise. *Show that the shape operator is* self-adjoint*; that is,*

$$\langle\mathrm{Shape}\,\mathrm{u},\mathrm{v}\rangle=\langle\mathrm{u},\mathrm{Shape}\,\mathrm{v}\rangle$$

[2]The following bilinear forms on a tangent plane

$$\mathrm{I}(\mathrm{v},\mathrm{w})=\langle\mathrm{v},\mathrm{w}\rangle,\qquad\mathrm{II}(\mathrm{v},\mathrm{w})=\langle\mathrm{Shape}\,\mathrm{v},\mathrm{w}\rangle,\qquad\mathrm{III}(\mathrm{v},\mathrm{w})=\langle\mathrm{Shape}\,\mathrm{v},\mathrm{Shape}\,\mathrm{w}\rangle$$

are called the *first, second, and third fundamental forms* respectively. These forms were introduced before the shape operator, but we will not touch them in the sequel.

for any $\mathrm{u}, \mathrm{v} \in \mathrm{T}_p$.

Let us denote by i, j and k the standard basis in the $\mathbb{R}^3$. Recall that the components ℓ, m, and n of the Hessian matrix are defined in Section 9A.

9.5. Corollary. *Let $z = f(x,y)$ be a local representation of a smooth surface Σ in the tangent-normal coordinates at p. Suppose that its the Hessian matrix at p is $\left(\begin{smallmatrix} \ell & m \\ m & n \end{smallmatrix}\right)$. Then*

$$\operatorname{Shape}\mathrm{i} = \ell\cdot\mathrm{i} + m\cdot\mathrm{j}, \qquad \operatorname{Shape}\mathrm{j} = m\cdot\mathrm{i} + n\cdot\mathrm{j};$$

that is, the multiplication by the Hessian matrix at p describes its shape operator.

In particular, the principal curvatures of Σ at p are the eigenvalues of Shape_p *and the principal directions are the eigendirections of* Shape_p.

Since the Hessian matrix is symmetric, the corollary also implies that Shape is self-adjoint which gives another way to solve 9.4.

Proof. Note that $s\colon (u,v) \mapsto (u,v,f(u,v))$ is a chart of Σ that covers p. Further note that

$$\begin{aligned} s_u(0,0) &= \mathrm{i}, & s_v(0,0) &= \mathrm{j}, & \nu(0,0) &= \mathrm{k}, \\ s_{uu}(0,0) &= \ell\cdot\mathrm{k}, & s_{uv}(0,0) &= m\cdot\mathrm{k}, & s_{vv}(0,0) &= n\cdot\mathrm{k}. \end{aligned}$$

It remains to apply 9.3. □

9.6. Corollary. *Let Σ be a smooth surface with orientation defined by a unit normal field ν. Suppose the spherical map $\nu\colon \Sigma \to \mathbb{S}^2$ is injective. Then*

$$\iint\limits_{\Sigma} |K| = \operatorname{area}[\nu(\Sigma)].$$

Proof. Observe that the tangent planes $\mathrm{T}_p\Sigma = \mathrm{T}_{\nu(p)}\mathbb{S}^2$ are parallel for any $p \in \Sigma$. Indeed both of these planes are perpendicular to $\nu(p)$.

Choose an orthonormal basis of T_p consisting of principal directions, so the shape operator can be expressed by the matrix $\left(\begin{smallmatrix} k_1 & 0 \\ 0 & k_2 \end{smallmatrix}\right)$.

Since $\operatorname{Shape}_p = -d_p\nu$, 9.5 implies that

$$\operatorname{jac}_p \nu = \left|\det\left(\begin{smallmatrix} k_1 & 0 \\ 0 & k_2 \end{smallmatrix}\right)\right| = |K(p)|.$$

By the area formula (8.7), the statement follows. □

9.7. Exercise. *Let Σ be a smooth surface with orientation defined by a unit normal field ν. Suppose that Σ has unit principal curvatures at any point.*

(a) *Show that* $\mathrm{Shape}_p(\mathrm{w}) = \mathrm{w}$ *for any* $p \in \Sigma$ *and* $\mathrm{w} \in \mathrm{T}_p\Sigma$.
(b) *Show that* $p + \nu(p)$ *is constant; that is, the point* $c = p + \nu(p)$ *does not depend on* $p \in \Sigma$. *Conclude that* Σ *is a subset of the unit sphere centered at* c.

We define the *angle* between two oriented surfaces at a point of their intersection p as the angle between their normal vectors at p. The following exercise is a result by Ferdinand Joachimsthal [44] generalized by Ossian Bonnet [12].

9.8. Exercise. *Assume that two smooth oriented surfaces* Σ_1 *and* Σ_2 *intersect at constant angle along a smooth regular curve* γ. *Show that if* γ *is a curvature line in* Σ_1, *then it is also a curvature line in* Σ_2.

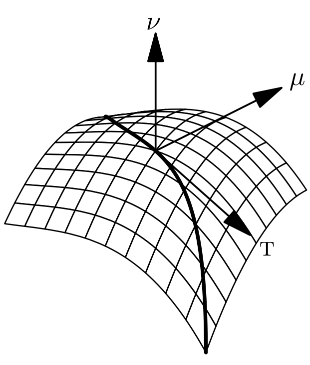
Conclude that if a smooth surface Σ *intersects a plane or sphere at constant angle along a smooth regular curve* γ, *then* γ *is a curvature line of* Σ.

9.9. Exercise. *Let* Σ *be a closed smooth surface with orientation defined by a unit normal field* ν. *Assume* t *is sufficiently close to zero.*
(a) *Show that the set of points of the form* $p + t{\cdot}\nu(p)$ *with* $p \in \Sigma$ *forms a smooth surface; denote it by* Σ_t.
(b) *Let* $a(t)$ *be the area of* Σ_t. *Show that*

$$a'(0) = -\iint\limits_{\Sigma} H,$$

where H *denotes mean curvature of* Σ.

E Curves in a surface



Suppose γ is a regular smooth curve in a smooth oriented surface Σ. As usual we denote by ν the unit normal field on Σ.

Without loss of generality we may assume that γ is unit-speed; in this case $\mathrm{T}(s) = \gamma'(s)$ is its tangent indicatrix. Let us use the shortcut notation $\nu(s) = \nu(\gamma(s))$. Note that the unit vectors $\mathrm{T}(s)$ and $\nu(s)$ are orthogonal. Therefore there is a unique unit vector $\mu(s)$ such that $\mathrm{T}(s), \mu(s), \nu(s)$ is an oriented orthonormal basis; it is called the *Darboux frame* of γ in Σ. Since $\mathrm{T}_{\gamma(s)} \perp \nu(s)$, the vector $\mu(s)$ is tangent to Σ at $\gamma(s)$. In fact $\mu(s)$ is a counterclockwise rotation of $\mathrm{T}(s)$ by the angle $\frac{\pi}{2}$ in the tangent plane $\mathrm{T}_{\gamma(s)}$. This vector can also be defined as the vector product $\mu(s) = \nu(s) \times \mathrm{T}(s)$.

Since γ is unit-speed, we have that $\gamma'' \perp \gamma'$ (see 3.1). Therefore the acceleration of γ can be written as a linear combination of μ and ν; that is,

$$\gamma''(s) = k_g(s)\cdot\mu(s) + k_n(s)\cdot\nu(s).$$

The values $k_g(s)$ and $k_n(s)$ are called *geodesic* and *normal curvatures* of γ at s, respectively. Since the frame $\mathrm{T}(s), \mu(s), \nu(s)$ is orthonormal, these curvatures can be also written as the following scalar products:

$$\begin{aligned} k_g(s) &= \langle\gamma''(s), \mu(s)\rangle = \\ &= \langle\mathrm{T}'(s), \mu(s)\rangle. \\ k_n(s) &= \langle\gamma''(s), \nu(s)\rangle = \\ &= \langle\mathrm{T}'(s), \nu(s)\rangle. \end{aligned}$$

Since $0 = \langle\mathrm{T}(s), \nu(s)\rangle$ we have that

$$\begin{aligned} 0 &= \langle\mathrm{T}(s), \nu(s)\rangle = \\ &= \langle\mathrm{T}'(s), \nu(s)\rangle + \langle\mathrm{T}(s), \nu'(s)\rangle = \\ &= k_n(s) + \langle\mathrm{T}(s), D_{\mathrm{T}(s)}\nu\rangle. \end{aligned}$$

Applying the definition of shape operator, we get the following:

9.10. Proposition. *Assume γ is a smooth unit-speed curve in a smooth surface Σ. Let $p = \gamma(s_0)$ and $\mathrm{V} = \gamma'(s_0)$. Then*

$$k_n(s_0) = \langle\mathrm{Shape}_p(\mathrm{V}), \mathrm{V}\rangle,$$

where k_n denotes the normal curvature of γ at s_0 and Shape_p is the shape operator at p.

Note that according to the proposition, the normal curvature of a regular smooth curve in Σ is completely determined by the velocity vector V at the point p. By that reason the normal curvature is also denoted by k_{V}; according to the proposition,

$$k_{\mathrm{V}} = \langle\mathrm{Shape}_p(\mathrm{V}), \mathrm{V}\rangle$$

for any unit vector V in T_p.

Let p be a point on a smooth surface Σ. Assume we choose tangent-normal coordinates at p so that the Hessian matrix is diagonalized, then we have

$$M_p = \begin{pmatrix} k_1(p) & 0 \\ 0 & k_2(p) \end{pmatrix}.$$

Consider a vector $\mathrm{W} = a\cdot\mathrm{I} + b\cdot\mathrm{J}$ in the (x,y)-plane. Then by 9.5, we have

$$\langle\mathrm{Shape}\,\mathrm{W}, \mathrm{W}\rangle = a^2\cdot k_1(p) + b^2\cdot k_2(p).$$

If w is unit, then $a^2+b^2=1$ which implies the following:

9.11. Observation. *For any point p on an oriented smooth surface Σ, the principal curvatures $k_1(p)$ and $k_2(p)$ are respectively the minimum and maximum of the normal curvatures at p. Moreover, if θ is the angle between a unit vector $\mathrm{w}\in\mathrm{T}_p$ and the first principal direction at p, then*

$$k_{\mathrm{w}}(p)=k_1(p)\cdot(\cos\theta)^2+k_2(p)\cdot(\sin\theta)^2.$$

The last identity is called *Euler's formula.*

9.12. Exercise. *Let Σ be a smooth surface. Show that the sum of the normal curvatures for any pair of orthogonal directions, at a point $p\in\Sigma$ equals $H(p)$ — the mean curvature at p.*

9.13. Meusnier's theorem. *Let γ be a regular smooth curve that runs along a smooth oriented surface Σ. Let $p=\gamma(t_0)$, $\mathrm{v}=\gamma'(t_0)$, and $\alpha=\measuredangle(\nu(p),\mathrm{N}(t_0))$; that is, α is the angle between the unit normal to Σ at p and the unit normal vector in the Frenet frame of γ at t_0. Then the following identity holds true*

$$\kappa(t_0)\cdot\cos\alpha=k_n(t_0);$$

here $\kappa(t_0)$ and $k_n(t_0)$ are the curvature and the normal curvature of γ at t_0, respectively.

Proof. Since $\gamma''=\mathrm{T}'=\kappa\cdot\mathrm{N}$, we get that

$$\begin{aligned}k_n(t_0)&=\langle\gamma'',\nu\rangle=\\&=\kappa(t_0)\cdot\langle\mathrm{N},\nu\rangle=\\&=\kappa(t_0)\cdot\cos\alpha.\end{aligned}$$

□

The theorem above, as well as the statement in the following exercise were proved by Jean Baptiste Meusnier [59].

9.14. Exercise. *Let Σ be a smooth surface, $p\in\Sigma$ and $\mathrm{v}\in\mathrm{T}_p\Sigma$ a unit vector. Assume that $k_{\mathrm{v}}(p)\neq0$; that is, the normal curvature of Σ at p in the direction of v does not vanish.*

Show that the osculating circles at p of smooth regular curves in Σ that run in the direction v sweep out a sphere S with center $p+\frac{1}{k_{\mathrm{v}}}\cdot\nu$ and radius $r=\frac{1}{|k_{\mathrm{v}}|}$.

9.15. Exercise. *Let $\gamma(s)=(x(s),y(s))$ be a smooth unit-speed simple plane curve in the upper half-plane, and Σ be the surface of revolution around the x-axis with generatrix γ.*

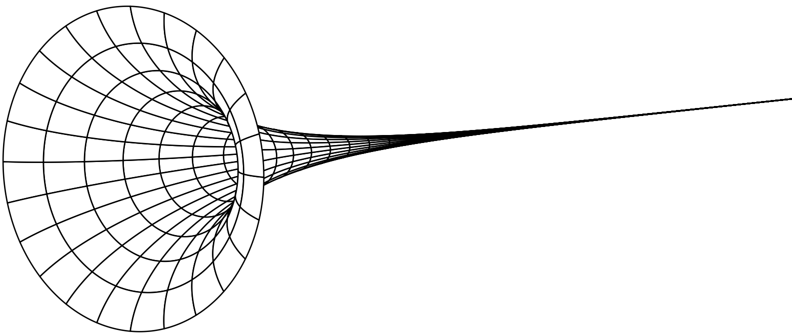
(a) *Show that the parallels and meridians are lines of curvature on* Σ.

(b) *Show that*

$$\frac{|x'(s)|}{y(s)} \quad \textit{and} \quad \frac{-y''(s)}{|x'(s)|}$$

are the principal curvatures of Σ *at* $(x(s), y(s), 0)$ *in the direction of the corresponding parallel and meridian respectively.*

(c) *Show that* Σ *has Gauss curvature* -1 *at all points if and only if* y *satisfies the differential equation* $y'' = y$. *The case* $y = e^{-s}$ *is shown; this is the so-called* pseudosphere.



9.16. Exercise. *Show that the* catenoid *defined implicitly by the equation*

$$(\operatorname{ch} z)^2 = x^2 + y^2$$

is a minimal surface.

9.17. Exercise. *Show that the* helicoid *defined by the following parametric equation*

$$s(u, v) = (u \cdot \sin v, u \cdot \cos v, v)$$

is a minimal surface.

F Lagunov's example

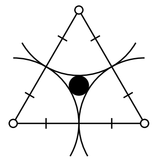
9.18. Exercise. *Assume* V *is a body in* $\mathbb{R}^3$ *bounded by a smooth surface of revolution with principal curvatures at most 1 in absolute value. Show that* V *contains a unit ball.*

The following question is a 3-dimensional analog of the moon in a puddle problem (6.14).

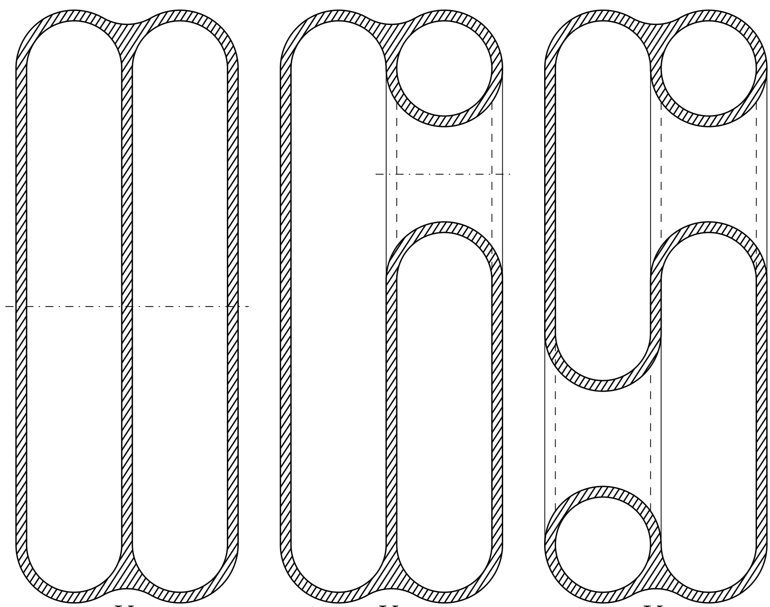
9.19. Question. *Assume a set* $V \subset \mathbb{R}^3$ *is bounded by a closed connected surface* Σ *with principal curvatures bounded in absolute value by 1. Is it true that* V *contains a ball of radius 1?*

According to 9.18, the answer is "yes" for surfaces of revolution. Later (see 10.7) we will show that the answer is "yes" for convex surfaces. Now we are going to show by an example that the answer is "no" in the general case; this example was constructed by Vladimir Lagunov [49].

Construction. Let us start with a body of revolution V_1 whose cross section is shown on the diagram below. The boundary curve of the cross section consists of 6 long vertical line segments included into 3 closed simple smooth curves. (To make the curves smooth, one has to use cutoffs and mollifiers from Section 0E.) The boundary of V_1 has 3 components, each of which is a smooth sphere.



We assume that the curves have curvature at most 1. Moreover with the exception of the almost vertical parts, the curve has absolute curvature close to 1 all the time. The only thick part in V is the place where all three boundary components come close together; the remaining part of V is assumed to be very thin. It could be arranged that the radius r of the maximal ball in V is just a little bit above $r_2 = \frac{2}{\sqrt{3}} - 1$. (The small black disc on the diagram has radius r_2, assuming that the three big circles are unit.) In particular, we may assume that $r < \frac{1}{6}$.



Exercise 9.15 gives formulas for the principal curvatures of the boundary of V; which imply that both principal curvatures are at most 1 in

absolute value.

It remains to modify V_1 to make its boundary connected without increasing the bounds on its principal curvatures and without allowing larger balls inside.

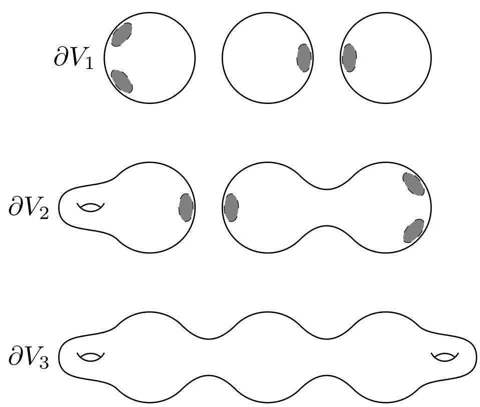
Note that each sphere in the boundary contains two flat discs; they come into pairs closely lying to each other. Let us drill thru two of such pairs and reconnect the holes by another body of revolution whose axis is shifted but stays parallel to the axis of V_1. Denote the obtained body by V_2; the cross section of the obtained body is shown on the diagram.

Then repeat the operation for the other two pairs. Denote the obtained body by V_3.

Note that the boundary of V_3 is connected. Assuming that the holes are large, its boundary can be made so that its principal curvatures are still at most 1; the latter can be proved in the same way as for V_1. $\square$

Remarks

Note that the surface of V_3 in the Lagunov's example has genus 2; that is, it can be parametrized by a sphere with two handles.



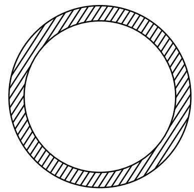
Indeed, the boundary of V_1 consists of three smooth spheres.

When we drill a hole, we make one hole in two spheres and two holes in one sphere. We reconnect two spheres with a tube and obtain one sphere. By connecting the two holes of the other sphere with a tube we get a torus; it is on the right side in the picture of V_2. That is, the boundary of V_2 is formed by one sphere and one torus.

To construct V_3 from V_2, we make a torus from the remaining sphere and connect it to the other torus by a tube. This way we get a sphere with two handles; that is, it has genus 2.

9.20. Exercise. *Modify Lagunov's construction to make the boundary surface a sphere with 4 handles.*

Question 9.19 can be asked differently: what is the maximal radius r of the ball that has to be included in any body bounded by a smooth surface with principal curvatures bounded in absolute value by 1.



One may also consider bodies bounded by more than one smooth surface. In this case the example of a region between two large concentric spheres with almost equal radii shows that in the general case there is no bound. Indeed, this region can be made arbitrarily thin while the curvature of the boundary can be made arbitrarily close to zero.

Recall that the Lagunov example shows that $r \leqslant r_2$, where r_2 is the radius of the smallest circle tangent to three unit circles that are tangent to each other, so

$$r_2 = \tfrac{2}{\sqrt{3}} - 1 < \tfrac{1}{6}.$$

The statement in the following exercise is due to Vladimir Lagunov [48]; it implies that this bound is optimal.

9.21. Advanced exercise. *Suppose a connected body $V \subset \mathbb{R}^3$ is bounded by a finite number of closed smooth surfaces with principal curvatures bounded in absolute value by 1. Assume that V does not contain a ball of radius r_2. Show that its boundary has two diffeomorphic connected components.*

Let r_3 be the radius of the smallest sphere tangent to four unit spheres that are tangent to each other. Direct calculations show that

$$r_3 = \sqrt{\tfrac{3}{2}} - 1 > \tfrac{1}{5}.$$

In particular $r_3 > r_2$.

The statement in the following exercise is a special case of a theorem by Vladimir Lagunov and Abram Fet [50, 51].

9.22. Very advanced exercise. *Suppose a body $V \subset \mathbb{R}^3$ is bounded by a smooth sphere with principal curvatures bounded in absolute value by 1. Show that V contains a ball of radius r_3.*

Show that this bound is sharp; that is, show that for every $\varepsilon > 0$, there are examples of V as above not containing a ball of radius $r_3 + \varepsilon$.

Chapter 10

Supporting surfaces

A Definitions

Assume two surfaces Σ_1 and Σ_2 have a common point p. If there is a neighborhood U of p such that $\Sigma_1 \cap U$ lies on one side from Σ_2 in U, then we say that Σ_2 *locally supports* Σ_1 at p.

Let us describe Σ_2 locally at p as a graph $z = f_2(x, y)$ in tangent-normal coordinates at p. If Σ_2 locally supports Σ_1 at p, then all points of Σ_1 near p lie either above or below the graph $z = f_2(x, y)$.

In both cases the surfaces Σ_1 and Σ_2 have common tangent plane at p, so we can write both as graphs $z = f_1(x, y)$ and $z = f_2(x, y)$ in the common tangent-normal coordinates at p. Note that Σ_2 locally supports Σ_1 at p if and only if

$$f_1(x, y) \geqslant f_2(x, y) \quad \text{or} \quad f_1(x, y) \leqslant f_2(x, y)$$

for all (x, y) sufficiently close to the origin.

If the surfaces are orientable, we can assume that they are *cooriented* at p; that is, they have a common unit normal vector at p in the direction of the z-axis. If the normal vectors are opposite, we say that Σ_1 and Σ_2 are *counteroriented* at p; in this case reverting the orientation of one of the surfaces makes them cooriented.

If Σ_2 locally supports Σ_1 and they are cooriented at p, then we can say that Σ_1 supports Σ_2 *from inside* or *from outside*, assuming that the normal vector points *inside* the domain bounded by the surface Σ_2 in U. Using the above notations, Σ_1 locally supports Σ_2 from inside (from outside) if $f_1(x, y) \geqslant f_2(x, y)$ (respectively $f_1(x, y) \leqslant f_2(x, y)$) for (x, y) in a sufficiently small neighborhood of the origin.

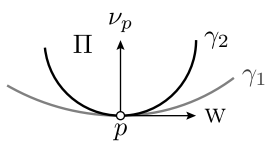
10.1. Proposition. *Let Σ_1 and Σ_2 be oriented surfaces. Assume Σ_1*

locally supports Σ_2 from inside at the point p (equivalently Σ_2 locally supports Σ_1 from outside). Then

$$k_1(p)_{\Sigma_1} \geqslant k_1(p)_{\Sigma_2} \quad \textit{and} \quad k_2(p)_{\Sigma_1} \geqslant k_2(p)_{\Sigma_2}.$$

10.2. Exercise. *Give an example of two surfaces Σ_1 and Σ_2 that have a common point p with a common unit normal vector ν_p such that $k_1(p)_{\Sigma_1} > > k_1(p)_{\Sigma_2}$ and $k_2(p)_{\Sigma_1} > k_2(p)_{\Sigma_2}$, but Σ_1 does not support Σ_2 locally at p.*

Proof. We can assume that Σ_1 and Σ_2 are graphs $z = f_1(x,y)$ and $z = f_2(x,y)$ in common tangent-normal coordinates at p, so we have $f_1 \geqslant f_2$.



Fix a unit vector $\mathrm{w} \in \mathrm{T}_p\Sigma_1 = \mathrm{T}_p\Sigma_2$. Consider the plane Π passing thru p and spanned by the normal vector ν_p and w. Let γ_1 and γ_2 be the curves of intersection of Σ_1 and Σ_2 with Π.

Let us orient Π so that the common normal vector ν_p for both surfaces at p points to the left from w. Further, let us parametrize both curves so that they are running in the direction of w at p and are therefore cooriented. Note that in this case the curve γ_1 supports the curve γ_2 from the right.

By 6.3, we have the following inequality for the normal curvatures of Σ_1 and Σ_2 at p in the direction of w:

$$❶ \qquad k_{\mathrm{w}}(p)_{\Sigma_1} \geqslant k_{\mathrm{w}}(p)_{\Sigma_2}.$$

According to 9.11,

$$k_1(p)_{\Sigma_i} = \min\{\, k_{\mathrm{w}}(p)_{\Sigma_i} \;:\; \mathrm{w} \in \mathrm{T}_p, |\mathrm{w}| = 1\,\}$$

for $i = 1, 2$. Choose w so that $k_1(p)_{\Sigma_1} = k_{\mathrm{w}}(p)_{\Sigma_1}$. Then by ❶, we have that

$$\begin{aligned} k_1(p)_{\Sigma_1} &= k_{\mathrm{w}}(p)_{\Sigma_1} \geqslant \\ &\geqslant k_{\mathrm{w}}(p)_{\Sigma_2} \geqslant \\ &\geqslant \min_{\mathrm{v}}\{\, k_{\mathrm{v}}(p)_{\Sigma_2}\,\} = \\ &= k_1(p)_{\Sigma_2}; \end{aligned}$$

here we assume that $\mathrm{v} \in \mathrm{T}_p$ and $|\mathrm{v}| = 1$. That is, $k_1(p)_{\Sigma_1} \geqslant k_1(p)_{\Sigma_2}$.

Similarly, by 9.11, we have that

$$k_2(p)_{\Sigma_i} = \max_{\mathrm{w}}\{\, k_{\mathrm{w}}(p)_{\Sigma_i}\,\}.$$

Let us fix w so that $k_2(p)_{\Sigma_2} = k_{\mathrm{w}}(p)_{\Sigma_2}$. Then

$$\begin{aligned} k_2(p)_{\Sigma_2} &= k_{\mathrm{w}}(p)_{\Sigma_2} \leqslant \\ &\leqslant k_{\mathrm{w}}(p)_{\Sigma_1} \leqslant \\ &\leqslant \max_{\mathrm{v}} \{ k_{\mathrm{v}}(p)_{\Sigma_1} \} = \\ &= k_2(p)_{\Sigma_1}; \end{aligned}$$

that is, $k_2(p)_{\Sigma_1} \geqslant k_2(p)_{\Sigma_2}$. □

10.3. Corollary. *Let Σ_1 and Σ_2 be oriented surfaces. Assume Σ_1 locally supports Σ_2 from inside at the point p. Then*
(a) $H(p)_{\Sigma_1} \geqslant H(p)_{\Sigma_2}$;
(b) If $k_1(p)_{\Sigma_2} \geqslant 0$, then $K(p)_{\Sigma_1} \geqslant K(p)_{\Sigma_2}$.

Proof; (a) The statement follows from 10.1 and the definition of mean curvature

$$H(p)_{\Sigma_i} = k_1(p)_{\Sigma_i} + k_2(p)_{\Sigma_i}.$$

(b). Since $k_2(p)_{\Sigma_i} \geqslant k_1(p)_{\Sigma_i}$ and $k_1(p)_{\Sigma_2} \geqslant 0$, we get that all the principal curvatures $k_1(p)_{\Sigma_1}$, $k_1(p)_{\Sigma_2}$, $k_2(p)_{\Sigma_1}$, and $k_2(p)_{\Sigma_2}$ are nonnegative. By 10.1, it implies that

$$\begin{aligned} K(p)_{\Sigma_1} &= k_1(p)_{\Sigma_1} \cdot k_2(p)_{\Sigma_1} \geqslant \\ &\geqslant k_1(p)_{\Sigma_2} \cdot k_2(p)_{\Sigma_2} = \\ &= K(p)_{\Sigma_2}. \end{aligned}$$

□

10.4. Exercise. *Show that any closed surface in a unit ball has a point with Gauss curvature at least 1.*

10.5. Exercise. *Show that any closed surface that lies at distance at most 1 from a straight line has a point with Gauss curvature at least 1.*

B Convex surfaces

A proper surface without boundary that bounds a convex region is called *convex.*

10.6. Exercise. *Show that the Gauss curvature of any convex smooth surface is nonnegative at each point.*

10.7. Exercise. *Assume R is a convex body in $\mathbb{R}^3$ bounded by a surface with principal curvatures at most 1. Show that R contains a unit ball.*

Recall that a region R in the Euclidean space is called *strictly convex* if for any two points $x, y \in R$, any point z between x and y lies in the interior of R.

Clearly any open convex set is strictly convex; the cube (as well as any convex polyhedron) gives an example of a non-strictly convex set. Recall that a closed convex region is strictly convex if and only if its boundary does not contain a line segment.

10.8. Lemma. *Let $z = f(x, y)$ be the local description of a smooth surface Σ in tangent-normal coordinates at some point $p \in \Sigma$. Assume both principal curvatures of Σ are positive at p. Then the function f is strictly convex in a neighborhood of the origin and has a local minimum at the origin.*

In particular the tangent plane T_p locally supports Σ from outside at p.

Proof. Since both principal curvatures are positive, by 9.5, we have

$$D^2_{\mathrm{w}} f(0,0) = \langle \mathrm{Shape}_p(\mathrm{w}), \mathrm{w} \rangle \geqslant k_1(p) > 0$$

for any unit tangent vector $\mathrm{w} \in \mathrm{T}_p\Sigma$ (which is the (x, y)-plane).

Since the set of unit vectors is compact, we have that

$$D^2_{\mathrm{w}} f(0,0) > \varepsilon$$

for some fixed $\varepsilon > 0$ and any unit tangent vector $\mathrm{w} \in \mathrm{T}_p\Sigma$.

By continuity of the function $(x, y, \mathrm{w}) \mapsto D^2_{\mathrm{w}} f(x, y)$, we have that $D^2_{\mathrm{w}} f(x, y) > 0$ if $\mathrm{w} \neq 0$ and (x, y) lies in a sufficiently small neighborhood of the origin. This property implies that f is a strictly convex function in a neighborhood of the origin in the (x, y)-plane (see Section 0E).

Finally since $\nabla f(0,0) = 0$ and f is strictly convex in a neighborhood of the origin it has a strict local minimum at the origin. $\square$

10.9. Exercise. *Let Σ be a smooth surface (without boundary) with positive Gauss curvature. Show that any connected component of intersection of Σ with a plane Π is either a single point or a smooth regular plane curve whose signed curvature has constant sign.*

The following theorem gives a global description of surfaces with positive Gauss curvature.

10.10. Theorem. *Suppose Σ is a proper smooth surface with positive Gauss curvature. Then Σ bounds a strictly convex region.*

In fact the statement holds for surfaces with possible self-intersections; it was stated and proved by James Stoker [73] who attributed it to Jacques Hadamard, who proved a closely relevant statement in [38, item 23].

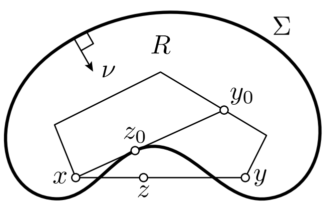
Note that in the proof we have to use that the surface is a connected set; otherwise a pair of disjoint spheres which bound two disjoint balls would give a counterexample.

Proof. Since the Gauss curvature is positive, we can choose unit normal field ν on Σ so that the principal curvatures are positive at any point. Denote by R the region bounded by Σ that lies on the side of ν; that is, ν points inside of R at any point of Σ. (The region R exists by 7.1.)

Let us show that R is *locally strictly convex*; that is, for any point $p \in R$, the intersection of R with a small ball centered at p is strictly convex.

Indeed, suppose that $z = f(x, y)$ is a local description of Σ in the tangent-normal coordinates at p. By 10.8, f is strictly convex in a neighborhood of the origin. In particular the intersection of a small ball centered at p with the epigraph $z \geqslant f(x, y)$ is strictly convex.

Since Σ is connected, so is R; moreover any two points in the interior of R can be connected by a polygonal line in the interior of R.



Assume the interior of R is not convex; that is, there are points $x, y \in R$ and a point z between x and y that does not lie in the interior of R. Consider a polygonal line β from x to y in the interior of R. Let y_0 be the first point on β such that the chord $[x, y_0]$ touches Σ at some point, say z_0.

Since R is locally strictly convex, $R \cap$ $\cap B(z_0, \varepsilon)$ is strictly convex for all sufficiently small $\varepsilon > 0$. On the other hand z_0 lies between two points in the intersection $[x, y_0] \cap B(z_0, \varepsilon)$. Since $[x, y_0] \subset R$, we arrived to a contradiction.

Therefore the interior of R is a convex set. Note that the region R is the closure of its interior, therefore R is convex as well.

Since R is locally strictly convex, its boundary Σ contains no line segments. Therefore R is strictly convex. □

Remark. We proved a more general statement. Namely, *any closed connected locally convex region in the Euclidean space is convex.*

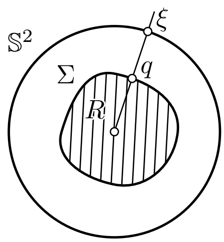
10.11. Exercise. *Assume that a closed convex surface Σ surrounds a unit circle. Show that there is a point $p \in \Sigma$ with $K(p) \leqslant 1$.*

10.12. Exercise. *Let Σ be a closed convex smooth surface of diameter at least π; that is, there is a pair of points $p, q \in \Sigma$ such that $|p - q| \geqslant \pi$. Show that Σ has a point with Gauss curvature at most 1.*

10.13. Theorem. *Suppose Σ is a closed smooth convex surface. Then it is a smooth sphere; that is, Σ admits a smooth regular parametrization by $\mathbb{S}^2$.*

The following exercise will guide you thru the proof of the theorem.

10.14. Exercise. *Assume a convex compact region R contains the origin in its interior and is bounded by a smooth surface Σ.*



(a) *Show that any half-line that starts at the origin intersects Σ at a single point; that is, there is a positive function $\rho\colon \mathbb{S}^2 \to \mathbb{R}$ such that Σ consists of the points $q = \rho(\xi)\cdot\xi$ for $\xi \in \mathbb{S}^2$.*

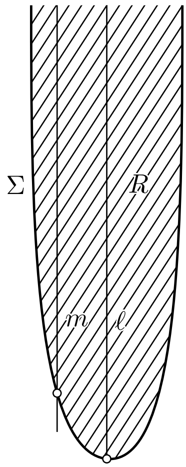
(b) *Show that $\rho\colon \mathbb{S}^2 \to \mathbb{R}$ is a smooth function.*

(c) *Conclude that $\xi \mapsto \rho(\xi)\cdot\xi$ is a smooth regular parametrization $\mathbb{S}^2 \to \Sigma$.*

10.15. Theorem. *Suppose Σ is an open smooth strictly convex surface. Then there is a coordinate system in which Σ is the graph $z = f(x, y)$ of a convex function f defined on a convex open region Ω of the (x, y)-plane.*

Moreover, $f(x_n, y_n) \to \infty$ if $(x_n, y_n) \in \Omega$ and $(x_n, y_n) \to (x_\infty, y_\infty) \in \in \partial\Omega$.

10.16. Exercise. *Assume a strictly convex closed noncompact region R contains the origin in its interior and is bounded by a smooth surface Σ.*



(a) *Show that R contains a half-line ℓ.*

(b) *Show that any line m parallel to ℓ intersects Σ at most at one point.*

(c) *Consider an (x, y, z)-coordinate system such that the z-axis points in the direction of ℓ. Show that the projection of Σ to the (x, y) plane is an open convex set; denote it by Ω.*

(d) *Conclude that Σ is a graph $z = f(x, y)$ of a convex function f defined on Ω. (It proves the main statement in 10.15.)*

(e) *Prove the last statement in 10.15.*

10.17. Exercise. *Show that any open surface Σ with positive Gauss curvature is a topological plane; that is, there is an embedding $\mathbb{R}^2 \to \mathbb{R}^3$ with image Σ.*

Try to show that Σ *is a smooth plane; that is, the embedding* f *can be made smooth and regular.*

10.18. Exercise. *Show that any open smooth surface* Σ *with positive Gauss curvature lies inside an infinite circular cone. In other words, there is an* (x, y, z)*-coordinate system in which* Σ *lies in the region* $z \geqslant m\cdot\sqrt{x^2 + y^2}$ *for some* $m > 0$.

10.19. Exercise. *Assume* Σ *is a smooth convex surface with positive Gauss curvature. Show that*

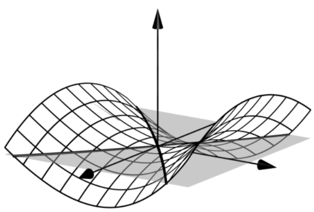
(a) *If* Σ *is closed, then the spherical map* $\nu\colon \Sigma \to \mathbb{S}^2$ *is a bijection. Conclude that*

$$\iint\limits_{\Sigma} K = 4\cdot\pi.$$

(b) *If* Σ *is open, then the spherical map* $\nu\colon \Sigma \to \mathbb{S}^2$ *maps* Σ *bijectively into a subset of a hemisphere. Conclude that*

$$\iint\limits_{\Sigma} K \leqslant 2\cdot\pi.$$

C Saddle surfaces



A surface is called *saddle* if its Gauss curvature at each point is nonpositive; in other words the principal curvatures at each point have opposite signs or at least one of them is zero.

If the Gauss curvature is negative at each point, then the surface is said to be *strictly saddle*; equivalently, this means that the principal curvatures have opposite signs at each point. Note that in this case the tangent plane cannot support the surface even locally — moving along the surface in the principal directions at a given point, one goes above and below the tangent plane at this point.

10.20. Exercise. *Let* $f\colon \mathbb{R} \to \mathbb{R}$ *be a smooth positive function. Show that the surface of revolution of the graph* $y = f(x)$ *around the* x*-axis is saddle if and only if* f *is convex; that is, if* $f''(x) \geqslant 0$ *for any* x.

A surface Σ is called *ruled* if for every point $p \in \Sigma$ there is a line segment $\ell_p \subset \Sigma$ passing thru p that is infinite or its endpoint(s) lie on the boundary line of Σ.

10.21. Exercise. *Show that any ruled surface Σ is saddle.*

10.22. Exercise. *Let Σ be an open strictly saddle surface and $f\colon\mathbb{R}^3\to$ $\to\mathbb{R}$ be a smooth convex function. Show that the restriction of f to Σ does not have a point of strict local maximum.*

A tangent direction on a smooth surface with vanishing normal curvature is called *asymptotic*. A smooth regular curve that always runs in an asymptotic direction is called an *asymptotic line*.

Recall that a set R in the plane is called *star-shaped* if there is a point $p\in R$ such that for any $x\in R$ the line segment $[p,x]$ belongs to R.

The statement in the following exercise is due to Dmitri Panov [66].

10.23. Advanced exercise. *Let γ be a closed smooth asymptotic line in such graph $z=f(x,y)$ of a smooth function f. Assume that the graph is strictly saddle in a neighborhood of γ. Show that the region in the (x,y)-plane bounded by the projection of γ to such plane cannot be star-shaped.*

10.24. Advanced exercise. *Let Σ be a smooth surface and $p\in\Sigma$. Assume $K(p)<0$. Show that there is a neighborhood Ω of p in Σ such that the intersection of Ω with the tangent plane T_p is a union of two smooth curves* intersecting transversally *at p.*

D Hats

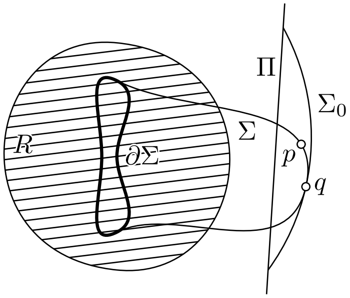
Note that a closed surface cannot be saddle. Indeed, let Σ be a closed surface and consider a smallest sphere that contains Σ in its interior; it supports Σ at some point p and at this point the principal curvatures must have the same sign. The following more general statement is proved using the same idea.

10.25. Lemma. *Assume Σ is a compact saddle surface and its boundary line lies in a closed convex region R. Then the entire surface Σ lies in R.*

Remark. Note that in the case when Σ is strictly saddle, the follows from 10.22.

Proof. Arguing by contradiction, assume there is a point $p\in\Sigma$ that does not lie in R. Let Π be a plane that separates p from R; it exists by 0.14. Denote by Σ' the part of Σ that lies with p on the same side of Π.

Since Σ is compact, it is surrounded by a sphere; let σ be the circle of intersection of this sphere and Π. Consider the smallest spherical dome Σ_0 with boundary σ that surrounds Σ'.



Note that Σ_0 supports Σ at some point q. Without loss of generality we may assume that Σ_0 and Σ are cooriented at q and Σ_0 has positive principal curvatures. In this case Σ_0 supports Δ from outside. By 10.3, we have $K(q)_\Sigma \geqslant \geqslant K(q)_{\Sigma_0} > 0$, a contradiction. $\square$

10.26. Exercise. *Let Δ be a compact smooth regular saddle surface with boundary and $p \in \Delta$. Assume the boundary line of Δ lies in the unit sphere centered at p. Show that if Δ is a disc, then* $\mathrm{length}(\partial\Delta) \geqslant 2\cdot\pi$.

Remark. Show that the statement does not hold without assuming that Δ is a disc.

If Δ is as in the exercise, then in fact $\mathrm{area}\,\Delta \geqslant \pi$. The proof of this statement can be obtained by applying the so-called *coarea fromula* together with the inequality in the exercise.

10.27. Exercise. *Show that an open saddle surface cannot lie inside of an infinite circular cone.*

A disc Δ in a surface Σ is called a *hat* of Σ if its boundary line $\partial\Delta$ lies in a plane Π and $\Delta\backslash\partial\Delta$ lies on one side of Π.

10.28. Proposition. *A smooth surface Σ is saddle if and only if it has no hats.*

Note that a saddle surface can contain a closed plane curve. For example the hyperboloid $x^2 + y^2 - z^2 = 1$ contains the unit circle in the (x, y)-plane. However, according to the proposition (as well as the lemma), a plane curve cannot bound a disc (as well any compact set) in a saddle surface.

Proof. Since a plane is convex, the "only if" part follows from 10.25; it remains to prove the "if" part.

Assume Σ is not saddle; that is, it has a point p with strictly positive Gauss curvature; or equivalently, the principal curvatures $k_1(p)$ and $k_2(p)$ have the same sign.

Let $z = f(x, y)$ be a graph representation of Σ in tangent-normal coordinates at p. Consider the set F_ε in the (x, y)-plane defined by the inequality $f(x, y) \leqslant \varepsilon$. By 10.8, f is convex in a small neighborhood of $(0, 0)$. Therefore F_ε is convex, for sufficiently small $\varepsilon > 0$. In particular, F_ε is a topological disc.

Note that $(x,y) \mapsto (x,y,f(x,y))$ is a homeomorphism from F_ε to

$$\Delta_\varepsilon = \left\{ (x,y,f(x,y)) \in \mathbb{R}^3 \ : \ f(x,y) \leqslant \varepsilon \right\};$$

so Δ_ε is a topological disc for any sufficiently small $\varepsilon > 0$. Note that the boundary line of Δ_ε lies on the plane $z = \varepsilon$ and the entire disc lies below it; that is, Δ_ε is a hat of Σ. □

The following exercise shows that Δ_ε is in fact a smooth disc. This can be used to prove a slightly stronger version of 10.28; namely in the definition of hats one can assume that the disc is smooth.

10.29. Exercise. *Let $f\colon \mathbb{R}^2 \to \mathbb{R}$ be a smooth strictly convex function with minimum at the origin. Show that the set F_ε in the graph $z = f(x,y)$ defined by the inequality $f(x,y) \leqslant \varepsilon$ is a smooth disc for any $\varepsilon > 0$; that is, there is a diffeomorphism F_ε to the unit disc $\Delta = \left\{ (x,y) \in \mathbb{R}^2 \ : \ x^2 + y^2 \leqslant 1 \right\}$.*

10.30. Exercise. *Let $L\colon \mathbb{R}^3 \to \mathbb{R}^3$ be an affine transformation; that is, L is an invertible map $\mathbb{R}^3 \to \mathbb{R}^3$ that sends any plane to a plane. Show that for any saddle surface Σ the image $L(\Sigma)$ is also a saddle surface.*

E Saddle graphs

The following theorem was proved by Sergei Bernstein [9].

10.31. Theorem. *Let $f\colon \mathbb{R}^2 \to \mathbb{R}$ be a smooth function. Assume its graph $z = f(x,y)$ is a strictly saddle surface in $\mathbb{R}^3$. Then f is not bounded; that is, there is no constant C such that $|f(x,y)| \leqslant C$ for any $(x,y) \in \mathbb{R}^2$.*

The theorem states that a saddle graph cannot lie between two parallel horizontal planes; applying 10.30 we get that saddle graphs cannot lie between two parallel planes, not necessarily horizontal. The following exercise shows that the theorem does not hold for saddle surfaces which are not graphs.

10.32. Exercise. *Construct an open strictly saddle surface that lies between two parallel planes.*

Since $\exp(x - y^2) > 0$, the following exercise shows that there are strictly saddle graphs with functions bounded on one side; that is, both (upper and lower) bounds are needed in the proof of Bernshtein's theorem.

10.33. Exercise. *Show that the graph $z = \exp(x - y^2)$ is strictly saddle.*

The following exercise gives a condition that guarantees that a saddle surface is a graph; it can be used in combination with Bernshtein's theorem.

10.34. Advanced exercise. *Let Σ be an open smooth strictly saddle disc in $\mathbb{R}^3$. Assume there is a compact subset $K \subset \Sigma$ such that the complement $\Sigma\backslash K$ is the graph $z = f(x, y)$ of a smooth function defined in an open domain of the (x, y)-plane. Show that the surface Σ is a graph.*

Note that according to 10.25, there are no proper saddle surfaces inside a parallelepiped with boundary line on one of its facets. The following lemma gives an analogous statement for a parallelepiped with an infinite side.

10.35. Lemma. *There is no proper strictly saddle smooth surface that has its boundary line in a plane Π and lies at a bounded distance from a line contained in Π.*

Proof. Note that by 10.30, the statement can be reformulated in the following way: *There is no proper strictly saddle smooth surface with boundary line in the (x, y)-plane and contained in a region of the form:*

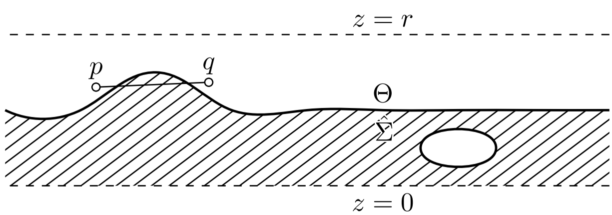
$$R = \left\{ (x, y, z) \in \mathbb{R}^3 \, : \, 0 \leqslant z \leqslant r, 0 \leqslant y \leqslant r \right\}.$$

Let us prove this statement.

Assume the contrary, let Σ be such a surface. Consider the projection $\hat{\Sigma}$ of Σ to the (x, z)-plane. It lies in the upper half-plane and below the line $z = r$.

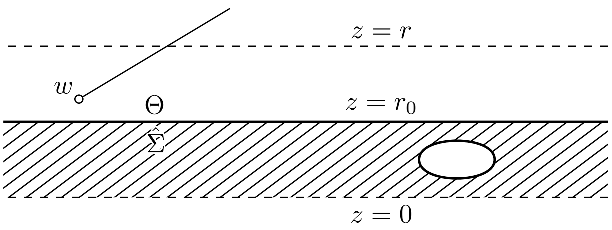
Consider the open upper half-plane $H = \left\{ (x, z) \in \mathbb{R}^2 \, : \, z > 0 \right\}$. Let Θ be the connected component of the complement $H\backslash\hat{\Sigma}$ that contains all the points above the line $z = r$.

Note that Θ is convex. If not, then there is a line segment $[p, q]$ for some $p, q \in \Theta$ that cuts from $\hat{\Sigma}$ a compact piece. Consider the plane



Π thru $[p, q]$ that is perpendicular to the (x, z)-plane. Note that Π cuts from Σ a compact region Δ. By a general position argument (see 8.11) we can assume that Δ is a compact surface with boundary line in Π and the remaining part of Δ lies on one side from Π. Since the plane Π is convex, this statement contradicts 10.25.

Summarizing, Θ is an open convex set of H that contains all points above $z = r$. By convexity, together with any point w, the set Θ contains all points on the half-lines that start as w and *point up*; that is, in directions with positive z-coordinate. In other words, with any point w, the set Θ contains all points with larger z-coordinates. Since Θ is open it

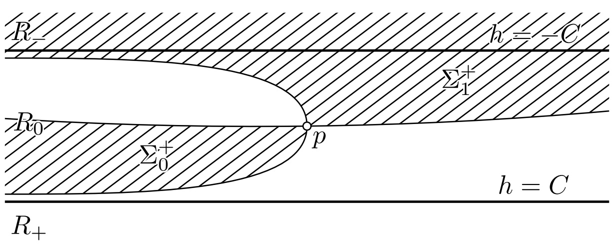


can be described by an inequality $z > r_0$. It follows that the plane $z = r_0$ supports Σ at some point (in fact at many points). By 10.1, the latter is impossible — a contradiction. □

Proof of 10.31. Denote by Σ the graph $z = f(x, y)$. Assume the contrary; that is, Σ lies between two planes $z = \pm C$.

Note that the function f cannot be constant. It follows that the tangent plane T_p at some point $p \in \Sigma$ is not horizontal.

Denote by Σ^+ the part of Σ that lies above T_p. Note that it has at least two connected components which are approaching p from both sides in the principal direction with positive principal curvature. Indeed, if there was a curve that runs in Σ^+ and approaches p from both sides, then it would cut a disc from Σ with boundary line above T_p and some points below it; the latter would contradict 10.25.



The surface Σ seeing from above.

Summarizing, Σ^+ has at least two connected components, denote them by Σ_0^+ and Σ_1^+. Let $z = h(x, y) = a{\cdot}x + b{\cdot}y + c$ be the equation of T_p. Note that Σ^+ contains all points in the region

$$R_- = \{\,(x, y, f(x, y)) \in \Sigma \,:\, h(x, y) < -C\,\}$$

which is a connected set and no points in

$$R_+ = \{\,(x, y, f(x, y)) \in \Sigma \,:\, h(x, y) > C\,\}$$

Whence one of the connected components, say Σ_0^+, lies in

$$R_0 = \{\, (x, y, f(x, y)) \in \Sigma \,:\, |h(x, y)| \leqslant C \,\}\,.$$

This set lies on a bounded distance from the line of intersection of T_p with the (x, y)-plane.

Let us move T_p slightly upward and cut from Σ_0^+ the piece above the obtained plane, say $\bar{\Sigma}_0^+$. By the general position argument (8.11), we can assume that $\bar{\Sigma}_0^+$ is a surface with smooth boundary line; by construction the boundary line lies in the plane. Note that the obtained surface $\bar{\Sigma}_0^+$ still lies on a bounded distance to a line. The latter is impossible by 10.35. □

Remarks

Note that Bernstein's theorem and the lemma in its proof do not hold for nonstrictly saddle surfaces; counterexamples can be found among infinite cylinders over smooth regular curves. In fact it can be shown that these are the only counterexamples; a proof is based on the same idea, but it is much more technical.

By 10.28, saddle surfaces can be defined as smooth surfaces without hats. This definition can be used for arbitrary surfaces, not necessarily smooth. Some results, for example Bernshtein's characterization of saddle graphs can be extended to generalized saddle surfaces, but this class of surfaces is far from being understood; see [5, Chapter 4] and the references therein.

Chapter 11

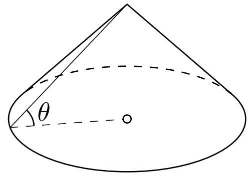
Shortest paths

A Intrinsic geometry

We begin to study the *intrinsic geometry* of surfaces. A property is called *intrinsic* if it can be checked by measuring things inside the surface, for example lengths of curves or angles between the curves that lie in the surface. Otherwise, if a definition of a property essentially uses the ambient space, then it is called *extrinsic*.

For instance, the mean curvature as well as the Gauss curvature are defined via principal curvatures, which are extrinsic. Later (15.14) it will be shown that *remarkably* the Gauss curvature is actually intrinsic — it can be calculated via measurements inside the surface. The mean curvature is not intrinsic, for example, the intrinsic geometry of the (x,y)-plane is not distinguishable from the intrinsic geometry of the graph $z = (x+y)^2$. However, while the mean curvature of the former vanishes at all points, the mean curvature of the latter does not vanish, say at $p=(0,0,1)$.

The following exercise should help you get in the right mood; it might look like a tedious problem in calculus, but actually it is an easy problem in geometry. We learned this problem from Joel Fine, who attributed it so Frederic Bourgeois [30].



11.1. Exercise. *A cowboy stands at the bottom of a frictionless ice-mountain formed by a cone with a circular base with angle of inclination θ. He wants to climb the mountain; he throws up his lasso which slips neatly over the tip of the cone, pulls it tight and starts to climb.*

What is the critical angle θ at which the cowboy can no longer climb the ice-mountain?

B Definition

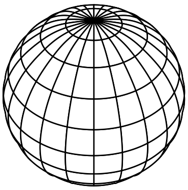
Let p and q be two points on a surface Σ. Recall that $|p-q|_\Sigma$ denotes the induced length distance from p to q; that is, the greatest lower bound of the lengths of paths in Σ from p to q.

Note that if Σ is smooth, then any two points in Σ can be joined by a piecewise smooth path. Since any such path is rectifiable, the value $|p-q|_\Sigma$ is finite for any pair $p, q\in\Sigma$.

A path γ from p to q in Σ that minimizes the length is called a *shortest path* from p to q.

The image of a shortest path between p and q in Σ is usually denoted by $[p,q]$ or by $[p,q]_\Sigma$. In general there might be no shortest path between two given points on a surface and there might be many of them; this is shown in the following two examples.

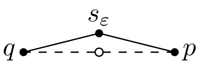
Usually, if we write $[p,q]_\Sigma$, we are assuming that such a shortest path exists and we have made a choice of one of them.



Nonuniqueness. There are plenty of shortest paths between the poles of a round sphere — each meridian is a shortest path.

Nonexistence. Let Σ be the (x,y)-plane with the origin removed. Consider two points $p=(1,0,0)$ and $q=(-1,0,0)$ in Σ.

11.2. Claim. *There is no shortest path from p to q in Σ.*



Proof. Note that $|p-q|_\Sigma=2$. Indeed, given $\varepsilon>0$, consider the point $s_\varepsilon=(0,\varepsilon,0)$. Observe that the polygonal path $ps_\varepsilon q$ lies in Σ and its length $2\cdot\sqrt{1+\varepsilon^2}$ approaches 2 as $\varepsilon\to0$. It follows that $|p-q|_\Sigma\leqslant2$. Since $|p-q|_\Sigma\geqslant|p-q|_{\mathbb{R}^3}=2$, we get $|p-q|_\Sigma=2$.

It follows that a shortest path from p to q, if it exists, must have length 2. By the triangle inequality any curve of length 2 from p to q must run along the line segment $[p,q]$; in particular it must pass thru the origin. Since the origin does not lie in Σ, there is no shortest from p to q in Σ $\square$

11.3. Proposition. *Any two points in a proper smooth surface can be joined by a shortest path.*

Proof. Fix a proper smooth surface Σ and two points p and q. Set $\ell=|p-q|_\Sigma$.

By the definition of induced length-metric (Section 2G), there is a sequence of paths γ_n from p to q in Σ such that

$$\operatorname{length}\gamma_n\to\ell\quad\text{as}\quad n\to\infty.$$

Without loss of generality, we may assume that $\operatorname{length}\gamma_n < \ell + 1$ for any n and each γ_n is parametrized proportional to its arc-length. In particular each path $\gamma_n\colon [0,1] \to \Sigma$ is $(\ell+1)$-Lipschitz; that is,

$$|\gamma(t_0) - \gamma(t_1)| \leqslant (\ell+1)\cdot|t_0 - t_1|$$

for any $t_0, t_1 \in [0,1]$.

Note that the image of γ_n lies in the closed ball $\bar{B}[p, \ell+1]$ for any n. It follows that the coordinate functions of γ_n are uniformly equicontinuous and uniformly bounded. By the Arzelá–Ascoli theorem (0.19) there is a convergent subsequence of γ_n and its limit, say $\gamma_\infty\colon [0,1] \to \mathbb{R}^3$, is continuous; that is, γ_∞ is a path. Evidently γ_∞ runs from p to q; in particular

$$\operatorname{length}\gamma_\infty \geqslant \ell.$$

Since Σ is a closed set, γ_∞ lies in Σ. Finally, since length is semicontinuous (2.16), we get that

$$\operatorname{length}\gamma_\infty \leqslant \ell.$$

Therefore $\operatorname{length}\gamma_\infty = \ell$ or, equivalently, γ_∞ is a shortest path from p to q. □

C Closest point projection

11.4. Lemma. *Let R be a closed convex set in $\mathbb{R}^3$. Then for every point $p \in \mathbb{R}^3$ there is a unique point $\bar{p} \in R$ that minimizes the distance to R; that is, $|p - \bar{p}| \leqslant |p - x|$ for any point $x \in R$.*

Moreover the map $p \mapsto \bar{p}$ is short; that is,

❶ $$|p - q| \geqslant |\bar{p} - \bar{q}|$$

for any pair of points $p, q \in \mathbb{R}^3$.

The map $p \mapsto \bar{p}$ is called the *closest point projection*; it maps the Euclidean space to R. Note that if $p \in R$, then $\bar{p} = p$.

Proof. Fix a point p and set

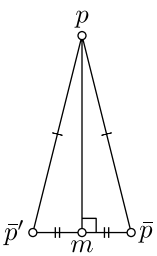
$$\ell = \inf\{\, |p - x| \;:\; x \in R \,\}.$$

Choose a sequence $x_n \in R$ such that $|p - x_n| \to \ell$ as $n \to \infty$.

Without loss of generality, we can assume that all the points x_n lie in a ball of radius $\ell + 1$ centered at p. Therefore we can pass to a *partial limit* $\bar{p}$ of x_n; that is, $\bar{p}$ is a limit of a subsequence of x_n. Since R is closed, $\bar{p} \in R$. By construction

$$|p - \bar{p}| = \lim_{n\to\infty} |p - x_n| = \ell.$$

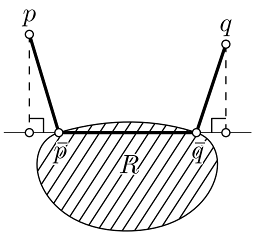
Hence the existence follows.



Assume there are two distinct points $\bar p, \bar p' \in R$ that minimize the distance to p. Since R is convex, their midpoint $m = \frac{1}{2}\cdot(\bar p + \bar p')$ lies in R. Note that $|p - \bar p| = |p - \bar p'| = \ell$; that is, the triangle $[p\bar p\bar p']$ is isosceles and therefore the triangle $[p\bar p m]$ is right with the right angle at m. Since a leg of a right triangle is shorter than its hypotenuse, we have $|p - m| < \ell$, a contradiction.

It remains to prove ❶. We can assume that $\bar p \neq \bar q$, otherwise there is nothing to prove.

Note that if $\measuredangle[\bar p\,{}^{p}_{\bar q}] < \frac{\pi}{2}$, then $|p-x| < |p-\bar p|$ for some point $x \in [\bar p, \bar q]$. Since $[\bar p, \bar q] \subset K$, the latter is impossible.



Therefore $p = \bar p$ or $\measuredangle[\bar p\,{}^{p}_{\bar q}] \geqslant \frac{\pi}{2}$. In both cases the orthogonal projection of p to the line $\bar p\bar q$ lies behind $\bar p$, or coincides with $\bar p$. The same way we show that the orthogonal projection of q to the line $\bar p\bar q$ lies behind $\bar q$, or coincides with $\bar q$. It implies that the orthogonal projection of the line segment $[p, q]$ to the line $\bar p\bar q$ contains the line segment $[\bar p, \bar q]$. In particular

$$|p - q| \geqslant |\bar p - \bar q|. \qquad \square$$

11.5. Corollary. *Assume a surface Σ bounds a closed convex region R and $p, q \in \Sigma$. Denote by W the outer closed region of Σ; in other words W is the union of Σ and the complement of R. Then*

$$\operatorname{length}\gamma \geqslant |p - q|_\Sigma$$

for any path γ in W from p to q. Moreover if γ does not lie in Σ, then the inequality is strict.

Proof. The first part of the corollary follows from the lemma and the definition of length. Indeed, consider the closest point projection $\bar\gamma$ of γ. Note that $\bar\gamma$ lies in Σ and connects p to q therefore

$$\operatorname{length}\bar\gamma \geqslant |p - q|_\Sigma.$$

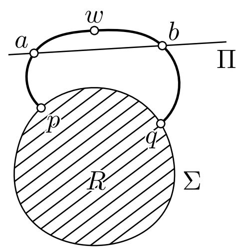
To prove the first statement, it remains to show that

❷ $$\operatorname{length}\gamma \geqslant \operatorname{length}\bar\gamma.$$

Consider a polygonal line $\bar p_0 \dots \bar p_n$ inscribed in $\bar\gamma$. Let $p_0 \dots p_n$ be the corresponding polygonal line inscribed in γ; that is $p_i = \gamma(t_i)$ if $\bar p_i = \bar\gamma(t_i)$. By 11.4 $|p_i - p_{i-1}| \geqslant |\bar p_i - \bar p_{i-1}|$ for any i. Therefore

$$\operatorname{length} p_0 \dots p_n \geqslant \operatorname{length} \bar p_0 \dots \bar p_n.$$

Taking the least upper bound of each side of the inequality for all inscribed polygonal lines $p_0 \dots p_n$ in γ, we get ❷.



It remains to prove the second statement. Suppose there is a point $w = \gamma(t_1) \notin \Sigma$; note that $w \notin R$. By the separation lemma (0.14) there is a plane Π that cuts w from Σ. The curve γ must intersect Π at two points: one point before t_1 and one after. Let $a = \gamma(t_0)$ and $b = \gamma(t_2)$ be these points. Note that the arc of γ from a to b is strictly longer that $|a-b|$; indeed its length is at least $|a-w|+|w-b|$ and $|a-w|+|w-b| > |a-b|$ since $w \notin [a,b]$.

Remove from γ the arc from a to b and replace it with the line segment $[a,b]$; denote the obtained curve by γ_1. From above, we have that

$$\operatorname{length}\gamma > \operatorname{length}\gamma_1$$

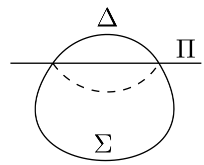
Note that γ_1 runs in W. Therefore by the first part of the corollary, we have

$$\operatorname{length}\gamma_1 \geqslant |p-q|_\Sigma.$$

Whence the second statement follows. □

11.6. Exercise. *Suppose Σ is a proper smooth surface with positive Gauss curvature and ν is the unit normal field on Σ. Show that for any two points $p, q \in \Sigma$ we have the following inequality:*

$$|p-q|_\Sigma \leqslant 2\cdot\frac{|p-q|}{|\nu(p)+\nu(q)|}.$$



11.7. Exercise. *Suppose Σ is a closed smooth surface that bounds a convex region R in $\mathbb{R}^3$ and Π is a plane that cuts a hat Δ from Σ. Assume that the reflection of the interior of Δ across Π lies in the interior of R. Show that Δ is* convex *with respect to the intrinsic metric of Σ; that is, if both ends of a shortest path in Σ lie in Δ, then the entire path lies in Δ.*

Let us define the *intrinsic diameter* of a closed surface Σ as the least upper bound of the lengths of shortest paths in the surface.

11.8. Exercise. *Assume a closed smooth surface Σ with positive Gauss curvature lies in a unit ball B.*

(a) Show that the intrinsic diameter of Σ cannot exceed π.

(b) Show that the area of Σ cannot exceed $4\cdot\pi$.

Chapter 12

Geodesics

A Definition

A smooth curve γ in a smooth surface Σ is called a *geodesic* if for any t, the acceleration $\gamma''(t)$ is perpendicular to the tangent plane $\mathrm{T}_{\gamma(t)}$.

Physically, geodesics can be understood as the trajectories of particles that slide on Σ without friction. Indeed, since there is no friction, the force that keeps the particle on Σ must be perpendicular to Σ. Therefore, by Newton's second law of motion, we get that the acceleration γ'' is perpendicular to $\mathrm{T}_{\gamma(t)}$.

From a physics point of view, the following lemma is a corollary of the law of conservation of energy.

12.1. Lemma. *Let γ be a geodesic in a smooth surface Σ. Then $|\gamma'|$ is constant.*

Moreover, for any $\lambda \in \mathbb{R}$, the curve $\gamma_\lambda := \gamma(\lambda{\cdot}t)$ is a geodesic as well. In other words, any geodesic has constant speed, and multiplying its parameter by a constant yields another geodesic.

Proof. Since $\gamma'(t)$ is a tangent vector at $\gamma(t)$, we have that $\gamma''(t) \perp \gamma'(t)$, or equivalently $\langle \gamma'', \gamma' \rangle = 0$ for any t. Whence

$$\langle \gamma', \gamma' \rangle' = 2{\cdot}\langle \gamma'', \gamma' \rangle = 0.$$

That is, $|\gamma'|^2 = \langle \gamma', \gamma' \rangle$ is constant.

The second part, follows since $\gamma_\lambda''(t) = \lambda^2{\cdot}\gamma''(\lambda t)$. $\square$

The statement in the following exercise is called *Clairaut's relation*; it can be obtained from the lemma above and the conservation of angular momentum.

12.2. Exercise. *Let γ be a geodesic on a smooth surface of revolution. Suppose that $r(t)$ denotes the distance from $\gamma(t)$ to the axis of rotation and $\theta(t)$ — the angle between $\gamma'(t)$ and the latitudinal circle thru $\gamma(t)$.*

Show that the value $r(t)\cdot\cos\theta(t)$ is constant.

Recall that an asymptotic line is a curve in the surface with vanishing normal curvature.

12.3. Exercise. *Assume a curve γ is a geodesic and, at the same time, is an asymptotic line of a smooth surface Σ. Show that γ is a straight line segment.*

12.4. Exercise. *Assume a smooth surface Σ is mirror symmetric with respect to a plane Π. Suppose that Σ and Π intersect along a smooth regular curve γ. Show that γ, parametrized by arc-length, is a geodesic on Σ.*

B Existence and uniqueness

The following proposition says that the position of a particle that travels without friction depends smoothly on its initial position, initial velocity, and the amount of time it has traveled.

12.5. Proposition. *Let Σ be a smooth surface without boundary. Given a tangent vector v to Σ at a point p, there is a unique geodesic $\gamma\colon\mathbb{I}\to\Sigma$ defined on a maximal open interval $\mathbb{I}\ni 0$ that starts at p with velocity vector v; that is, $\gamma(0)=p$ and $\gamma'(0)=\mathrm{v}$.*

Moreover

(a) the map $(p,\mathrm{v},t)\mapsto\gamma(t)$ is smooth in its domain of definition.

(b) if Σ is proper, then $\mathbb{I}=\mathbb{R}$; that is, the maximal interval is the entire real line.

A surface that satisfies the conclusion of *(b)* for any tangent vector v is said to be *geodesically complete*. So part *(b)* says that any proper surface is geodesically complete. The latter statement is a part of the *Hopf–Rinow theorem* [40].

The proof of this proposition uses the theorem on existence of solutions for an initial value problem (0.25). The following lemma shows that the condition of being a geodesic can be stated as a second order differential equation.

12.6. Lemma. *Let f be a smooth function defined on an open domain in $\mathbb{R}^2$. A smooth curve $t\mapsto\gamma(t)=(x(t),y(t),z(t))$ is a geodesic in the* *graph* $z = f(x, y)$ *if and only if* $z(t) = f(x(t), y(t))$ *for any* t *and the functions* $t \mapsto x(t)$ *and* $t \mapsto y(t)$ *satisfy a differential equation*

$$\begin{cases} x'' = g(x, y, x', y'), \\ y'' = h(x, y, x', y'), \end{cases}$$

where g *and* h *are smooth functions of four variables determined by* f*.*

The proof of the lemma is done by means of direct calculations.

Proof. In the following calculations, we often omit the arguments — we may write x instead of $x(t)$ and f instead of $f(x, y)$ or $f(x(t), y(t))$ and so on.

First let us calculate $z''(t)$ in terms of f, $x(t)$, and $y(t)$.

$$\begin{aligned} z'' &= f(x, y)'' = \\ &= (f_x \cdot x' + f_y \cdot y')' = \qquad ❶ \\ &= f_{xx} \cdot (x')^2 + f_x \cdot x'' + 2 \cdot f_{xy} \cdot x' \cdot y' + f_{yy} \cdot (y')^2 + f_y \cdot y''. \end{aligned}$$

Now observe that the equation

$$❷ \qquad \gamma''(t) \perp \mathrm{T}_{\gamma(t)}$$

means that γ'' is perpendicular to the two basis vectors in $\mathrm{T}_{\gamma(t)}$. Therefore the vector equation ❷ can be rewritten as the following system of two real equations

$$\begin{cases} \langle \gamma'', s_x \rangle = 0, \\ \langle \gamma'', s_y \rangle = 0, \end{cases}$$

where $s(x, y) := (x, y, f(x, y))$, $x = x(t)$, and $y = y(t)$.

Observe that $s_x = (1, 0, f_x)$ and $s_y = (0, 1, f_y)$. Since $\gamma'' = (x'', y'', z'')$, we can rewrite the system in the following way.

$$\begin{cases} x'' + f_x \cdot z'' = 0, \\ y'' + f_y \cdot z'' = 0, \end{cases}$$

It remains use expression ❶ for z'', combine similar terms and simplify. □

Proof of 12.5. Let $z = f(x, y)$ be a description of Σ in tangent-normal coordinates at p. By Lemma 12.6, the condition $\gamma''(t) \perp \mathrm{T}_{\gamma(t)}$ can be written as a second order differential equation. Applying the existence and uniqueness of the initial value problem (0.25) we get existence and uniqueness of a geodesic γ in a small interval $(-\varepsilon, \varepsilon)$ for some $\varepsilon > 0$.

Let us extend γ to a maximal open interval $\mathbb{I}$. Suppose there is another geodesic γ_1 with the same initial data that is defined on a maximal open interval $\mathbb{I}_1$. Suppose γ_1 splits from γ at some time $t_0 > 0$; that is, γ_1 coincides with γ on the interval $[0, t_0)$, but they are different on the interval $[0, t_0 + \varepsilon)$ for any $\varepsilon > 0$. By continuity $\gamma_1(t_0) = \gamma(t_0)$ and $\gamma'(t_0) = \gamma'(t_0)$. Applying uniqueness of the initial value problem (0.25) again, we get that γ_1 coincides with γ in a small neighborhood of t_0 — a contradiction.

The case $t_0 < 0$ can be proved along the same lines. It follows that $\gamma_1 = \gamma$; in particular, $\mathbb{I}_1 = \mathbb{I}$.

Part *(a)* follows since the solution of the initial value problem depends smoothly on the initial data (0.25).

Assume *(b)* does not hold; that is, the maximal interval $\mathbb{I}$ is a proper subset of the real line $\mathbb{R}$. Without loss of generality, we may assume that $b = \sup \mathbb{I} < \infty$. (If not, switch the direction of γ.)

By 12.1 $|\gamma'|$ is constant, in particular $t \mapsto \gamma(t)$ is a uniformly continuous function. Therefore the limit point, $q = \lim_{t \to b} \gamma(t)$ is defined. Since Σ is a proper surface, $q \in \Sigma$.

Applying the argument above in a tangent-normal coordinate chart at q, we conclude that γ can be extended as a geodesic beyond q. Therefore $\mathbb{I}$ is not a maximal interval — a contradiction. □

12.7. Exercise. *Let Σ be a smooth torus of revolution; that is, a smooth surface of revolution with closed generatrix. Show that any closed geodesic on Σ is noncontractible.*

(In other words, if $s\colon \mathbb{R}^2 \to \Sigma$ is the natural bi-periodic parametrization of Σ, then there is no closed curve γ in $\mathbb{R}^2$ such that $s \circ \gamma$ is a geodesic.)

C Exponential map

Let Σ be a smooth regular surface and $p \in \Sigma$. Given a tangent vector $\mathrm{v} \in \mathrm{T}_p$, consider a geodesic γ_{v} in Σ that starts at p with initial velocity v; that is, $\gamma(0) = p$ and $\gamma'(0) = \mathrm{v}$.

The map

$$\exp_p\colon \mathrm{v} \mapsto \gamma_{\mathrm{v}}(1)$$

is called *exponential*.[1] By 12.5, the map $\exp_p\colon \mathrm{T}_p \to \Sigma$ is smooth and defined in a neighborhood of zero in the tangent plane T_p. Moreover, if Σ is proper, then $\exp_p$ is defined on the entire plane T_p.

The exponential map is defined on the tangent plane, which is a smooth surface, and its target is another smooth surface. Observe that one can identify the plane T_p with its tangent plane $\mathrm{T}_0\mathrm{T}_p$ so the differential $d_0(\exp_p)\colon \mathrm{v} \mapsto D_{\mathrm{v}} \exp_p$ maps T_p to itself. Furthermore, note that by

[1]There is a good reason to call this map *exponential*, but it is far from the subject.

Lemma 12.1, this differential is the identity map; that is, $d_0 \exp_p(\mathrm{v}) = \mathrm{v}$ for any $\mathrm{v} \in \mathrm{T}_p$.

Summarizing, we get the following statement:

12.8. Observation. *Let Σ be a smooth surface and $p \in \Sigma$. Then*

(a) $\exp_p$ *is a smooth map and its domain contains a neighborhood of the origin in* T_p*,*

(b) the differential $d_0(\exp_p)\colon \mathrm{T}_p \to \mathrm{T}_p$ *is the identity map.*

In fact it is easy to see that $\mathrm{Dom}(\exp_p)$ — the domain of definition of $\exp_p$ — is an open *star-shaped* region of T_p; the latter means that if $\mathrm{v} \in \mathrm{Dom}(\exp_p)$, then $\lambda{\cdot}\mathrm{v} \in \mathrm{Dom}(\exp_p)$ for any $0 \leqslant \lambda \leqslant 1$. Note also that 12.5*b* implies that *if Σ is proper, then* $\mathrm{Dom}(\exp_p) = \mathrm{T}_p$.

12.9. Proposition. *Let p be a point on smooth surface Σ (without boundary). Then there is $r_p > 0$ such that the exponential map* $\exp_p$ *is defined on the open ball $B_p = B(0, r_p)_{\mathrm{T}_p}$ and the restriction* $\exp_p|_{B_p}$ *is a smooth regular parametrization of a neighborhood of p in Σ.*

Moreover we have a local control *on r_p; that is, for any $q \in \Sigma$ there is $\varepsilon > 0$ such that if $|p - q|_\Sigma < \varepsilon$, then $r_p \geqslant \varepsilon$.*

The proof of the proposition uses the observation and the inverse function theorem (0.21).

Proof. Let $z = f(x, y)$ be a local graph representation of Σ in tangent-normal coordinates at p. Note that the (x, y)-plane coincides with the tangent plane T_p.

Denote by s the composition of the exponential map $\exp_p$ with the orthogonal projection $(x, y, z) \mapsto (x, y)$. By 12.8, the differential $d_0 s$ is the identity; in other words, the Jacobian matrix of this map at 0 is the identity. Applying the inverse function theorem (0.21) we get the first part of the proposition.

The second part of the inverse function theorem (0.21) guarantees that for a given point $q \in \Sigma$, the radii of the balls B_p for all points p sufficiently close to q can be taken uniformly bounded below by a positive number $\varepsilon_q > 0$. $\square$

Given $p \in \Sigma$, the least upper bound on r_p that satisfies 12.9 is called the *injectivity radius* of Σ at p; it is denoted by $\mathrm{inj}(p)$. The proposition states that *injectivity radius is positive and locally bounded away from zero.* In fact, the function $\mathrm{inj}\colon \Sigma \to (0, \infty]$ *is continuous*; the latter was proved by Wilhelm Klingenberg [35, 5.4].

The proof of the following statement will be indicated in 15.5.

12.10. Proposition. *Let p be a point on smooth surface Σ (without boundary). If* $\exp_p$ *is injective in $B_p = B(0, r)_{\mathrm{T}_p}$, then the restriction* $\exp_p|_{B_p}$ *is a diffeomorphism between B_p and its image in Σ.*

In other words, the injectivity radius at p can be defined as the least upper bound on the r such that $\exp_p$ *is injective in the ball* $B(0,r)_{\mathrm{T}_p}$.

D Shortest paths are geodesics

12.11. Proposition. *Let Σ be a smooth regular surface. Then any shortest path γ in Σ parametrized proportional to its arc-length is a geodesic in Σ. In particular, γ is a smooth curve.*

A partial converse to the first statement also holds: a sufficiently short arc of any geodesic is a shortest path. More precisely, any point p in Σ has a neighborhood U such that any geodesic that lies entirely in U is a shortest path.

As one can see from the following exercise, a geodesics might fail to be a shortest path. A geodesic that is also a shortest path is called a *minimizing geodesic*.

12.12. Exercise. *Let Σ be the cylindrical surface described by the equation $x^2+y^2=1$. Show that the helix $\gamma\colon[0,2{\cdot}\pi]\to\Sigma$ defined by $\gamma(t)=$ $=(\cos t,\sin t,t)$ is a geodesic, but not a shortest path on Σ.*

A formal proof of the proposition will be given later; see Section 15B.

The following informal physical explanation might be sufficiently convincing. In fact, if one assumes that γ is smooth, then it is easy to convert this explanation into a rigorous proof.

Informal explanation. Let us think about a shortest path γ as a stable position of a stretched elastic thread that is forced to lie on a frictionless surface. Since it is frictionless, the force density $\mathrm{N}=\mathrm{N}(t)$ that keeps γ in the surface must be proportional to the normal vector to the surface at $\gamma(t)$.

The tension in the thread has to be the same at all points (otherwise the thread would move back or forth and it would not be stable). Denote the tension by τ.

We can assume that γ has unit speed; in this case the net force from tension along the arc $\gamma_{[t_0,t_1]}$ is $\tau{\cdot}(\gamma'(t_1)-\gamma'(t_0))$. Hence the density of the net force from tension at t_0 is

$$\begin{aligned}\mathrm{F}(t_0)&=\lim_{t_1\to t_0}\tau{\cdot}\frac{\gamma'(t_1)-\gamma'(t_0)}{t_1-t_0}=\\&=\tau{\cdot}\gamma''(t_0).\end{aligned}$$

According to Newton's second law of motion, we have $\mathrm{F}+\mathrm{N}=0$. The latter implies that $\gamma''(t)\perp\mathrm{T}_{\gamma(t)}\Sigma$. $\square$

12.13. Corollary. *Let Σ be a smooth regular surface, $p \in \Sigma$ and $r \leqslant \mathrm{inj}(p)$. Then the exponential map $\exp_p$ is a diffeomorphism from $B(0,r)_{\mathrm{T}_p}$ to $B(p,r)_\Sigma$.*

Proof. By 12.10, the restriction of $\exp_p$ to $B_p = B(0,r)_{\mathrm{T}_p}$ is a diffeomorphism to its image $\exp_p(B_p) \subset \Sigma$.

Evidently $B(p,r)_\Sigma \supset \exp_p(B_p)$. By 12.11, $B(p,r)_\Sigma \subset \exp_p(B_p)$, hence the result. □

According to the corollary, the restriction $\exp_p|_{\mathrm{T}_p}$ admits an inverse map called the *logarithmic map at p*; it is denoted by

$$\log_p \colon B(p,r)_\Sigma \to B(0,r)_{\mathrm{T}_p}.$$

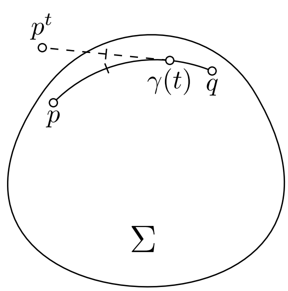
Note that according to the proposition above, any shortest path parametrized by its arc-length is a smooth curve. This observation should help the reader to solve the following two exercises.

12.14. Exercise. *Show that two distinct shortest paths can cross each other at most once. More precisely, if two shortest paths have two distinct common points p and q, then either these points are the ends of both shortest paths or both shortest paths share a common arc from p to q.*

Show by an example that nonoverlapping geodesics can cross each other an arbitrary number of times.

12.15. Exercise. *Assume that a smooth regular surface Σ is mirror symmetric with respect to a plane Π. Show that no shortest path α in Σ can* cross Π *more than once.*

In other words, if you travel along α, then you go from one side of Π *to the other at most once.*



12.16. Advanced exercise. *Let Σ be a smooth closed strictly convex surface in $\mathbb{R}^3$ and $\gamma\colon [0,\ell] \to \Sigma$ be a unit-speed minimizing geodesic. Set $p = \gamma(0)$, $q = \gamma(\ell)$ and*

$$p^t = \gamma(t) - t\cdot\gamma'(t),$$

where $\gamma'(t)$ denotes the velocity vector of γ at t.

Show that for any $t \in (0,\ell)$, one cannot see *q from p^t; that is, the line segment $[p^t, q]$ intersects Σ at a point distinct from q.*

Show that the statement does not hold without assuming that γ is minimizing.

E Liberman's lemma

The following lemma is a smooth analog of a lemma proved by Joseph Liberman [56].

12.17. Lemma. *Let f be a smooth convex function defined on an open subset of the plane. Suppose that $t \mapsto \gamma(t) = (x(t), y(t), z(t))$ is a unit-speed geodesic on the graph $z = f(x, y)$. Then $t \mapsto z(t)$ is a convex function; that is, $z''(t) \geqslant 0$ for any t.*

Proof. Choose the orientation on the graph so that the unit normal vector ν always points up; that is, ν has positive z-coordinate. Let us use the shortcut $\nu(t)$ for $\nu(\gamma(t))$.

Since γ is a geodesic, we have $\gamma''(t) \perp \mathrm{T}_{\gamma(t)}$, or equivalently $\gamma''(t)$ is proportional to $\nu(t)$ for any t. Furthermore,

$$\gamma'' = k \cdot \nu,$$

where $k = k(t)$ is the normal curvature at $\gamma(t)$ in the direction of $\gamma'(t)$.

Therefore

$$❶ \qquad z'' = k \cdot \cos\theta,$$

where $\theta = \theta(t)$ denotes the angle between $\nu(t)$ and the z-axis.

Since ν points up, we have $\theta(t) < \frac{\pi}{2}$, or equivalently

$$\cos\theta > 0.$$

Since f is convex, we have that the tangent plane supports the graph from below at any point; in particular $k(t) \geqslant 0$ for any t. It follows that the right hand side in ❶ is nonnegative; whence the statement follows. □

12.18. Exercise. *Assume γ is a unit-speed geodesic on a smooth convex surface Σ and a point p lies in the interior of the convex set bounded by Σ. Set $\rho(t) = |p - \gamma(t)|^2$. Show that $\rho''(t) \leqslant 2$ for any t.*

F Total curvature of geodesics

Recall that $\Phi(\gamma)$ denotes the total curvature of curve γ.

12.19. Exercise. *Let γ be a geodesic on an oriented smooth surface Σ with unit normal field ν. Show that*

$$\mathrm{length}(\nu \circ \gamma) \geqslant \Phi(\gamma).$$

12.20. Theorem. *Assume Σ is the graph $z = f(x,y)$ of a convex ℓ-Lipschitz function f defined on an open set in the (x,y)-plane. Then the total curvature of any geodesic in Σ is at most $2{\cdot}\ell$.*

This theorem was first proved by Vladimir Usov [81], an amusing generalization was found by David Berg [8].

Proof. Let $t \mapsto \gamma(t) = (x(t), y(t), z(t))$ be a unit-speed geodesic on Σ. According to Liberman's lemma, the function $t \mapsto z(t)$ is convex.

Since the slope of f is at most ℓ, we have

$$|z'(t)| \leqslant \frac{\ell}{\sqrt{1+\ell^2}}.$$

If γ is defined on the interval $[a,b]$, then

❶
$$\begin{aligned}\int\limits_a^b z''(t)dt = z'(b) - z'(a) \leqslant \\ \leqslant 2{\cdot}\frac{\ell}{\sqrt{1+\ell^2}}.\end{aligned}$$

Also note that z'' is the projection of γ'' to the z-axis. Since f is ℓ-Lipschitz, the tangent plane $\mathrm{T}_{\gamma(t)}\Sigma$ cannot have a slope greater than ℓ for any t. Because γ'' is perpendicular to that plane, we have that

$$|\gamma''(t)| \leqslant z''(t){\cdot}\sqrt{1+\ell^2}.$$

By ❶, we get that

$$\begin{aligned}\Phi(\gamma) = \int\limits_a^b |\gamma''(t)|{\cdot}dt \leqslant \\ \leqslant \sqrt{1+\ell^2}{\cdot}\int\limits_a^b z''(t){\cdot}dt \leqslant \\ \leqslant 2{\cdot}\ell.\end{aligned}$$

□

12.21. Exercise. *Note that the graph $z = \ell{\cdot}\sqrt{x^2+y^2}$ with the origin removed is a smooth surface; denote it by Σ. Show that any both-side-infinite geodesic γ in Σ has total curvature exactly $2{\cdot}\ell$.*

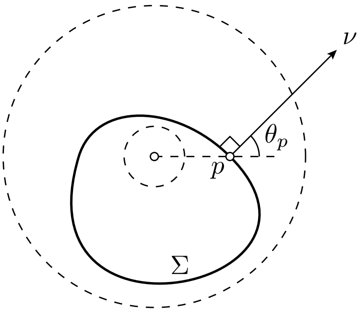
Ihe previous exercise implies that the estimate in Usov's theorem is optimal. To see this, mollify the function $f(x,y) = \ell{\cdot}\sqrt{x^2+y^2}$ in a small neighborhood of the origin while keeping it convex and ℓ-Lipschitz. Note that we can assume that the geodesic γ does not enter the smoothed part of the graph.

12.22. Exercise. *Assume f is a smooth convex $\tfrac{3}{2}$-Lipschitz function defined on the (x,y)-plane. Show that any geodesic γ on the graph $z = f(x,y)$ is simple; that is, it has no self-intersections.*

Construct a convex 2-Lipschitz function defined on the (x,y)-plane with a nonsimple geodesic γ in its graph $z = f(x,y)$.

12.23. Theorem. *Suppose a smooth surface Σ bounds a convex set K in the Euclidean space. Assume $B(0,\varepsilon) \subset K \subset B(0,1)$. Then the total curvatures of any shortest path in Σ can be bounded in terms of ε.*

The following exercise will guide you thru the proof of the theorem.



12.24. Exercise. *Let Σ be as in the theorem and γ be a unit-speed shortest path in Σ. Denote by ν_p the unit normal vector that points outside of Σ; denote by θ_p the angle between ν_p and the direction from the origin to the point $p \in \Sigma$. Set $\rho(t) = |\gamma(t)|^2$; denote by $k(t)$ the curvature of γ at t.*

(a) Show that $\cos\theta_p \geqslant \varepsilon$ for any $p \in \Sigma$.

(b) Show that $|\rho'(t)| \leqslant 2$ for any t.

(c) Show that

$$\rho''(t) = 2 - 2{\cdot}k(t){\cdot}\cos\theta_{\gamma(t)}{\cdot}|\gamma(t)|$$

for any t.

(d) Use the closest-point projection from the unit sphere to Σ to show that

$$\operatorname{length}\gamma \leqslant \pi.$$

(e) Conclude that $\Phi(\gamma) \leqslant 100/\varepsilon^2$.

The obtained bound of the total curvature goes to infinity as $\varepsilon \to 0$, but there is a bound that is independent of ε; it is a result of Nina Lebedeva and the first author [54]. According to Exercise 12.19, this statement would also follow if the length of the spherical image of γ could be bounded above; that is, if $\operatorname{length}(\nu\circ\gamma) \leqslant C$ for a universal constant C. The latter was conjectured by Aleksei Pogorelov [69]; counterexamples to the different forms of this conjecture were found by Viktor Zalgaller in [83], Anatoliy Milka in [60] and Vladimir Usov in [81]; these results were partly rediscovered later by János Pach [64].

Chapter 13

Parallel transport

A Parallel tangent fields

Let Σ be a smooth surface in the Euclidean space and $\gamma\colon [a,b]\to\Sigma$ be a smooth curve. A smooth vector-valued function $t\mapsto \mathrm{v}(t)\in\mathbb{R}^3$ is called a *tangent field* along γ if the vector $\mathrm{v}(t)$ lies in the tangent plane $\mathrm{T}_{\gamma(t)}\Sigma$ for each t.

A tangent field $\mathrm{v}(t)$ on γ is called *parallel* if $\mathrm{v}'(t)\perp\mathrm{T}_{\gamma(t)}$ for any t.

In general, the family of tangent planes $\mathrm{T}_{\gamma(t)}\Sigma$ is not parallel. Therefore one cannot expect to have a *truly* parallel family $\mathrm{v}(t)$ with $\mathrm{v}'\equiv 0$. The condition $\mathrm{v}'(t)\perp\mathrm{T}_{\gamma(t)}$ means that the family is as parallel as possible — it rotates together with the tangent plane, but does not rotate inside the plane.

Note that by the definition of geodesic, the velocity field $\mathrm{v}(t)=\gamma'(t)$ of any geodesic γ is parallel on γ.

13.1. Exercise. *Let Σ be a smooth regular surface in the Euclidean space, $\gamma\colon[a,b]\to\Sigma$ a smooth curve. Suppose that $\mathrm{v}(t)$, $\mathrm{w}(t)$ are parallel vector fields along γ.*

(a) Show that $|\mathrm{v}(t)|$ is constant.

(b) Show that the angle $\theta(t)$ between $\mathrm{v}(t)$ and $\mathrm{w}(t)$ is constant.

B Parallel transport

Let Σ be a smooth surface in the Euclidean space and $\gamma\colon[a,b]\to\Sigma$ be a smooth curve. Assume $p=\gamma(a)$ and $q=\gamma(b)$.

Given a tangent vector $\mathrm{v}\in\mathrm{T}_p$ there is a unique parallel field $\mathrm{v}(t)$ along γ such that $\mathrm{v}(a)=\mathrm{v}$. The latter follows from 0.25; the uniqueness also follows from Exercise 13.1.

The vector $\mathrm{w} = \mathrm{v}(b) \in \mathrm{T}_q$ is called the *parallel transport* of v along γ in Σ.

The parallel transport along γ will be denoted by ι_γ; so we can write $\mathrm{w} = \iota_\gamma(\mathrm{v})$ or $\mathrm{w} = \iota_\gamma(\mathrm{v})_\Sigma$ if we need to emphasize that γ lies in the surface Σ. From Exercise 13.1, it follows that parallel transport $\iota_\gamma\colon \mathrm{T}_p \to \mathrm{T}_q$ is an isometry. In general, the parallel transport $\iota_\gamma\colon \mathrm{T}_p \to \mathrm{T}_q$ depends on the choice of γ; that is, for another curve γ_1 connecting p to q in Σ, the parallel transports ι_{γ_1} and ι_γ might be different.

Suppose that γ_1 and γ_2 are two smooth curves in two smooth surfaces Σ_1 and Σ_2. Denote by $\nu_i\colon \Sigma_i \to \mathbb{S}^2$ the Gauss maps of Σ_1 and Σ_2. If $\nu_1 \circ \gamma_1(t) = \nu_2 \circ \gamma_2(t)$ for any t, then we say that the curves γ_1 and γ_2 have *identical spherical images* in Σ_1 and Σ_2 respectively.

In this case the tangent plane $\mathrm{T}_{\gamma_1(t)}\Sigma_1$ is parallel to $\mathrm{T}_{\gamma_2(t)}\Sigma_2$ for any t and so we can identify $\mathrm{T}_{\gamma_1(t)}\Sigma_1$ with $\mathrm{T}_{\gamma_2(t)}\Sigma_2$. In particular, if $\mathrm{v}(t)$ is a tangent vector field along γ_1, then it is also a tangent vector field along γ_2. Moreover $\mathrm{v}'(t) \perp \mathrm{T}_{\gamma_1(t)}\Sigma_1$ is equivalent to $\mathrm{v}'(t) \perp \mathrm{T}_{\gamma_2(t)}\Sigma_2$; that is, if $\mathrm{v}(t)$ is a parallel vector field along γ_1, then it is also a parallel vector field along γ_2.

The dicussion above leads to the following observation that will play a key role in the sequel.

13.2. Observation. *Let γ_1 and γ_2 be two smooth curves in two smooth surfaces Σ_1 and Σ_2. Suppose that γ_1 and γ_2 have identical spherical images in Σ_1 and Σ_2 respectively. Then the parallel transports ι_{γ_1} and ι_{γ_2} are identical.*

13.3. Exercise. *Let Σ_1 and Σ_2 be two surfaces with a common curve γ. Suppose that Σ_1 bounds a region that contains Σ_2. Show that the parallel translation along γ in Σ_1 coincides with the parallel translation along γ in Σ_2.*

C Bike wheel and projections

In this section we describe two interpretations of parallel transport; they might help to build the right intuition, but will not help in writing rigorous proofs. The first interpretation, the *bike wheel*, comes from physics and it was suggested by Mark Levi [55]. The second one is via orthogonal projections of tangent planes.

Think of walking along γ and carrying a perfectly balanced horizontal bike wheel. Imagine that you keep its axis normal to Σ and touch only the tip of the axis. By Newton's first law of motion, if the wheel is not spinning at the starting point p, then it will not be spinning after stopping at q. (Indeed, by pushing the axis one cannot produce torque to spin the wheel.) The map that sends the initial position of the wheel to the final position is the parallel transport ι_γ.

The observation above essentially states that *moving the axis of the wheel without changing its direction does not change the direction of the wheel's spikes.*

On a more formal level, one can choose a partition $a = t_0 < \ldots < < t_n = b$ of $[a, b]$ and consider the sequence of orthogonal projections $\varphi_i \colon \mathrm{T}_{\gamma(t_{i-1})} \to \mathrm{T}_{\gamma(t_i)}$. For a fine partition, the composition

$$\varphi_n \circ \cdots \circ \varphi_1 \colon \mathrm{T}_p \to \mathrm{T}_q$$

gives an approximation of ι_γ.

(Note that each φ_i does not increase the magnitude of a vector and neither the composition. It is straightforward to see that if the partition is sufficiently fine, then the above composition is almost an isometry; in particular it almost preserves the magnitudes of tangent vectors.)

13.4. Exercise. *Construct a loop γ in $\mathbb{S}^2$ with base at p such that the parallel transport $\iota_\gamma \colon \mathrm{T}_p \to \mathrm{T}_p$ is not the identity map.*

D Total geodesic curvature

Recall that geodesic curvature is defined in Section 9E. It measures how much a given curve γ diverges from being a geodesic; it is positive where γ turns left and negative when γ turns right. In particular, by the following exercise, geodesics have vanishing geodesic curvature.

The total geodesic curvature is defined as the integral

$$\Psi(\gamma) := \int\limits_{\mathbb{I}} k_g(t) \cdot dt,$$

assuming that γ is a smooth unit-speed curve defined on the real interval $\mathbb{I}$.

Note that if Σ is a plane and γ lies in Σ, then the geodesic curvature of γ coincides with its signed curvature and therefore its total geodesic curvature is equal to its total signed curvature. By that reason, we use the same notation $\Psi(\gamma)$ as for total signed curvature; if we need to emphasize that we consider γ as a curve in Σ, we write $\Psi(\gamma)_\Sigma$.

If γ is a piecewise smooth regular curve in Σ, then its total geodesic curvature is defined as the sum of all total geodesic curvatures of its arcs plus the sum of the signed exterior angles of γ at the joints. More precisely, if γ is a concatenation of smooth regular curves $\gamma_1, \ldots, \gamma_n$, then

$$\Psi(\gamma) = \Psi(\gamma_1) + \cdots + \Psi(\gamma_n) + \theta_1 + \cdots + \theta_{n-1},$$

where θ_i is the signed external angle at the joint between γ_i and γ_{i+1}; it is positive if γ turns left and negative if γ turns right, it is undefined if γ turns to the opposite direction. If γ is closed, then

$$\Psi(\gamma) = \Psi(\gamma_1) + \dots + \Psi(\gamma_n) + \theta_1 + \dots + \theta_n,$$

where θ_n is the signed external angle at the joint γ_n and γ_1.

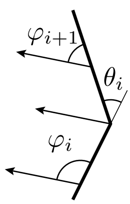
If each arc γ_i in the concatenation is a minimizing geodesic, then γ is called a *broken geodesic*. In this case $\Psi(\gamma_i) = 0$ for each i and therefore the total geodesic curvature of γ is the sum of its signed external angles.

13.5. Proposition. *Assume γ is a closed broken geodesic in a smooth oriented surface Σ that starts and ends at the point p. Then the parallel transport $\iota_\gamma\colon \mathrm{T}_p \to \mathrm{T}_p$ is a clockwise rotation of the plane T_p by the angle $\Psi(\gamma)$.*

Moreover, the same statement holds true for smooth closed curves and piecewise smooth curves.

Proof. Assume γ is a cyclic concatenation of geodesics $\gamma_1, \dots, \gamma_n$. Fix a tangent vector v at p and extend it to a parallel vector field along γ. Since $\mathrm{T}_i(t) = \gamma_i'(t)$ is parallel along γ_i, the angle φ_i from T_i to v stays constant along each γ_i.

If θ_i denotes the external angle at the vertex in which we switch from γ_i to γ_{i+1}, we have that



$$\varphi_{i+1} = \varphi_i - \theta_i \pmod{2\cdot\pi}.$$

Therefore after going over all vertices we get that

$$\varphi_{n+1} - \varphi_1 = -\theta_1 - \dots - \theta_n = -\Psi(\gamma).$$

Hence the first statement follows.

For a smooth unit-speed curve $\gamma\colon [a,b] \to \Sigma$, the proof is analogous. Denote by $\varphi(t)$ is the signed angle from $\mathrm{v}(t)$ to $\mathrm{T}(t)$. Let us show that

$$\text{❶} \qquad \varphi'(t) + k_g(t) \equiv 0$$

Recall that $\mu = \mu(t)$ denotes the counterclockwise rotation of $\mathrm{T} = \mathrm{T}(t)$ by angle $\frac{\pi}{2}$ in $\mathrm{T}_{\gamma(t)}$. Denote by $\mathrm{w} = \mathrm{w}(t)$ the counterclockwise rotation of $\mathrm{v} = \mathrm{v}(t)$ by angle $\frac{\pi}{2}$ in $\mathrm{T}_{\gamma(t)}$. Then

$$\mathrm{T} = \cos\varphi\cdot\mathrm{v} - \sin\varphi\cdot\mathrm{w},$$
$$\mu = \sin\varphi\cdot\mathrm{v} + \cos\varphi\cdot\mathrm{w}.$$

Note that W is a parallel vector field along γ; that is, $\mathrm{V}'(t), \mathrm{W}'(t) \perp \perp \mathrm{T}_{\gamma(t)}$. Therefore $\langle \mathrm{V}', \mu \rangle = \langle \mathrm{W}', \mu \rangle = 0$. It follows that

$$\begin{aligned} k_g &= \langle \mathrm{T}', \mu \rangle = \\ &= -(\sin^2\varphi + \cos^2\varphi)\cdot\varphi'. \end{aligned}$$

Whence ❶ follows.

By ❶ we get that

$$\begin{aligned} \varphi(b) - \varphi(a) &= \int\limits_a^b \varphi'(t)\cdot dt = \\ &= -\int\limits_a^b k_g\cdot dt = \\ &= -\Psi(\gamma) \end{aligned}$$

In the case when γ is a piecewise regular smooth curve, the result follows from a straightforward combination of the above two cases. $\square$

Chapter 14

Gauss–Bonnet formula

A Formulation

The following theorem was proved by Carl Friedrich Gauss [34] for geodesic triangles; Pierre Bonnet and Jacques Binet independently generalized the statement for arbitrary curves. A generalized formula (14.13) was proved by Walther von Dyck.

14.1. Theorem. *Let Δ be a topological disc in a smooth oriented surface Σ bounded by a simple piecewise smooth and regular curve $\partial\Delta$. Suppose that $\partial\Delta$ is oriented in such a way that Δ lies on its left. Then*

$$\Psi(\partial\Delta) + \iint\limits_{\Delta} K = 2{\cdot}\pi, \qquad ❶$$

where K denotes the Gauss curvature of Σ.

We will give an informal proof of this formula in a leading special case. A formal computational proof will be given in Section 15E.

Before going into the proofs, we suggest solving the following exercises using the Gauss–Bonnet formula.

14.2. Exercise. *Assume γ is a closed simple curve with constant geodesic curvature 1 in a smooth closed surface Σ with positive Gauss curvature. Show that*

$$\operatorname{length}\gamma \leqslant 2{\cdot}\pi;$$

that is, the length of γ cannot exceed the length of the unit circle in the plane.

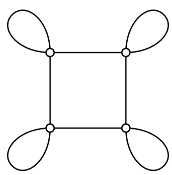
14.3. Exercise. *Let γ be a simple closed geodesic on a smooth closed surface Σ with positive Gauss curvature. Assume $\nu\colon \Sigma \to \mathbb{S}^2$ is a Gauss* *map. Show that the curve $\alpha=\nu\circ\gamma$ divides the sphere into two regions of equal area. Conclude that* length $\alpha\geqslant 2\cdot\pi$.

14.4. Exercise. *Let γ be a closed geodesic on a smooth closed surface Σ with positive Gauss curvature. Suppose that R is one of the regions that γ cuts from Σ. Show that*

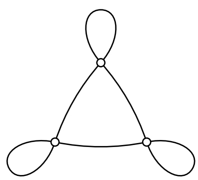
$$\iint_R K\leqslant 2\cdot\pi.$$

Conclude that any two closed geodesics on Σ have a common point.

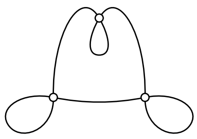
14.5. Exercise. *Let Σ be a smooth regular sphere with positive Gauss curvature and $p\in\Sigma$. Suppose γ is a closed geodesic whose image is covered by a single coordinate chart. Show that γ cannot look like one of the curves on the following diagrams.*



(easy)



(tricky)



In fact γ also cannot look like the curve on the right, but the proof requires a more advanced technique; see 16.9.

The following exercise gives the optimal bound on the Lipschitz constant of a convex function that guarantees that its geodesics have no self-intersections; compare to 12.22.

14.6. Exercise. *Suppose that $f\colon\mathbb{R}^2\to\mathbb{R}$ is a $\sqrt{3}$-Lipschitz smooth convex function. Show that any geodesic in the graph $z=f(x,y)$ has no self-intersections.*

A surface Σ is called *simply-connected* if any closed simple curve in Σ bounds a disc. Equivalently any closed curve in Σ can be continuously deformed into a *trivial curve*; that is, a curve that stands at one point all the time.

Observe that a plane or a sphere are examples of simply-connected surfaces, while the torus or the cylinder are not simply-connected.

14.7. Exercise. *Suppose that Σ is a simply-connected open surface with nonpositive Gauss curvature.*

(a) *Show that any two points in Σ are connected by a unique geodesic.*

(b) *Conclude that for any point $p \in \Sigma$, the exponential map $\exp_p$ is a diffeomorphism from the tangent plane T_p to Σ. In particular Σ is diffeomorphic to the plane.*

B Additivity

Let Δ be a topological disc in a smooth oriented surface Σ bounded by a simple piecewise smooth and regular curve $\partial\Delta$. As before we suppose that $\partial\Delta$ is oriented in such a way that Δ lies on its left. Set

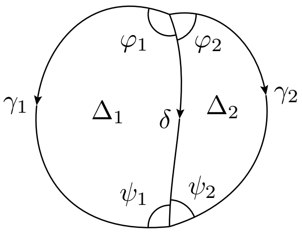
❶
$$\mathrm{GB}(\Delta) = \Psi(\partial\Delta) + \iint\limits_{\Delta} K - 2\cdot\pi,$$

where K denotes the Gauss curvature of Σ. Here GB stands for Gauss–Bonnet formula; our goal is to show that

$$\mathrm{GB}(\Delta) = 0.$$

14.8. Lemma. *Suppose that the disc Δ is subdivided into two discs Δ_1 and Δ_2 by a curve δ. Then*

$$\mathrm{GB}(\Delta) = \mathrm{GB}(\Delta_1) + \mathrm{GB}(\Delta_2).$$



Proof. Let us subdivide $\partial\Delta$ into two curves γ_1 and γ_2 that share endpoints with δ so that Δ_i is bounded by the arc γ_i and δ for $i = 1, 2$.

Denote by φ_1, φ_2, ψ_1, and ψ_2 the angles between δ and γ_i marked on the diagram. Suppose that the arcs γ_1, γ_2, and δ are oriented as on the diagram. Then

$$\begin{aligned}
\Psi(\partial\Delta) &= \Psi(\gamma_1) - \Psi(\gamma_2) + (\pi - \varphi_1 - \varphi_2) + (\pi - \psi_1 - \psi_2),\\
\Psi(\partial\Delta_1) &= \Psi(\gamma_1) - \Psi(\delta) + (\pi - \varphi_1) + (\pi - \psi_1),\\
\Psi(\partial\Delta_1) &= \Psi(\delta) - \Psi(\gamma_2) + (\pi - \varphi_2) + (\pi - \psi_2),\\
\iint\limits_{\Delta} K &= \iint\limits_{\Delta_1} K + \iint\limits_{\Delta_2} K.
\end{aligned}$$

It remains to plug in the results in the formulas for $\mathrm{GB}(\Delta)$, $\mathrm{GB}(\Delta_1)$, and $\mathrm{GB}(\Delta_2)$. $\square$

C Spherical case

Note that if Σ is a plane, then the Gauss curvature vanishes; therefore the Gauss–Bonnet formula ❶ can be written as

$$\Psi(\partial\Delta) = 2{\cdot}\pi,$$

and it follows from 5.5. In other words, $\mathrm{GB}(\Delta) = 0$ for any disc Δ in the plane.

If Σ is the unit sphere, then $K \equiv 1$; in this case Theorem 14.1 can be formulated in the following way:

14.9. Proposition. *Let P be a spherical polygon bounded by a simple closed broken geodesic ∂P. Assume ∂P is oriented such that P lies on the left of ∂P. Then*

$$\mathrm{GB}(P) = \Psi(\partial P) + \operatorname{area} P - 2{\cdot}\pi = 0.$$

Moreover the same formula holds true for any spherical region bounded by a piecewise smooth simple closed curve.

This proposition will be used in the informal proof given below.

Sketch of proof. Suppose that a spherical triangle Δ has angles α, β, and γ. According to 0.13,

$$\operatorname{area}\Delta = \alpha + \beta + \gamma - \pi.$$

Recall that $\partial\Delta$ is oriented so that Δ lies on its left. Then its oriented external angles are $\pi - \alpha$, $\pi - \beta$ and $\pi - \gamma$. Therefore

$$\Psi(\partial\Delta) = 3{\cdot}\pi - \alpha - \beta - \gamma.$$

It follows that $\Psi(\partial\Delta) + \operatorname{area}\Delta = 2{\cdot}\pi$ or, equivalently,

$$\mathrm{GB}(\Delta) = 0.$$

Note that we can subdivide a given spherical polygon P into triangles by dividing a polygon in two on each step. By the additivity lemma (14.8), we get

$$\mathrm{GB}(P) = 0$$

for any spherical polygon P.

The second statement can be proved by approximation. One has to show that one could approximate the total geodesic curvature of a piecewise smooth simple closed curve by the total geodesic curvature of inscribed broken geodesics. We omit the proof of the latter statement, but it goes along the same lines as 3.20. □

14.10. Exercise. *Assume γ is a simple piecewise smooth loop on $\mathbb{S}^2$ that divides its area in two equal parts. Denote by p the base point of γ. Show that the parallel transport $\iota_\gamma\colon \mathrm{T}_p\mathbb{S}^2 \to \mathrm{T}_p\mathbb{S}^2$ is the identity map.*

D Intuitive proof

In this section we prove a special case of the Gauss–Bonnet formula. This case is leading — the general case can be proved similarly, but one has to use the signed area counted with multiplicity.

Proof of 14.1 for proper surfaces with positive Gauss curvature. By 9.6, in this case, we have

$$\text{❶} \qquad \mathrm{GB}(\Delta) = \Psi(\partial\Delta) + \mathrm{area}[\nu(\Delta)] - 2{\cdot}\pi.$$

Fix $p \in \partial\Delta$; assume the loop α runs along $\partial\Delta$ so that Δ lies on the left from it. Consider the parallel translation $\iota_\alpha \colon \mathrm{T}_p \to \mathrm{T}_p$ along α. According to 13.5, ι_α is the clockwise rotation by the angle $\Psi(\alpha)_\Sigma$.

Set $\beta = \nu \circ \alpha$. By 13.2, we have $\iota_\alpha = \iota_\beta$, where β is considered as a curve in the unit sphere. In particular ι is a clockwise rotation by angle $\Psi(\beta)_{\mathbb{S}^2}$. By 14.9

$$\mathrm{GB}(\nu(\Delta)) = \Psi(\beta)_{\mathbb{S}^2} + \mathrm{area}[\nu(\Delta)] - 2{\cdot}\pi = 0.$$

Therefore ι is a counterclockwise rotation by $\mathrm{area}[\nu(\Delta)]$

Summarizing, the clockwise rotation by $\Psi(\alpha)_\Sigma$ is identical to a counterclockwise rotation by $\mathrm{area}[\nu(\Delta)]$. The rotations are identical if the angles are equal modulo $2{\cdot}\pi$. Therefore

$$\text{❷} \qquad \begin{aligned} \mathrm{GB}(\nu(\Delta)) &= \Psi(\partial\Delta)_\Sigma + \mathrm{area}[\nu(\Delta)] - 2{\cdot}\pi = \\ &= 2{\cdot}n{\cdot}\pi \end{aligned}$$

for some integer n.

It remains to show that $n = 0$. By 5.5, this is so for a topological disc in a plane. One can think of a general disc Δ as the result of a continuous deformation of a plane disc. The integer n cannot change in the process of deformation since the left hand side in ❷ is continuous along the deformation; whence $n = 0$ for the result of the deformation. □

E Simple geodesic

The following theorem provides an interesting application of the Gauss–Bonnet formula proved by Stephan Cohn-Vossen [Satz 9 in 23].

14.11. Theorem. *Any open smooth regular surface with positive Gauss curvature has a simple two-sided infinite geodesic.*

Proof. Let Σ be an open surface with positive Gauss curvature and γ a two-sided infinite geodesic in Σ.

If γ has a self-intersection, then it contains a simple loop; that is, a restriction $\ell = \gamma|_{[a,b]}$ is a simple loop for some closed interval $[a, b]$.

By 10.16, Σ is parametrized by an open convex region Ω in the plane. By Jordan's theorem (0.28), ℓ bounds a disc in Σ; denote it by Δ. If φ is the angle at the base of the loop, then by the Gauss–Bonnet formula,

$$\iint_{\Delta} K = \pi + \varphi.$$

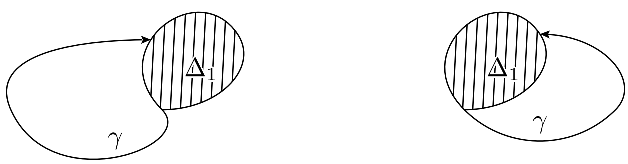
Recall that by 10.19*b*, we have

❶ $$\iint_{\Sigma} K \leqslant 2{\cdot}\pi.$$

Therefore $0 < \varphi < \pi$; that is, γ has no concave simple loops.

Assume γ has two simple loops, say ℓ_1 and ℓ_2 that bound discs Δ_1 and Δ_2. Then the discs Δ_1 and Δ_2 have to overlap, otherwise the curvature of Σ would exceed $2{\cdot}\pi$ contradicting ❶.

It follows that after leaving Δ_1, the geodesic γ has to enter it again before creating another simple loop. Consider the moment when γ enters



Δ_1 again; two possible scenarios are shown in the picture. On the left picture we get two nonoverlapping discs which, as we know, is an impossible scenario. The right picture is impossible as well — in this case we get a concave simple loop.

It follows that γ contains only one simple loop. This loop cuts a disc from Σ and goes around it either clockwise or counterclockwise. This way we divide all the self-intersecting geodesics on Σ into two sets which we call *clockwise* and *counterclockwise*.

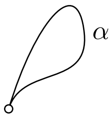
Note that the geodesic $t \mapsto \gamma(t)$ is clockwise if and only if the same geodesic traveled backwards $t \mapsto \gamma(-t)$ is counterclockwise.

Let us shoot a unit-speed geodesic at each directions from a given point $p = \gamma(0)$. this gives a one-parameter family of geodesics γ_s for $s \in [0, \pi]$ connecting the geodesic $t \mapsto \gamma(t)$ with $t \mapsto \gamma(-t)$; that is, $\gamma_0(t) = \gamma(t)$ and $\gamma_\pi(t) = \gamma(-t)$.

Observe that the subset of values $s \in [0, \pi]$ such that γ_s is right (or left) is open. That is, if γ_s is right, then so is γ_t for all t sufficiently close to s. Indeed, denote by φ_s the angle of the simple loop of γ_s. From above we have $0 < \varphi_s < \pi$. Therefore the self-intersection at the base of the loop of γ_s is transverse. It follows that the self-intersection survives in γ_t for all t sufficiently close to s.

Since $[0, \pi]$ is connected, it cannot be subdivided into two nonempty open sets. It follows that for some s, the geodesic γ_s is neither clockwise nor counterclockwise; that is, γ_s has no self-intersections. □

14.12. Exercise. *Let Σ be an open smooth regular surface with positive Gauss curvature. Suppose $\alpha\colon [0,1] \to \Sigma$ is a smooth regular loop such that $\alpha'(0) + \alpha'(1) = 0$. Show that there is a simple two-sided infinite geodesic γ that is tangent to α at some point.*



F General domains

14.13. Theorem. *Let Λ be a compact domain bounded by a finite collection (possibly empty) of simple piecewise smooth and regular curves $\gamma_1, \dots, \gamma_n$ in a smooth surface Σ. Suppose that each γ_i is oriented in such a way that Λ lies on its left. Then*

❶ $$\Psi(\gamma_1) + \dots + \Psi(\gamma_n) + \iint\limits_{\Lambda} K = 2{\cdot}\pi{\cdot}\chi$$

for an integer $\chi = \chi(\Lambda)$.

Moreover, if Λ can be subdivided into f discs by an embedded graph with v vertices and e edges then $\chi = v - e + f$.

The number $\chi = \chi(\Lambda)$ is called the *Euler characteristic* of Λ. Note that χ does not depend on the choice of the subdivision since the left hand side in ❶ does not.

Proof. Suppose a graph with v vertices and e edges subdivides Λ into f discs. Apply Gauss–Bonnet formula for each disc and sum up the results. Observe that each disc and each vertex contributes $2{\cdot}\pi$ and each edge contributes $-2{\cdot}\pi$ to the total sum. Whence ❶ follows. □

14.14. Exercise. *Find the integral of the Gauss curvature on each of the following surfaces:*

(a) Torus.
(b) Moebius band with geodesic boundary.
(c) Pair of pants with geodesic boundary components.
(d) Sphere with two handles.

Chapter 15

Semigeodesic charts

This chapter contains computational proofs of several statements discussed above, including

- ⋄ Proposition 12.10 — an alternative definition of injectivity radius.
- ⋄ Proposition 12.11 — shortest paths are geodesics.
- ⋄ The Gauss–Bonnet formula (14.1).

In addition, we discuss intrinsic isometries between surfaces and prove Gauss' remarkable theorem, stating that Gauss curvature can be defined intrinsically.

A Polar coordinates

The property of the exponential map in 12.9 can be used to define *polar coordinates* in a smooth surface Σ with respect to a point $p \in \Sigma$.

Namely, fix some polar coordinates $(r, \theta) \in \mathbb{R}^+ \times \mathbb{S}^1$ on the tangent plane T_p. If $\mathrm{v} \in \mathrm{T}_p$ has coordinates (r, θ), then we say that $s(r, \theta) = \exp_p \mathrm{v}$ is the point in Σ with polar coordinates (r, θ).

Since there might be many geodesics from p to a given point x, the point x might have many different polar coordinates. However, according to Proposition 12.9 polar coordinates behave in the usual way for small values of r. More precisely, the following statement holds:

15.1. Observation. *Let $(r, \theta) \mapsto s(r, \theta)$ describe polar coordinates with respect to a point p in a smooth surface Σ. Then there is $r_0 > 0$ such that s is regular at any pair (r, θ) with $0 < r < r_0$.*

Moreover if $0 \leqslant r_1, r_2 < r_0$, then $s(r_1, \theta_1) = s(r_2, \theta_2)$ if and only if $r_1 = r_2 = 0$ or $r_1 = r_2$ and $\theta_1 = \theta_2 + 2 \cdot n \cdot \pi$ for an integer n.

The following statement will play a key role in the formal proof that shortest paths are geodesics; see Section 15B.

15.2. Gauss lemma. *Let $(r, \theta) \mapsto s(r, \theta)$ describe polar coordinates with respect to a point p in a smooth surface Σ. Then $s_\theta \perp s_r$ for any r and θ.*

Proof. Choose θ. By the definition of exponential map, the curve $\gamma(t) = s(t, \theta)$ is a unit-speed geodesic that starts at p; in particular, we have the following two identities:

(i) Since the geodesic γ has unit speed, we have $|s_r| = |\gamma'| = 1$. In particular,
$$\tfrac{\partial}{\partial\theta}\langle s_r, s_r\rangle = 0$$

(ii) Since γ is a geodesic, we have $s_{rr}(r, \theta) = \gamma''(r) \perp \mathrm{T}_{\gamma(r)}$ and therefore
$$\langle s_\theta, s_{rr}\rangle = 0$$

It follows that

$$\begin{aligned} \tfrac{\partial}{\partial r}\langle s_\theta, s_r\rangle &= \langle s_{\theta r}, s_r\rangle + \cancel{\langle s_\theta, s_{rr}\rangle} = \\ &= \tfrac{1}{2}\cdot\tfrac{\partial}{\partial\theta}\langle s_r, s_r\rangle = \\ &= 0. \end{aligned} \qquad ❶$$

Therefore, for fixed θ, the value $\langle s_\theta, s_r\rangle$ does not depend on r.

Note that $s(0, \theta) = p$ for any θ. Therefore $s_\theta(0, \theta) = 0$ and in particular

$$\langle s_\theta, s_r\rangle = 0$$

for $r = 0$. By ❶ the same holds for any r. □

B Shortest paths are geodesics

In this section we use the construction of polar coordinates and the Gauss lemma 15.2 to prove Proposition 12.11.

Proof of 12.11. Let $\gamma\colon [0, \ell] \to \Sigma$ be a shortest path parametrized by arc-length. Suppose $\ell = \operatorname{length}\gamma$ is sufficiently small, so γ can be described in the polar coordinates at p; say as $\gamma(t) = s(r(t), \theta(t))$ for some functions $t \mapsto \theta(t)$ and $t \mapsto r(t)$ with $r(0) = 0$.

Note that by the chain rule, we have

$$\gamma' = s_\theta\cdot\theta' + s_r\cdot r' \qquad ❶$$

whenever the left side is defined. By the Gauss lemma 15.2, $s_\theta \perp s_r$ and by definition of polar coordinates $|s_r| = 1$. Therefore ❶ implies

$$|\gamma'(t)| \geqslant r'(t). \qquad ❷$$

for any t where $\gamma'(t)$ is defined.

Since γ parametrized by arc-length, we have

$$|\gamma(t_2) - \gamma(t_1)| \leqslant |t_2 - t_1|.$$

In particular, γ is Lipschitz. Therefore by Rademacher's theorem (0.16) the derivative γ' is defined almost everywhere. By 2.5a, we have that

$$\begin{aligned}\operatorname{length}\gamma &= \int\limits_0^\ell |\gamma'(t)|\cdot dt \geqslant \\ &\geqslant \int\limits_0^\ell r'(t)\cdot dt = \\ &= r(\ell).\end{aligned}$$

Note that by the definition of polar coordinates, there is a geodesic of length $r(\ell)$ from $p = \gamma(0)$ to $q = \gamma(\ell)$. Since γ is a shortest path, we get that $r(\ell) = \ell$ and moreover $r(t) = t$ for any t. This equality holds if and only if we have equality in ❷ for almost all t. The latter implies that γ is a geodesic.

It remains to prove the partial converse.

Fix a point $p \in \Sigma$. Let $\varepsilon > 0$ be as in 12.9. Assume a geodesic γ of length less than ε from p to q does not minimize the length between its endpoints. Then there is a shortest path from p to q, which becomes a geodesic when it is parametrized by its arc-length. That is, there are two geodesics from p to q of length smaller than ε. In other words there are two vectors $\mathrm{v}, \mathrm{w} \in \mathrm{T}_p$ such that $|\mathrm{v}| < \varepsilon$, $|\mathrm{w}| < \varepsilon$ and $q = \exp_p \mathrm{v} = \exp_p \mathrm{w}$. But according to 12.9, the exponential map $T_p \to \Sigma$ is injective in the ε-neighborhood of zero — a contradiction. □

C Gauss curvature

Let s be a smooth map from a (possibly unbounded) coordinate rectangle in the (u, v)-plane to a smooth surface Σ. The map s is called *semigeodesic* if for any fixed v the map $u \mapsto s(u, v)$ is a unit-speed geodesic and $s_u \perp s_v$ for any (u, v).

Note that according to the Gauss lemma (15.2), the polar coordinates on Σ are described by a semigeodesic map.

Note that we can choose a unit vector field $\nu = \nu(u, v)$ that is normal to Σ at the point $s(u, v)$. For each pair (u, v), consider an orthonormal frame ν, $\mathrm{u} = s_u$ and $\mathrm{v} = \nu \times \mathrm{u}$. Recall that since the vector $s_v(u, v)$ is tangent to Σ at $s(u, v)$, $s_v \perp \mathrm{u}$ and $s_v \perp \nu$. Therefore we have that $s_v = b\cdot\mathrm{v}$ for some smooth function $(u, v) \mapsto b(u, v)$.

(For a fixed value v_0, the vector field $s_v = b{\cdot}\mathrm{V}$ describes the difference between γ_0 and an *infinitesimally close* geodesic $\gamma_1\colon u \mapsto s(u, v_1)$. The fields with this property are called *Jacobi fields* along γ_0.)

15.3. Proposition. *Suppose $(u, v) \mapsto s(u, v)$ is a semigeodesic map to a smooth surface Σ and ν, U, V and b are as described above. Then*

$$b{\cdot}K + b_{uu} = 0.$$

where $K = K(u, v)$ is the Gauss curvature of Σ at the point $s(u, v)$.

Moreover,

$$\langle \mathrm{U}_u, \mathrm{U}\rangle = \langle \mathrm{U}_u, \mathrm{V}\rangle = \langle \mathrm{U}_v, \mathrm{U}\rangle = 0, \quad \textit{and} \quad \langle \mathrm{U}_v, \mathrm{V}\rangle = b_u.$$

The proof is done by lengthy, but straightforward computations.

Proof. Suppose that $\ell = \ell(u, v)$, $m = m(u, v)$, and $n = n(u, v)$ be the components of the matrix describing the shape operator in the frame U, V; that is,

❶ $$\operatorname{Shape}\mathrm{U} = \ell{\cdot}\mathrm{U} + m{\cdot}\mathrm{V}, \quad \operatorname{Shape}\mathrm{V} = m{\cdot}\mathrm{U} + n{\cdot}\mathrm{V}.$$

Recall that (see Section 9C)

$$K = \ell{\cdot}n - m^2.$$

The following four identities are the key to the proof:

❷ $$\begin{aligned} \mathrm{U}_u &= \ell{\cdot}\nu, & \mathrm{U}_v &= b_u{\cdot}\mathrm{V} + b{\cdot}m{\cdot}\nu, \\ \mathrm{V}_u &= m{\cdot}\nu, & \mathrm{V}_v &= -b_u{\cdot}\mathrm{U} + b{\cdot}n{\cdot}\nu. \end{aligned}$$

Suppose that the identities in ❷ are proved already. Then the proposition can be proved via the following calculations:

$$\begin{aligned} b{\cdot}K &= b{\cdot}(\ell{\cdot}n - m^2) = \\ &= \langle \mathrm{U}_u, \mathrm{V}_v\rangle - \langle \mathrm{U}_v, \mathrm{V}_u\rangle = \\ &= \left(\tfrac{\partial}{\partial v}\langle \mathrm{U}_u, \mathrm{V}\rangle - \cancel{\langle \mathrm{U}_{uv}, \mathrm{V}\rangle}\right) - \left(\tfrac{\partial}{\partial u}\langle \mathrm{U}_v, \mathrm{V}\rangle - \cancel{\langle \mathrm{U}_{uv}, \mathrm{V}\rangle}\right) = \\ &= -b_{uu}. \end{aligned}$$

It remains to prove the four identities in ❷.

Proof of $\mathrm{U}_u = \ell{\cdot}\nu$. Since the frame ν, U and V is orthonormal, this vector identity can be rewritten as the following three scalar identities:

❸ $$\langle \mathrm{U}_u, \mathrm{U}\rangle = 0, \quad \langle \mathrm{U}_u, \mathrm{V}\rangle = 0, \quad \langle \mathrm{U}_u, \nu\rangle = \ell.$$

Since $u\mapsto s(u,v)$ is a geodesic we have that $\mathrm{U}_u=s_{uu}(u,v)\perp \mathrm{T}_{s(u,v)}$. Hence the first two identities follow.

The remaining identity $\langle \mathrm{U}_u,\nu\rangle=\ell$ follows from 9.3 and ❶. Indeed

$$\begin{aligned}\langle \mathrm{U}_u,\nu\rangle &= \langle s_{uu},\nu\rangle = \\ &= \langle \mathrm{Shape}\, s_u, s_u\rangle = \\ &= \langle \mathrm{Shape}\, \mathrm{U}, \mathrm{U}\rangle = \\ &= \ell.\end{aligned}$$

Proof of $\mathrm{U}_v=-b_u\cdot\mathrm{V}+b\cdot m\cdot\nu$. This vector identity can be rewritten as the following three scalar identities:

❹ $$\langle \mathrm{U}_v,\mathrm{U}\rangle=0,\quad \langle \mathrm{U}_v,\mathrm{V}\rangle=b_u,\quad \langle \mathrm{U}_v,\nu\rangle=b\cdot m.$$

Since $\langle \mathrm{U},\mathrm{U}\rangle=1$, we get $0=\tfrac{\partial}{\partial v}\langle \mathrm{U},\mathrm{U}\rangle=2\cdot\langle \mathrm{U}_v,\mathrm{U}\rangle$; hence the first identity in ❹ follows.

Further, since

$$\langle \mathrm{V},\mathrm{V}\rangle=1,\quad 0=\tfrac{\partial}{\partial u}\langle \mathrm{V},\mathrm{V}\rangle=2\cdot\langle \mathrm{V}_u,\mathrm{V}\rangle,\quad \text{and}\quad s_v=b\cdot\mathrm{V},$$

we get

$$\begin{aligned}\langle \mathrm{U}_v,\mathrm{V}\rangle &= \langle s_{vu},\mathrm{V}\rangle = \\ &= \langle \tfrac{\partial}{\partial u}(b\cdot\mathrm{V}),\mathrm{V}\rangle = \\ &= b_u\cdot\langle \mathrm{V},\mathrm{V}\rangle + b\cdot\langle \mathrm{V}_u,\mathrm{V}\rangle = \\ &= b_u;\end{aligned}$$

hence the first identity in ❹ follows.

Finally, applying 9.3, ❶, and $s_v=b\cdot\mathrm{V}$, we get

$$\begin{aligned}\langle \mathrm{U}_v,\nu\rangle &= \langle s_{uv},\nu\rangle = \\ &= \langle \mathrm{Shape}\, s_u, s_v\rangle = \\ &= \langle \mathrm{Shape}\, \mathrm{U}, b\cdot\mathrm{V}\rangle = \\ &= b\cdot m.\end{aligned}$$

Proof of $\mathrm{V}_u=m\cdot\nu$ *and* $\mathrm{V}_v=-b_u\cdot\mathrm{U}+b\cdot n\cdot\nu$. Recall that $\mathrm{V}=\nu\times\mathrm{U}$. Therefore

❺ $$\mathrm{V}_u=\nu_u\times\mathrm{U}+\nu\times\mathrm{U}_u,\quad \mathrm{V}_v=\nu_v\times\mathrm{U}+\nu\times\mathrm{U}_v,$$

Expressions for U_u and U_v in ❷ are proved already. Further

$$\begin{aligned}-\nu_u &= \mathrm{Shape}\, s_u = \\ &= \mathrm{Shape}\, \mathrm{U} = \\ &= \ell\cdot\mathrm{U}+m\cdot\mathrm{V},\end{aligned}\qquad\qquad \begin{aligned}-\nu_v &= \mathrm{Shape}\, s_v = \\ &= b\cdot\mathrm{Shape}\, \mathrm{V} = \\ &= b\cdot(m\cdot\mathrm{U}+n\cdot\mathrm{V}),\end{aligned}$$

It remains to plug these four expressions in ❺. □

A chart $(u,v)\mapsto s(u,v)$ is called *semigeodesic* if the map $(u,v)\mapsto$ $\mapsto s(u,v)$ is semigeodesic. Note that the function $b=b(u,v)$ for a semigeodesic chart s has constant sign. Therefore, by changing the sign of ν, we can (and always will) assume that $b>0$; in other words, $b=|s_v|$.

15.4. Exercise. *Show that any point p in a smooth surface Σ can be covered by a semigeodesic chart.*

15.5. Exercise. *Let p be a point on a smooth surface Σ. Assume that $\exp_p$ is injective in the ball $B=B(0,r_0)_{\mathrm{T}_p}$. Suppose the semigeodesic map $(r,\theta)\mapsto s(r,\theta)$ describes polar coordinates with respect to p and the function $(r,\theta)\mapsto b(r,\theta)$ is as above.*

Prove the following statements:

(a) $b(r,\theta)$ does not change sign for $0\leqslant r<r_0$.
(b) $b(r,\theta)\neq 0$ if $0\leqslant r<r_0$.
(c) Apply (a) and (b) to prove 12.10.

A chart $(u,v)\mapsto s(u,v)$ is called *orthogonal* if $s_u\perp s_v$ for any (u,v). Note that any semigeodesic chart is orthogonal.

A solution of the following exercise is very similar to 15.3.

15.6. Exercise. *Let $(u,v)\mapsto s(u,v)$ be an orthogonal chart of a smooth surface Σ. Denote by $K=K(u,v)$ the Gauss curvature of Σ at $s(u,v)$. Set*

$$\begin{aligned} a&=a(u,v):=|s_u|, & b&=b(u,v):=|s_v|,\\ \mathrm{U}&=\mathrm{U}(u,v):=\tfrac{s_u}{a}, & \mathrm{V}&=\mathrm{V}(u,v):=\tfrac{s_v}{b}.\end{aligned}$$

Let $\nu=\nu(u,v)$ be the unit normal vector at $s(u,v)$.

(a) Show that

$$\begin{aligned} \mathrm{U}_u&=-\tfrac1b\cdot a_v\cdot\mathrm{V}+a\cdot\ell\cdot\nu, & \mathrm{V}_u&=\tfrac1b\cdot a_v\cdot\mathrm{U}+a\cdot m\cdot\nu\\ \mathrm{U}_v&=\tfrac1a\cdot b_u\cdot\mathrm{V}+b\cdot m\cdot\nu, & \mathrm{V}_v&=-\tfrac1a\cdot b_u\cdot\mathrm{U}+b\cdot n\cdot\nu,\end{aligned}$$

where $\ell=\ell(u,v)$, $m=m(u,v)$, and $n=n(u,v)$ be the components of the matrix describing the shape operator in the frame U,V.

(b) Show that

$$K=-\frac{1}{a\cdot b}\cdot\left(\frac{\partial}{\partial u}\left(\frac{b_u}{a}\right)+\frac{\partial}{\partial v}\left(\frac{a_v}{b}\right)\right).$$

15.7. Exercise. *Suppose that $(u,v)\mapsto s(u,v)$ is a* conformal chart*; that is, $s_u\perp s_v$ and $b=|s_u|=|s_v|$ for any (u,v); in this case the function $(u,v)\mapsto b(u,v)$ is called a* conformal factor *of s.*

Use 15.6 to show that the Gauss curvature can be expressed as

$$K = -\frac{\triangle(\ln b)}{b^2},$$

where $\triangle$ *denotes the* laplacian*; that is,* $\triangle = \frac{\partial^2}{\partial u^2} + \frac{\partial^2}{\partial v^2}$ *and* $K = K(u,v)$ *is the Gauss curvature of* Σ *at* $s(u,v)$.

D Rotation of a vector field

Let Σ be a smooth oriented surface and γ a simple closed path in Σ. Suppose that U is a field of unit tangent vectors to Σ defined in a neighborhood of γ. Denote by V the field obtained from U by a counterclockwise rotation by $\frac{\pi}{2}$ of the tangent plane at each point; it could also be defined as $\mathrm{V} = \nu \times \mathrm{U}$. Then the *rotation* of U around γ is defined as the integral

$$\operatorname{rot}_\gamma \mathrm{U} := \int_0^1 \langle \mathrm{U}'(t), \mathrm{V}(t)\rangle \cdot dt.$$

15.8. Lemma. *Suppose that* γ *is a loop based at a point* p *in a smooth oriented surface* Σ *and* U *is a field of tangent unit vectors to* Σ *defined in a neighborhood of* γ*. Then the parallel transport* $\iota_\gamma\colon \mathrm{T}_p \to \mathrm{T}_p$ *is a* clockwise *rotation by the angle* $\operatorname{rot}_\gamma \mathrm{U}$.

In particular rotations of different vector fields around γ *may only differ by a multiple of* $2\cdot\pi$.

Proof. As above, set $\mathrm{V} = \nu \times \mathrm{U}$. Denote by $\mathrm{U}(t)$ and $\mathrm{V}(t)$ the vectors at $\gamma(t)$ of the fields U and V respectively.

Let $t \mapsto \mathrm{X}(t) \in \mathrm{T}_{\gamma(t)}$ be the parallel vector field along γ with $\mathrm{X}(0) = \mathrm{U}(0)$. Set $\mathrm{Y} = \nu \times \mathrm{X}$.

Note that there is a continuous function $t \mapsto \varphi(t)$ such that $\mathrm{U}(t)$ is a counterclockwise rotation of $\mathrm{X}(0)$ by angle $\varphi(t)$. Since $\mathrm{X}(0) = \mathrm{U}(0)$, we can (and will) assume that $\varphi(0) = 0$.

Note that

$$\begin{aligned}
\mathrm{U} &= \cos\varphi\cdot\mathrm{X} + \sin\varphi\cdot\mathrm{Y}\\
\mathrm{V} &= -\sin\varphi\cdot\mathrm{X} + \cos\varphi\cdot\mathrm{Y}
\end{aligned}$$

It follows that

$$\begin{aligned}
\langle \mathrm{U}', \mathrm{V}\rangle &= \varphi'\cdot\Big((\sin\varphi)^2\cdot\langle \mathrm{X},\mathrm{X}\rangle + (\cos\varphi)^2\cdot\langle \mathrm{Y},\mathrm{Y}\rangle\Big) =\\
&= \varphi'.
\end{aligned}$$

Therefore

$$\begin{aligned}\operatorname{rot}_\gamma \mathrm{U} &= \int\limits_0^1 \langle \mathrm{U}'(t), \mathrm{V}(t)\rangle \cdot dt = \\ &= \int\limits_0^1 \varphi'(t)\cdot dt = \\ &= \varphi(1).\end{aligned}$$

Observe that

- ◇ $\iota_\gamma(\mathrm{X}(0)) = \mathrm{X}(1)$,
- ◇ $\mathrm{X}(0) = \mathrm{U}(0) = (1)$,
- ◇ $\mathrm{U}(1)$ is a counterclockwise rotation of $\mathrm{X}(1)$ by the angle $\varphi(1) = = \operatorname{rot}_\gamma \mathrm{U}$,
- ◇ $\mathrm{U}(0) = \mathrm{U}(1)$.

It follows that $\mathrm{X}(1)$ is a *clockwise* rotation of $\mathrm{X}(0)$ by angle $\operatorname{rot}_\gamma \mathrm{U}$, and the result follows. □

The following lemma will play a key role in the proof of the Gauss–Bonnet formula given in the next section.

15.9. Lemma. *Let $(u,v)\mapsto s(u,v)$ be a semigeodesic chart on a smooth surface Σ. Suppose that a simple loop γ bounds a disc Δ that is covered completely by s. Then*

$$\operatorname{rot}_\gamma \mathrm{U} + \iint\limits_\Delta K = 0,$$

where $\mathrm{U} = s_u$ and K denote the Gauss curvature of Σ.

The proof is done by a calculation using the so-called *Green formula* which can be formulated the following way:

Let D be a compact region in the (u,v)-coordinate plane that is bounded by a piecewise smooth simple closed curve α. Suppose that α is oriented in such a way that D lies on its left. Then for any two smooth functions P and Q defined on D we have

$$\iint\limits_D (Q_u - P_v)\cdot du\cdot dv = \int\limits_\alpha (P\cdot du + Q\cdot dv).$$

Note that Green and Gauss–Bonnet formulas are similar — they relate the integral along a disc and its boundary curve. So it shouldn't be surprising that Green helps to prove Gauss–Bonnet.

Proof. Let us write γ in the (u,v)-coordinates: $\gamma(t) = s(u(t), v(t))$. Set $\mathrm{V} = \frac{s_v}{b}$, note that V is a unit vector field orthogonal to U and we can assume that it is the counterclockwise rotation of U by the angle $\frac{\pi}{2}$.

Therefore

$$\operatorname{rot}_\gamma \mathrm{U} = \int\limits_0^1 \langle \mathrm{U}', \mathrm{V}\rangle \cdot dt =$$

by the chain rule

$$= \int\limits_0^1 [\langle \mathrm{U}_u, \mathrm{V}\rangle \cdot u' + \langle \mathrm{U}_v, \mathrm{V}\rangle \cdot v'] \cdot dt =$$

by 15.3,

$$= \int\limits_0^1 b_u \cdot v' \cdot dt =$$

$$= \int\limits_{s^{-1}\circ\gamma} b_u \cdot dv =$$

by the Green formula

$$= \iint\limits_{s^{-1}(R)} b_{uu} \cdot du \cdot dv =$$

Since $\operatorname{jac} s = b$, we get

$$= \iint\limits_R \frac{b_{uu}}{b} =$$

by 15.3, $K = -\frac{b_{uu}}{b}$, so we get

$$= -\iint\limits_R K.$$

$\square$

E Gauss–Bonnet formula: a formal proof

Recall that for a topological disc Δ in a smooth oriented surface Σ we set

$$\mathrm{GB}(\Delta) = \Psi(\partial\Delta) + \iint\limits_{\Delta} K - 2\cdot\pi,$$

where we assume that $\partial\Delta$ is oriented in such a way that Δ lies on its left. So the Gauss–Bonnet formula can be written as $\mathrm{GB}(\Delta) = 0$.

Proof of the Gauss–Bonnet formula (14.1). First assume that Δ is covered by a semigeodesic chart. Note that the following weaker formula follows from 13.5, 15.8, and 15.9:

❶ $$\mathrm{GB}(\Delta) = 2\cdot n\cdot\pi,$$

where $n = n(\Delta)$ is an integer.

By 15.4, any point can be covered by a semigeodesic chart. Therefore applying 14.8 finite number of times, we get that ❶ holds for any disc Δ in Σ.

Assume that Δ lies in a local graph realization $z = f(x, y)$ of Σ. Consider the one-parameter family Σ_t of graphs $z = t\cdot f(x, y)$ and denote by Δ_t the corresponding disc in Σ_t, so $\Delta_1 = \Delta$ and Δ_0 is its projection to the (x, y)-plane. Note that the function $f\colon t \mapsto \mathrm{GB}(\Delta_t)$ is continuous. From above, $f(t)$ is a multiple of $2\cdot\pi$ for any t. It follows that f is a constant function. In particular

$$\begin{aligned}\mathrm{GB}(\Delta) &= \mathrm{GB}(\Delta_0) = \\ &= 0,\end{aligned}$$

where the last equality follows from 5.5.

We proved that

❷ $$\mathrm{GB}(\Delta) = 0$$

if Δ lies in a graph $z = f(x, y)$ for some (x, y, z)-coordinate system. Since any point of Σ has a neighborhood that can be covered by such a graph, applying Lemma 14.8 as above we get that ❷ holds for any disc Δ in Σ. □

F Rauch comparison

The following proposition is a special case of the so-called *Rauch comparison theorem.*

15.10. Proposition. *Suppose that p is a point on a smooth surface Σ and $r \leqslant \operatorname{inj}(p)$. Given a curve $\tilde\gamma$ in the r-neighborhood of 0 in T_p, set*

$$\gamma = \exp_p \circ \tilde\gamma \quad \textit{or, equivalently} \quad \log_p \circ \gamma = \tilde\gamma;$$

note that γ is a curve in Σ.

(a) If Σ has nonnegative Gauss curvature, then the exponential map $\exp_p$ is length nonexpanding in the r-neighborhood of 0 in T_p; that is,

$$\operatorname{length}\gamma \leqslant \operatorname{length}\tilde\gamma$$

for any curve $\tilde\gamma$ in the open ball $B(0,r)_{\mathrm{T}_p}$.

(b) If Σ has nonpositive Gauss curvature, then the logarithmic map $\log_p$ is length nonexpanding in the r-neighborhood of p in Σ; that is,

$$\operatorname{length}\gamma \geqslant \operatorname{length}\tilde\gamma$$

for any curve γ in the open ball $B(p,r)_\Sigma$.

Proof. Suppose $(r(t),\theta(t))$ are the polar coordinates of $\tilde\gamma(t)$. Note that $\gamma(t) = s(r(t),\theta(t))$; that is, $(r(t),\theta(t))$ are the polar coordinates of $\gamma(t)$ with respect to p on Σ.

Set $b(r,\theta) := |s_\theta|$. By 15.3

$$b_{rr} = -K\cdot b.$$

If $K \geqslant 0$, then $r \mapsto b(r,\theta)$ is concave and if $K \leqslant 0$, then $r \mapsto b(r,\theta)$ is convex for any fixed θ. Note that $b(0,\theta) = 0$ and by 12.8, $b_\theta(0,\theta) = 1$. Therefore

❶
$$\begin{aligned} b(r,\theta) &\leqslant r \quad \text{if} \quad K \geqslant 0 \quad \text{and} \\ b(r,\theta) &\geqslant r \quad \text{if} \quad K \leqslant 0. \end{aligned}$$

Without loss of generality we may assume that $\tilde\gamma\colon [a,b] \to \mathrm{T}_p$ is parametrized by arc-length; in particular it is a Lipschitz curve. Note that

$$\operatorname{length}\tilde\gamma = \int\limits_a^b \sqrt{r'(t)^2 + r(t)^2\cdot\theta'(t)^2}.$$

Applying 15.2, we get

$$\operatorname{length}\gamma = \int\limits_a^b \sqrt{r'(t)^2 + b(r(t),\theta(t))^2\cdot\theta'(t)^2}.$$

Both statements follow from ❶. □

G Intrinsic isometries

Let Σ and Σ^* be two smooth regular surfaces in the Euclidean space. A map $f\colon \Sigma \to \Sigma^*$ is called *length-preserving* if for any curve γ in Σ the curve $\gamma^* = f\circ\gamma$ in Σ^* has the same length. If in addition f is smooth and bijective, then it is called an *intrinsic isometry.*

15.11. Exercise. *Suppose that the Gauss curvature of a smooth surface Σ vanishes. Show that Σ is* locally flat*; that is, a neighborhood of any point in Σ admits an intrinsic isometry to an open domain in the Euclidean plane.*

15.12. Exercise. *Suppose that a smooth surface Σ has unit Gauss curvature at every point. Show that a neighborhood of any point in Σ admits an intrinsic isometry to an open domain in the unit sphere.*

A simple example of intrinsic isometry can be obtained by wrapping a plane into a cylinder with the map $f : \mathbb{R}^2 \to \mathbb{R}^3$ given by

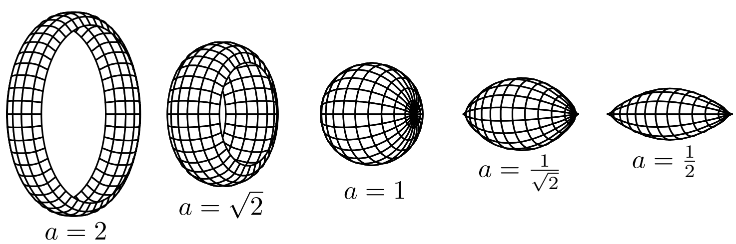
$$f(x,y) = (\cos x, \sin x, y).$$

The following exercise produces a more interesting example.

15.13. Exercise. *Let $a > 0$ be a constant and set*

$$\varepsilon = \arcsin\left(\min\{1, \frac{1}{a}\}\right) \in (0,\pi].$$

Construct $\gamma : (-\varepsilon,\varepsilon) \to \mathbb{R}^2$, $\gamma(t) = (x(t), y(t))$ a smooth unit speed curve with $y(t) = a\cos t$, and let Σ_a be the surface of revolution of γ around the x-axis. Show that the surface Σ_a has unit Gauss curvature at each point.



Use 15.12 to conclude that any small round disc Δ in $\mathbb{S}^2$ admits a smooth length-preserving deformation; that is, there is one-parameter family of surfaces with boundary Δ_t, such that $\Delta_0 = \Delta$ and Δ_t is not congruent to Δ_0 for $t \neq 0$.[1]

[1] In fact any disc in $\mathbb{S}^2$ admits a smooth length preserving deformation. However if

H The remarkable theorem

15.14. Theorem. *Suppose $f\colon\Sigma\to\Sigma^*$ is an intrinsic isometry between two smooth regular surfaces in the Euclidean space; $p\in\Sigma$ and $p^* = f(p)\in\Sigma$. Then*

$$K(p)_\Sigma = K(p^*)_{\Sigma^*};$$

that is, the Gauss curvature of Σ at p is the same as the Gauss curvature of Σ^* at p^*.

This theorem was proved by Carl Friedrich Gauss [34] who called it the *Remarkable theorem* (*Theorema Egregium*). The theorem is indeed remarkable because the Gauss curvature is defined as a product of principal curvatures which might be different at these points; however, according to the theorem, their product cannot change. In other words, the Gaussian curvature is an *intrinsic invariant.*

In fact the Gauss curvature of the surface at the given point can be found *intrinsically*, by measuring the lengths of curves in the surface. For example, the Gauss curvature $K(p)$ appears in the following formula for the circumference $c(r)$ of a geodesic circle centered at p in a surface:

$$c(r) = 2{\cdot}\pi{\cdot}r - \tfrac{\pi}{3}{\cdot}K(p){\cdot}r^3 + o(r^3).$$

The theorem implies for example that there is no smooth length-preserving map that sends an open region in the unit sphere to the plane.[2] This follows since the Gauss curvature of the plane is zero and the unit sphere has Gauss curvature 1. In other words, there is no map of any region on Earth without distortion.

Proof. Choose a chart $(u,v)\mapsto s(u,v)$ on Σ and set $s^* = f\circ s$. Note that s^* is a chart of Σ^* and

$$\langle s_u, s_u\rangle = \langle s^*_u, s^*_u\rangle, \qquad \langle s_u, s_v\rangle = \langle s^*_u, s^*_v\rangle, \qquad \langle s_v, s_v\rangle = \langle s^*_v, s^*_v\rangle$$

at any (u,v). Indeed the first and the third identity hold since otherwise f would not preserve the lengths of the coordinate lines $\gamma\colon t\mapsto s(t,v)$ or $\gamma\colon t\mapsto s(u,t)$. Taking this into account, the second identity holds since otherwise f would not preserve the lengths of the coordinate lines $\gamma\colon t\mapsto s(t,c-t)$ for some constant c.

It follows that if s is a semigeodesic chart, then so is s^*. It remains to apply 15.3 and 15.4. $\square$

---

the disc is larger than half-sphere, then the proof requires more; it can be obtained as a corollary of two deep results of Alexandr Alexandrov: the gluing theorem and the theorem on the existence of a convex surface with an abstractly given metric [69, p. 44].

[2]There are plenty of non-smooth length-preserving maps from the sphere to the plane; see [67] and the references therein.

Chapter 16

Comparison theorems

This chapter is based on material in the book by Stephanie Alexander, Vitali Kapovitch and the first author [6].

A Triangles and hinges

Recall that a shortest path between points x and y in a surface Σ is denoted by $[x,y]$ or $[x,y]_\Sigma$, and $|x-y|_\Sigma$ denotes the *intrinsic distance* from x to y in Σ.

A *geodesic triangle* in a surface Σ is defined as a triple of points $x,y,z\in$ $\in\Sigma$ with a choice of minimizing geodesics $[x,y]_\Sigma$, $[y,z]_\Sigma$ and $[z,x]_\Sigma$. The points x,y,z are called the *vertices* of the geodesic triangle, the minimizing geodesics $[x,y]$, $[y,z]$ and $[z,x]$ are called its *sides*; the triangle itself is denoted by $[xyz]$, or by $[xyz]_\Sigma$, if we need to emphasize that it lies in the surface Σ.

A triangle $[\tilde x\tilde y\tilde z]$ in the plane $\mathbb{R}^2$ is called a *model triangle* of the triangle $[xyz]$ if its corresponding sides are equal; that is,

$$|\tilde x-\tilde y|_{\mathbb{R}^2}=|x-y|_\Sigma,\quad |\tilde y-\tilde z|_{\mathbb{R}^2}=|y-z|_\Sigma,\quad |\tilde z-\tilde x|_{\mathbb{R}^2}=|z-x|_\Sigma.$$

In this case we write $[\tilde x\tilde y\tilde z]=\tilde\triangle xyz$.

A pair of minimizing geodesics $[x,y]$ and $[x,z]$ starting from one point x is called a *hinge* and is denoted by $[x\,{}^y_z]$. The angle between these geodesics at x is denoted by $\measuredangle[x\,{}^y_z]$. The corresponding angle $\measuredangle[\tilde x\,{}^{\tilde y}_{\tilde z}]$ in a model triangle $[\tilde x\tilde y\tilde z]=\tilde\triangle xyz$ is denoted by $\tilde\measuredangle(x\,{}^y_z)$; it is called *model* or *comparison angle* of the triangle $[xyz]$ at x.

By side-side-side congruence condition, the model triangle $[\tilde x\tilde y\tilde z]$ is uniquely defined up to congruence. Therefore the model angle $\tilde\theta=\tilde\measuredangle(x\,{}^y_z)$

is uniquely defined as well. In fact by cosine rule, we get

$$\cos\tilde\theta = \frac{a^2+b^2-c^2}{2\cdot a\cdot b},$$

where $a = |x-y|_\Sigma$, $b = |x-z|_\Sigma$, and $c = |y-z|_\Sigma$.

16.1. Exercise. *Let $[x_n y_n z_n]$ be a sequence of triangles in a smooth surface Σ. Set $a_n = |x_n - y_n|_\Sigma$, $b_n = |x_n - z_n|_\Sigma$, and $c_n = |y_n - z_n|_\Sigma$, and $\tilde\theta_n = \tilde\measuredangle(x_n\,{}^{y_n}_{z_n})$. Suppose that the sequences a_n and b_n are bounded away from zero; that is, $a_n > \varepsilon$ and $b_n > \varepsilon$ for a fixed $\varepsilon > 0$ and any n. Show that*

$$(a_n + b_n - c_n) \to 0 \qquad \Longleftrightarrow \qquad \tilde\theta_n \to \pi$$

as $n \to \infty$

B Formulations

Part *(b)* of the following theorem is called the *Toponogov comparison theorem* and sometimes the *Alexandrov comparison theorem*; it was proved by Paolo Pizzetti [68] and rediscovered by Alexandr Alexandrov [2]; generalizations were obtained by Victor Toponogov [79], Mikhael Gromov, Yuri Burago and Grigory Perelman [14].

Part *(a)* is called the *Cartan–Hadamard theorem*; it was proved by Hans von Mangoldt [57] and generalized by Elie Cartan [16], Jacques Hadamard [38], Herbert Busemann [15], Willi Rinow [70], Mikhael Gromov [36, p. 119], Stephanie Alexander and Richard Bishop [4].

Recall that a surface Σ is called *simply-connected* if any closed simple curve in Σ bounds a disc.

16.2. Comparison theorems. *Let Σ be a proper smooth regular surface.*

(a) If Σ is simply-connected and has nonpositive Gauss curvature, then

$$\measuredangle[x\,{}^y_z] \leqslant \tilde\measuredangle(x\,{}^y_z)$$

for any geodesic triangle $[xyz]$.

(b) If Σ has nonnegative Gauss curvature, then

$$\measuredangle[x\,{}^y_z] \geqslant \tilde\measuredangle(x\,{}^y_z)$$

for any geodesic triangle $[xyz]$.

The proof of part *(a)* will be given at the end of Section 16D. The proof of part *(b)* will be finished in Section 16E.

Let us show that the statement *(a)* does not hold without assuming that Σ is simply-connected. Consider the hyperboloid

$$\{\,(x,y,z)\in\mathbb{R}^3\;:\;x^2+y^2-z^2=1\,\};$$

it has negative Gauss curvature, but it is not simply-connected. The points $(1,0,0)$, $(-\frac12,\frac{\sqrt3}{2},0)$, $(-\frac12,-\frac{\sqrt3}{2},0)$ form a triangle with all angles equal to π, but all comparison angles equal to $\frac{\pi}{3}$.

Let us discuss the relationship between the Gauss–Bonnet formula and the comparison theorems. Suppose that a disc Δ is bounded by a geodesic triangle $[xyz]$ with internal angles α, β and γ. Then Gauss–Bonnet implies that

$$\alpha+\beta+\gamma-\pi=\iint\limits_{\Delta}K;$$

in particular both sides of the equation have the same sign. It follows that

- $\diamond$ if $K_\Sigma\geqslant0$ then $\alpha+\beta+\gamma\geqslant\pi$, and
- $\diamond$ if $K_\Sigma\leqslant0$ then $\alpha+\beta+\gamma\leqslant\pi$.

Now set $\hat\alpha=\measuredangle[x\,{}^y_z]$, $\hat\beta=\measuredangle[y\,{}^z_x]$, and $\hat\gamma=\measuredangle[z\,{}^x_y]$. Since the angles of any plane triangle add up π, from the comparison theorems, we get that

- $\diamond$ if $K_\Sigma\geqslant0$ then $\hat\alpha+\hat\beta+\hat\gamma\geqslant\pi$, and
- $\diamond$ if $K_\Sigma\leqslant0$ then $\hat\alpha+\hat\beta+\hat\gamma\leqslant\pi$.

The triples of angles $\{\alpha,\beta,\gamma\}$, $\{\hat\alpha,\hat\beta,\hat\gamma\}$ are related by the identities:

$$\hat\alpha=\min\{\,\alpha,2\cdot\pi-\alpha\,\},\quad\hat\beta=\min\{\,\beta,2\cdot\pi-\beta\,\},\quad\hat\gamma=\min\{\,\gamma,2\cdot\pi-\gamma\,\}.$$

One can use them to see that despite the Gauss–Bonnet formula and the comparison theorems are closely related, this relationship is not straightforward.

For example, suppose $K\geqslant0$. Then the Gauss–Bonnet formula does not forbid the internal angles α, β, and γ to be simultaneously close to $2\cdot\pi$. But if α, β, and γ are each close to $2\cdot\pi$, then $\hat\alpha$, $\hat\beta$, and $\hat\gamma$ are close to 0. The latter is impossible by the comparison theorem.

The following exercise is a lemma from the note by Arseniy Akopyan and the first author [1].

16.3. Exercise. *Let p and q be points on a closed convex surface Σ that lie at maximal intrinsic distance from each other; that is, $|p-q|_\Sigma\geqslant|x-y|_\Sigma$ for any $x,y\in\Sigma$. Show that*

$$\measuredangle[x\,{}^p_q]\geqslant\tfrac{\pi}{3}$$

for any point $x\in\Sigma\backslash\{p,q\}$.

16.4. Exercise. *Let Σ be a closed (or open) regular surface with nonnegative Gauss curvature. Show that*

$$\tilde{\measuredangle}(p\,{}^x_y) + \tilde{\measuredangle}(p\,{}^y_z) + \tilde{\measuredangle}(p\,{}^z_x) \leqslant 2\cdot\pi.$$

for any four distinct points p, x, y, z in Σ.

C Local comparisons

The following local version of the comparison theorem follows from the Rauch comparison (15.10). It will be used in the proof of the global version (16.2).

16.5. Theorem. *The comparison theorem (16.2) holds in a small neighborhood of any point.*

Moreover, let Σ be a smooth regular surface without boundary. Then for any $p \in \Sigma$ there is $r > 0$ such that if $|p - x|_\Sigma < r$, then $\operatorname{inj}(x)_\Sigma > r$ and the following statements hold:

(a) If Σ has nonpositive Gauss curvature, then

$$\measuredangle[x\,{}^y_z] \leqslant \tilde{\measuredangle}(x\,{}^y_z)$$

for any geodesic triangle $[xyz]$ in $B(p, \frac{r}{4})_\Sigma$.

(b) If Σ has nonnegative Gauss curvature, then

$$\measuredangle[x\,{}^y_z] \geqslant \tilde{\measuredangle}(x\,{}^y_z)$$

for any geodesic triangle $[xyz]$ in $B(p, \frac{r}{4})_\Sigma$.

Proof. The existence of $r > 0$ follows from 12.9. Let $[xyz]$ be a geodesic triangle in $B(p, \frac{r}{4})$.

Note that $y = \exp_x \mathrm{v}$ and $z = \exp_x \mathrm{w}$ for two vectors $\mathrm{v}, \mathrm{w} \in \mathrm{T}_x$ with

$$\begin{aligned}\measuredangle[0\,{}^{\mathrm{v}}_{\mathrm{w}}]_{\mathrm{T}_x} &= \measuredangle[x\,{}^y_z]_\Sigma,\\ |\mathrm{v}|_{\mathrm{T}_x} &= |x - y|_\Sigma,\\ |\mathrm{w}|_{\mathrm{T}_x} &= |x - z|_\Sigma;\end{aligned}$$

in particular, $|\mathrm{v}|, |\mathrm{w}| < \frac{r}{2}$.

(b). Consider the line segment $\tilde\gamma$ joining v to w in the tangent plane T_x and set $\gamma = \exp_x \circ \tilde\gamma$. By Rauch comparison (15.10a), we have

$$\operatorname{length}\gamma \leqslant \operatorname{length}\tilde\gamma.$$

Since $|\mathrm{v} - \mathrm{w}|_{\mathrm{T}_x} = \operatorname{length}\tilde\gamma$ and $|y - z|_\Sigma \leqslant \operatorname{length}\gamma$, we get

$$|\mathrm{v} - \mathrm{w}|_{\mathrm{T}_x} \geqslant |y - z|_\Sigma.$$

By the angle monotonicity (0.11), we obtain

$$\tilde{\measuredangle}(x\,{}^{z}_{y}) \leqslant \measuredangle[0\,{}^{\mathrm{v}}_{\mathrm{w}}]_{\mathrm{T}_x},$$

whence the result.

(a). Consider a minimizing geodesic γ joining y to z in Σ. Since $|x - - y|_\Sigma, |x - z|_\Sigma < \frac{r}{2}$, the triangle inequality implies that γ lies in the r-neighborhood of x. In particular, $\log_x \circ\gamma$ is defined, and the curve $\tilde{\gamma} = \log_x \circ\gamma$ lies in a r-neighborhood of zero in T_x that corresponds to γ. Note that $\tilde{\gamma}$ connects v to w in T_x.

By Rauch comparison (15.10b), we have

$$\operatorname{length}\gamma \geqslant \operatorname{length}\tilde{\gamma}.$$

Since $|\mathrm{v} - \mathrm{w}|_{\mathrm{T}_x} \leqslant \operatorname{length}\tilde{\gamma}$ and $|y - z|_\Sigma = \operatorname{length}\gamma$, we get

$$|\mathrm{v} - \mathrm{w}|_{\mathrm{T}_x} \geqslant |y - z|_\Sigma.$$

By angle monotonicity (0.11), we get

$$\tilde{\measuredangle}(x\,{}^{z}_{y}) \geqslant \measuredangle[0\,{}^{\mathrm{v}}_{\mathrm{w}}]_{\mathrm{T}_x}.$$

whence the result. □

D Nonpositive curvature

Proof of 16.2a. Sine Σ is simply-connected, 14.7 implies that

$$\operatorname{inj}(p)_\Sigma = \infty$$

for any $p \in \Sigma$. Therefore *(a)* implies 16.2a. □

E Nonnegative curvature

We will prove 16.2b, first assuming that Σ is compact. The general case requires only minor modifications; they are indicated in Exercise 16.8 at the end of the section. The proof is taken from [6] and it is very close to the proof given by Urs Lang and Viktor Schroeder [53].

Proof of 16.2b in the compact case. Assume Σ is compact. From the local theorem (16.5), we get that there is $\varepsilon > 0$ such that the inequality

$$\measuredangle[x\,{}^{p}_{q}] \geqslant \tilde{\measuredangle}(x\,{}^{p}_{q}).$$

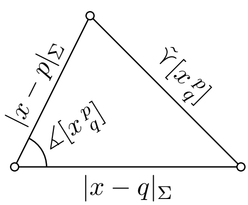
holds for any hinge $[x\,{}^p_q]$ with $|x-p|_\Sigma+|x-q|_\Sigma<\varepsilon$. The following lemma states that in this case the same holds true for any hinge $[x\,{}^p_q]$ such that $|x-p|_\Sigma+|x-q|_\Sigma<\frac{3}{2}\cdot\varepsilon$. Applying the key lemma (16.6) a few times we get that the comparison holds for an arbitrary hinge, which proves 16.2*b*. $\square$

16.6. Key lemma. *Let Σ be a proper smooth regular surface. Assume that the comparison*

$$❶ \qquad \measuredangle[x\,{}^y_z]\geqslant\tilde\measuredangle(x\,{}^y_z)$$

holds for any hinge $[x\,{}^y_z]$ with $|x-y|_\Sigma+|x-z|_\Sigma<\frac{2}{3}\cdot\ell$. Then the comparison ❶ holds for any hinge $[x\,{}^y_z]$ with $|x-y|_\Sigma+|x-z|_\Sigma<\ell$.

Given a hinge $[x\,{}^p_q]$ consider a triangle in the plane with angle $\measuredangle[x\,{}^p_q]$ and two adjacent sides $|x-p|_\Sigma$ and $|x-q|_\Sigma$. Let us denote by $\tilde\Upsilon[x\,{}^p_q]$ the third side of this triangle; let us call it the *model side* of the hinge.



The following comutational exercise plays a role in the proof of lemma.

16.7. Exercise.

(a) *Let u be a unit vector in the plane. Show that for any vector w, the function*

$$t\mapsto t+|\mathrm{w}|-|\mathrm{w}+t\cdot\mathrm{u}|$$

is concave.

(b) *Suppose that hinges $[x\,{}^p_q]$ and $[x\,{}^p_y]$ have a common side $[x,p]$ and $[x,y]\subset[x,q]$. Use part (a) to show that*

$$\frac{|x-p|+|x-q|-\tilde\Upsilon[x\,{}^p_q]}{|x-q|}\leqslant\frac{|x-p|+|x-y|-\tilde\Upsilon[x\,{}^p_y]}{|x-y|}.$$

Proof. Note that by angle monotonicity (0.11),

$$\measuredangle[x\,{}^p_q]\geqslant\tilde\measuredangle(x\,{}^p_q)\quad\iff\quad\tilde\Upsilon[x\,{}^p_q]\geqslant|p-q|_\Sigma.$$

Therefore it is sufficient to prove that

$$❷ \qquad \tilde\Upsilon[x\,{}^p_q]\geqslant|p-q|_\Sigma.$$

for a given hinge $[x\,{}^p_q]$ with $|x-p|_\Sigma+|x-q|_\Sigma<\ell$.

Let us describe a construction that produces a new hinge $[x'\,{}^p_q]$ for a given hinge $[x\,{}^p_q]$ such that

$$\tfrac{2}{3}\cdot\ell \leqslant |p-x|_\Sigma + |x-q|_\Sigma < \ell.$$

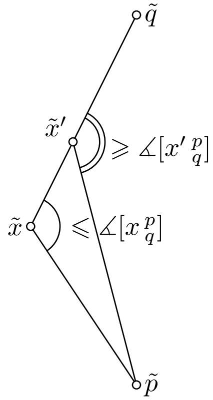
Assume $|x-q|_\Sigma \geqslant |x-p|_\Sigma$, otherwise switch the roles of p and q in the following construction. Take $x' \in [x,q]$ such that

❸ $$|p-x|_\Sigma + 3\cdot|x-x'|_\Sigma = \tfrac{2}{3}\cdot\ell$$

Choose a geodesic $[x',p]$ and consider the hinge $[x'\,{}^p_q]$ formed by $[x',p]$ and $[x',q] \subset [x,q]$.

By te triangle inequality we have

❹ $$|p-x|_\Sigma + |x-q|_\Sigma \geqslant |p-x'|_\Sigma + |x'-q|_\Sigma.$$



Let us show that

❺ $$\tilde\Upsilon[x\,{}^p_q] \geqslant \tilde\Upsilon[x'\,{}^p_q]$$

By ❸, we have that

$$|p-x|_\Sigma + |x-x'|_\Sigma < \tfrac{2}{3}\cdot\ell,$$
$$|p-x'|_\Sigma + |x'-x|_\Sigma < \tfrac{2}{3}\cdot\ell.$$

By the assumption, we get that

❻ $$\measuredangle[x\,{}^{p}_{x'}] \geqslant \tilde\measuredangle(x\,{}^{p}_{x'}),$$
$$\measuredangle[x'\,{}^{p}_{x}] \geqslant \tilde\measuredangle(x'\,{}^{p}_{x}).$$

Consider the model triangle $[\tilde x\tilde x'\tilde p] = \tilde\triangle x x' p$. Take $\tilde q$ on the extension of $[\tilde x, \tilde x']$ beyond x' such that $|\tilde x - \tilde q|_\Sigma = |x-q|_\Sigma$ and therefore $|\tilde x' - \tilde q|_\Sigma = |x'-q|_\Sigma$.

From ❻, we get

$$\measuredangle[x\,{}^p_q] = \measuredangle[x\,{}^{p}_{x'}] \geqslant \tilde\measuredangle(x\,{}^{p}_{x'}).$$

Therefore

$$\tilde\Upsilon[x\,{}^q_p] \geqslant |\tilde p - \tilde q|_{\mathbb{R}^2}.$$

Further, since $\measuredangle[x'\,{}^p_x] + \measuredangle[x'\,{}^p_q] = \pi$, ❻ implies

$$\pi - \tilde\measuredangle(x'\,{}^p_x) \geqslant \pi - \measuredangle[x'\,{}^p_x] \geqslant \measuredangle[x'\,{}^p_q].$$

Therefore

$$|\tilde p - \tilde q|_{\mathbb{R}^2} \geqslant \tilde\Upsilon[x'\,{}^q_p]$$

and ❺ follows.

Set $x_0 = x$; apply inductively the above construction to get a sequence of hinges $[x_n\,{}^p_q]$ with $x_{n+1} = x'_n$. By ❺ and ❹, both sequences

$$s_n = \tilde{\curlyvee}[x_n\,{}^p_q] \quad\text{and}\quad r_n = |p - x_n|_\Sigma + |x_n - q|_\Sigma$$

are nonincreasing.

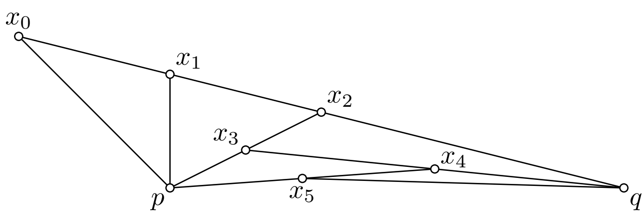
The sequence (x_n) might terminate at some n only if $r_n < \tfrac{2}{3}\cdot\ell$. In this case, by the assumptions of the lemma,

$$s_n = \tilde{\curlyvee}[x_n\,{}^p_q] \geqslant |p - q|_\Sigma.$$

Since the sequence (s_n) is nonincreasing, we get

$$\tilde{\curlyvee}[x\,{}^p_q] = s_0 \geqslant s_n \geqslant |p - q|_\Sigma;$$

whence ❷ follows.



Assume the sequence (x_n) does not terminate. By ❸, we have

❼ $$|x_n - x_{n-1}|_\Sigma \geqslant \tfrac{1}{100}\cdot\ell.$$

Evidently $|x_n - p|, |x_n - q| < \ell$ for any n. Applying 16.7b for the hinge $[x_n\,{}^p_q]$ and $[x_n\,{}^{p}_{x_{n+1}}]$ (or $[x_n\,{}^{x_{n+1}}_{q}]$, depending on n) we get that

$$r_n - s_n \leqslant 100\cdot(r_n - r_{n+1})$$

The sequences (r_n) and (s_n) are nonincreasing and nonnegative. Therefore they have to convege. In particular $r_n - r_{n+1} \to 0$ as $n \to \infty$. It follows that

$$\lim_{n\to\infty} s_n = \lim_{n\to\infty} r_n.$$

By the triangle inequality, $r_n \geqslant |p-q|_\Sigma$ for any n. Since s_n is nonincreasing, we get

$$\tilde{\curlyvee}[x\,{}^p_q] = s_0 \geqslant \lim_{n\to\infty} s_n = \lim_{n\to\infty} r_n \geqslant |p - q|_\Sigma.$$

which proves ❷ in the remaining case. □

16.8. Exercise. *Let Σ be an open surface with nonnegative Gauss curvature. Given $p \in \Sigma$, denote by R_p (the* comparison radius *at p) the maximal value (possibly ∞) such that the comparison*

$$\measuredangle[x\,{}^{p}_{y}] \geqslant \tilde{\measuredangle}(x\,{}^{p}_{y})$$

holds for any hinge $[x\,{}^{p}_{y}]$ with $|p-x|_\Sigma + |x-y|_\Sigma < R_p$.

(a) *Show that for any compact subset $K \subset \Sigma$, there is $\varepsilon > 0$ such that $R_p > \varepsilon$ for any $p \in K$.*

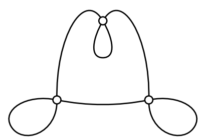
(b) *Use part (a) to show that there is a point $p \in \Sigma$ such that*

$$R_q > (1 - \tfrac{1}{100}) \cdot R_p,$$

for any $q \in B(p, 100 \cdot R_p)_\Sigma$.

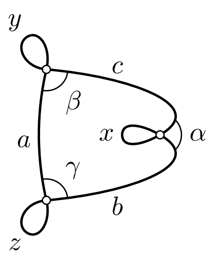
(c) *Use (b) to extend the proof of 16.2b (page 159) to open surfaces. (Show that $R_p = \infty$ for any $p \in \Sigma$.)*

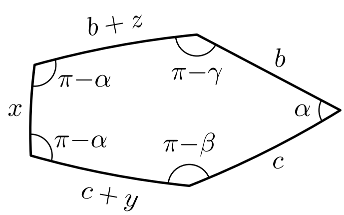
16.9. Advanced exercise. *Let Σ be a closed smooth regular surface with nonnegative Gauss curvature. The following sketch shows that a closed geodesic γ on Σ cannot have self-intersections as shown on the diagram; in other words, γ cannot cut Σ into 3 monogons, one quadrangle, and one pentagon.*



Make a complete proof from it.

Arguing by contradiction, suppose that such geodesic exists; assume that arcs and angles are labeled as on the left diagram.





(a) *Apply Gauss–Bonnet formula to show that*

$$2 \cdot \alpha < \beta + \gamma$$

and

$$2 \cdot \beta + 2 \cdot \gamma < \pi + \alpha.$$

Conclude that $\alpha < \frac{\pi}{3}$.

(b) *Consider the part of the geodesic γ without the arc a. It cuts from Σ a pentagon Δ with sides and angles as shown on the diagram. Show that there is a plane pentagon with convex sides of the same length and angles at most as big as the corresponding angles of Δ.*

(c) *Arrive to a contradiction using (a) and (b).*

F Alexandrov's lemma

A reformulation of the following lemma in plane geometry will be used in the next section to produce a few equivalent reformulations of the comparison theorems.

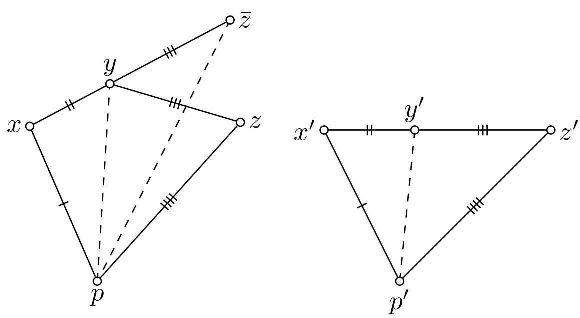
16.10. Lemma. *Assume $[pxyz]$ and $[p'x'y'z']$ be two quadrilaterals in the plane with equal corresponding sides. Assume that the sides $[x',y']$ and $[y',z']$ extend each other; that is, y' lies on the line segment $[x',z']$. Then the following expressions have the same signs:*

(i) $|p-y|-|p'-y'|$;

(ii) $\measuredangle[x\,{}^{p}_{y}]-\measuredangle[x'\,{}^{p'}_{y'}]$;

(iii) $\pi-\measuredangle[y\,{}^{p}_{x}]-\measuredangle[y\,{}^{p}_{z}]$;

Proof. Take a point $\bar z$ on the extension of $[x,y]$ beyond y so that $|y-\bar z|=|y-z|$ (and therefore $|x-\bar z|=|x'-z'|$).



From angle monotonicity (0.11), the following expressions have the same sign:

(i) $|p-y|-|p'-y'|$;

(ii) $\measuredangle[x\,{}^{y}_{p}]-\measuredangle[x'\,{}^{y'}_{p'}]=\measuredangle[x\,{}^{\bar z}_{p}]-\measuredangle[x'\,{}^{z'}_{p'}]$;

(iii) $|p-\bar z|-|p'-z'|=|p-\bar z|-|p-z|$;

(iv) $\measuredangle[y\,{}^{\bar z}_{p}]-\measuredangle[y\,{}^{z}_{p}]$;

The statement follows since

$$\measuredangle[y'\,{}^{z'}_{p'}]+\measuredangle[y'\,{}^{x'}_{p'}]=\pi$$

and

$$\measuredangle[y\,{}^{\bar z}_{p}] + \measuredangle[y\,{}^{x}_{p}] = \pi. \qquad \square$$

G Reformulations

In this section we formulate conditions equivalent to the conclusion of the comparison theorem (16.2).

For any triangle $[xyz]_\Sigma$ in a surface Σ, and $[\tilde x\tilde y\tilde z]_{\mathbb{R}^2}$ a comparison triangle, there is a map $f : [xyz] \to [\tilde x\tilde y\tilde z]$ that isometrically sends the geodesics $[x, y]$, $[y, z]$, $[z, x]$ to the segments $[\tilde x, \tilde y]$, $[\tilde y, \tilde z]$, $[\tilde z, \tilde x]$, respectively. The triangle $[xyz]$ is called *fat* (*thin*) if for any two points p and q in $[xyz]$, $|p - q|_\Sigma \geqslant |f(p) - f(q)|_{\mathbb{R}^2}$ (or, respectively, $|p - q|_\Sigma \leqslant |f(p) - f(q)|_{\mathbb{R}^2}$).

16.11. Proposition. *Let Σ be a proper smooth regular surface. Then the following three conditions are equivalent:*

(a) For any geodesic triangle $[xyz]$ in Σ we have

$$\measuredangle[x\,{}^{y}_{z}] \geqslant \tilde\measuredangle(x\,{}^{y}_{z}).$$

(b) For any geodesic triangle $[pxz]$ in Σ and y on the side $[x, z]$ we have

$$\tilde\measuredangle(x\,{}^{p}_{y}) \geqslant \tilde\measuredangle(x\,{}^{p}_{z}).$$

(c) Any geodesic triangle in Σ is fat.

Similarly, following three conditions are equivalent:

(A) For any geodesic triangle $[xyz]$ in Σ we have

$$\measuredangle[x\,{}^{y}_{z}] \leqslant \tilde\measuredangle(x\,{}^{y}_{z}).$$

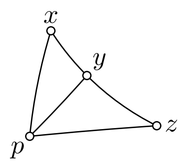
(B) For any geodesic triangle $[pxz]$ in Σ and y on the side $[x, z]$ we have

$$\tilde\measuredangle(x\,{}^{p}_{y}) \leqslant \tilde\measuredangle(x\,{}^{p}_{z}).$$

(C) Any geodesic triangle in Σ is thin.

In the proof we will use the following translation of the Alexandrov lemma to the language of comparison triangles and angles.

16.12. Reformulation of Alexandrov lemma. *Let $[pxz]$ be a triangle in a surface Σ and y a point on the side $[x, z]$. Consider its model triangle $[\tilde p\tilde x\tilde z] = \tilde\triangle pxz$ and let $\tilde y$ be the corresponding point on the side $[\tilde x, \tilde z]$. Then the following expressions have the same signs:*



(i) $|p - y|_\Sigma - |\tilde p - \tilde y|_{\mathbb{R}^2}$;

(ii) $\tilde\measuredangle(x\,{}^p_y) - \tilde\measuredangle(x\,{}^p_z)$;
(iii) $\pi - \tilde\measuredangle(y\,{}^p_x) - \tilde\measuredangle(y\,{}^p_z)$;

Proof or 16.11. We will prove the implications *(a)⇒(b)⇒(c)⇒(a)*. The implications *(A)⇒(B)⇒(C)⇒(A)* can be proved in the same way, but with the reverse inequalities.

(a)⇒(b). Note that $\measuredangle[y\,{}^p_x] + \measuredangle[y\,{}^p_z] = \pi$. By *(a)*,

$$\tilde\measuredangle(y\,{}^p_x) + \tilde\measuredangle(y\,{}^p_z) \leqslant \pi.$$

It remains to apply Alexandrov's lemma (16.12).

(b)⇒(c). Applying *(b)* twice, first for $y \in [x,z]$ and then for $w \in [p,x]$, we get that

$$\tilde\measuredangle(x\,{}^w_y) \geqslant \tilde\measuredangle(x\,{}^p_y) \geqslant \tilde\measuredangle(x\,{}^p_z)$$

and therefore

$$|w-y|_\Sigma \geqslant |\tilde w - \tilde y|_{\mathbb{R}^2},$$

where $\tilde w$ and $\tilde y$ are the points corresponding to w and y points on the sides of the model triangle.

(c)⇒(a). Since the triangle is fat, we have

$$\tilde\measuredangle(x\,{}^w_y) \geqslant \tilde\measuredangle(x\,{}^p_z)$$

for any $w \in\,]xp]$ and $y \in\,]xz]$. Note that $\tilde\measuredangle(x\,{}^w_y) \to \measuredangle[x\,{}^p_z]$ as $w, y \to x$. Whence the implication follows. □

16.13. Exercise. *Let Σ be a proper smooth regular surface, γ a unit-speed geodesic in Σ, and $p \in \Sigma$.*

Consider the function

$$h(t) = |p - \gamma(t)|^2_\Sigma - t^2.$$

(a) Show that if Σ is simply-connected and the Gauss curvature of Σ is nonpositive, then the function h is convex.
(b) Show that if the Gauss curvature of Σ is nonnegative, then the function h is concave.

16.14. Exercise. *Let $\bar x$ and $\bar y$ be the midpoints of minimizing geodesics $[p,x]$ and $[p,y]$ in an open smooth regular surface Σ.*
(a) Show that if Σ is simply-connected and has nonpositive Gauss curvature, then

$$2{\cdot}|\bar x - \bar y|_\Sigma \leqslant |x-y|_\Sigma.$$

(b) *Show that if the Gauss curvature of Σ is nonnegative, then*

$$2{\cdot}|\bar{x} - \bar{y}|_\Sigma \geqslant |x - y|_\Sigma.$$

16.15. Exercise. *Assume γ_1 and γ_2 are two geodesics in an open smooth regular simply-connected surface Σ with nonpositive Gauss curvature. Show that the function*

$$h(t) = |\gamma_1(t) - \gamma_2(t)|_\Sigma$$

is convex.

16.16. Advanced exercise. *Suppose that a smooth curve δ bounds a convex disc Δ in a proper smooth surface Σ; that is, for any two points $x, y \in \Delta$, any shortest path $[x, y]_\Sigma$ lies in Δ.*

Denote by $f\colon \Sigma \to \mathbb{R}$ the intrinsic distance function to δ; that is,

$$f(x) = \inf_{y\in\delta}\{\, |x - y|_\Sigma \,\}.$$

Show that

(a) *If Σ is simply-connected and $K_\Sigma \leqslant 0$, then the restriction of f to the complement of Δ is convex; that is, for any geodesic γ in $\Sigma\backslash\Delta$ the function $t \mapsto f\circ\gamma(t)$ is convex.*

(b) *If $K_\Sigma \geqslant 0$, then the restriction of f to Δ is concave; that is, for any geodesic γ in Δ the function $t \mapsto f\circ\gamma(t)$ is concave.*

H Busemann functions

A unit-speed geodesic $\lambda\colon [0, \infty) \to \mathcal{X}$ is called a *half-line* if it is minimizing on each interval $[a, b] \subset [0, \infty)$.

16.17. Proposition. *Suppose that $\lambda\colon [0, \infty) \to \Sigma$ is a half-line in a smooth regular surface Σ. Then the function*

❶ $$\operatorname{bus}_\lambda(x) = \lim_{t\to\infty} |\lambda(t) - x|_\Sigma - t$$

is defined.

Moreover,

(a) $\operatorname{bus}_\lambda$ *is a 1-Lipschitz function and*

$$\operatorname{bus}_\lambda \circ\lambda(t) + t = 0$$

for any t.

(b) *If Σ is an open simply-connected surface with nonpositive Gauss curvature, then $\operatorname{bus}_\lambda$ is convex; that is, for any geodesic α the real-to-real function $t \mapsto \operatorname{bus}_\lambda \circ\alpha(t)$ is convex.*

(c) *If Σ is an open surface with nonnegative Gauss curvature, then* bus_λ *is concave; that is, for any geodesic α the real-to-real function $t \mapsto$ $\mapsto \mathrm{bus}_\lambda \circ\alpha(t)$ is concave.*

The function $\mathrm{bus}_\lambda\colon \Sigma \to \mathbb{R}$ as in the proposition is called the *Busemann function associated to λ*. Intuitively the function bus_λ can be described as a distance function to the ideal point at infinity at the end of the half-line λ.

Proof. By the triangle inequality, the function

$$t \mapsto |\lambda(t) - x| - t$$

is nonincreasing in t. Also by the triangle inequality,

$$|\lambda(t) - x|_\Sigma - t \geqslant -|\lambda(0) - x|;$$

that is, for each x, the values $|\lambda(t) - x|_\Sigma - t$ are bounded below. Thus the limit in ❶ is defined.

Observe that each function $x \mapsto |\lambda(t)-x|_\Sigma - t$ is 1-Lipschitz. Therefore its limit $x \mapsto \mathrm{bus}_\lambda(x)$ is 1-Lipschitz as well. The second part of *(a)* is evident, since for $s \geqslant t$, we have $|\gamma(s) - \gamma(t)|_\Sigma = s - t$.

It remains to prove the last two statements. Choose a geodesic α. Given $t \geqslant 0$, consider the function $h_t(s) = |\lambda(t) - \alpha(s)|^2_\Sigma - s^2$.

Observe that for any fixed $x \in \Sigma$, we have $|\lambda(t) - x|/t \to 1$ as $t \to \infty$. Therefore

$$\mathrm{bus}_\lambda \circ\alpha(s) = \lim_{t\to\infty} \frac{|\lambda(t) - \alpha(s)|^2_\Sigma - s^2}{t} - t = \lim_{t\to\infty} \frac{h_t(s)}{t} - t$$

According to 16.13, the function $s \mapsto h_t(s)$ is convex or concave, assuming the conditions in *(b)* or *(c)* respectively. So is

$$s \to \frac{h_t(s)}{t} - t$$

for every t. Since pointwise limits of convex or concave functions are respectively convex or concave, *(b)* and *(c)* follow. □

16.18. Exercise. *Let Σ be an open surface and $p \in \Sigma$.*

(a) *Show that there is a half-line λ in Σ that starts at p.*
Moreover, if $K \ni p$ is a noncompact closed convex subset of Σ, then there is a half-line of Σ that starts at p and runs in K.

(b) *Suppose Σ has nonnegative Gauss curvature at any point. Consider the function*

$$f(x) = \inf_\lambda \mathrm{bus}_\lambda(x),$$

where the greatest lower bound is taken among all half-lines λ that start at p. Show that f is a concave function and its suplevel sets

$$S_c = \{\, x \in \Sigma \,:\, f(x) \geqslant c \,\}$$

are compact for any $c \in \mathbb{R}$.

(c) Let $s = \max\{\, f(x) : x \in \Sigma \,\}$. Show that the set S_s is either one-point, a geodesic arc or a closed geodesic. Show that all these possibilities can occur.

I Line splitting theorem

Let Σ be a smooth regular surface. A unit-speed geodesic $\lambda\colon \mathbb{R} \to \Sigma$ is called a *line* if it is length-minimizing on each interval $[a, b] \subset \mathbb{R}$.

16.19. Line splitting theorem. *Let Σ be an open smooth regular surface with nonnegative Gauss curvature and λ be a line in Σ. Then Σ admits an intrinsic isometry to the Euclidean plane or a circular cylinder $\{\,(x, y, z) \in \mathbb{R}^2 : x^2 + y^2 = r^2\,\}$ for some $r > 0$.*

In particular, Σ has zero Gauss curvature at every point.

This theorem was proved by Stefan Cohn-Vossen [Satz 5 in 23] and it has a sequence of variations in differential geometry:

- ◇ Victor Toponogov [78] proved a version of the splitting theorem for Riemannian manifolds with nonnegative sectional curvature;
- ◇ Jeff Cheeger and Detlef Gromoll [21] generalized it further to Riemannian manifolds with nonnegative Ricci curvature;
- ◇ Jost-Hinrich Eshenburg [27] proved a splitting theorem for space-time with nonnegative Ricci curvature in timelike directions.

Proof. Consider two Busemann functions bus_+ and bus_- associated with half-lines $\lambda : [0, \infty) \to \Sigma$ and $\lambda : (-\infty, 0] \to \Sigma$; that is,

$$\mathrm{bus}_\pm(x) = \lim_{t\to\infty} |\lambda(\pm t) - x|_\Sigma - t.$$

Step 1. Let us show and use that

❶ $$\mathrm{bus}_+(x) + \mathrm{bus}_-(x) = 0$$

for any $x \in \Sigma$.

Fix $x \in \Sigma$. Since λ is a line, the triangle inequality implies that

$$\begin{aligned}|\lambda(t) - x|_\Sigma + |\lambda(-t) - x|_\Sigma \geqslant |\lambda(t) - \lambda(-t)|_\Sigma = \\ = 2{\cdot}t.\end{aligned}$$

Passing to the limit as $t\to\infty$, we get

$$\operatorname{bus}_+(x)+\operatorname{bus}_-(x)\geqslant 0.$$

On the other hand, by 16.13, the function $h(t)=|\lambda(t)-x|^2_\Sigma-t^2$ is concave. In particular,

$$|\lambda(t)-x|_\Sigma\leqslant\sqrt{t^2+at+b}$$

for some constants $a,b\in\mathbb{R}$. Passing to the limit as $t\to\pm\infty$, we get

$$\operatorname{bus}_+(x)+\operatorname{bus}_-(x)\leqslant 0;$$

whence ❶ follows.

Conclusions. According to 16.17, both functions $\operatorname{bus}_\pm$ are concave and $\operatorname{bus}_\pm\circ\lambda(t)=\mp t$ for any t. By ❶ both functions $\operatorname{bus}_\pm$ are affine; that is, they are convex and concave at the same time. It follows that the differential of $\operatorname{bus}_\pm$ is defined at any point $x\in\Sigma$; that is, there is a linear function $\mathrm{T}_x\to\mathbb{R}$ that is defined by $\mathrm{v}\mapsto D_\mathrm{v}\operatorname{bus}_\pm$ for any tangent vector $\mathrm{v}\in T_x$.

Denote by u the gradient vector field of bus_-; that is, u is a tangent vector field such that for any tangent field v the following identity holds

$$\langle\mathrm{u},\mathrm{v}\rangle=D_\mathrm{v}(\operatorname{bus}_-).$$

Step 2. Let us show that, the surface Σ can be subdivided into lines that run in the direction of u.

Fix a point x. Given a real value a choose a shortest path $[x,\lambda(a)]$; denote by $\mathrm{w}^a\in\mathrm{T}_x$ the unit vector in the direction of the geodesic $[x,\lambda(a)]$. Since $|\mathrm{w}^a|=1$ and bus_- is affine, we get that

$$\begin{aligned}|\mathrm{u}|&\geqslant\varlimsup_{a\to\infty}\langle\mathrm{u},\mathrm{w}^a\rangle=\\&=\varlimsup_{a\to\infty}D_{\mathrm{w}^a}\operatorname{bus}_-=\\&=\varlimsup_{a\to\infty}\frac{\operatorname{bus}_-\circ\lambda(a)-\operatorname{bus}_-(x)}{|x-\lambda(a)|_\Sigma}=\\&=\varlimsup_{a\to\infty}\frac{a-\operatorname{bus}_-(x)}{a}=\\&=1.\end{aligned}$$

On the other hand, since bus_- is 1-Lipschitz, we have $|\mathrm{u}|\leqslant 1$. Whence

$$|\mathrm{u}|\equiv 1\quad\text{and}\quad\lim_{a\to\infty}\langle\mathrm{u},\mathrm{w}^a\rangle=1.$$

It follows that

$$\lim_{a\to\infty} \measuredangle(\mathrm{U}, \mathrm{W}^a) = 0;$$

analogously, we get

$$\lim_{a\to-\infty} \measuredangle(\mathrm{U}, \mathrm{W}^a) = \pi.$$

Set $b = \mathrm{bus}_-(x)$. Consider a unit-speed geodesic ζ such that $\zeta'(b) = = \mathrm{U}(x)$. Since bus_- is affine, we have that $\mathrm{bus}_- \circ\zeta(t) = t$ for any t. Since ζ is a unit-speed geodesic and bus_- is 1-Lipschitz, we get

$$\begin{aligned}|t_1 - t_0| &\geqslant |\zeta(t_1) - \zeta(t_0)|_\Sigma \geqslant \\ &\geqslant |\,\mathrm{bus}\circ\lambda(t_1) - \mathrm{bus}\circ\lambda(t_0)| = \\ &= |t_1 - t_0|.\end{aligned}$$

Whence ζ is a line for any x. Moreover ζ always runs in the direction of U.

Step 3. Let us show that the distances between points on two lines in the direction of U behave the same way as the distances between parallel lines in the Euclidean plane; here is a precise formulation:

❷ *Let ξ and ζ be two lines in Σ that run in the direction of* U*. Suppose that ξ and ζ are parametrized so that* $\mathrm{bus}_- \circ\xi(t) = \mathrm{bus}_- \circ\zeta(t) = t$ *for any t; further set $x_0 = \xi(0)$, $z_0 = \zeta(0)$, $x_1 = \xi(s)$, $z_1 = \zeta(t)$ for some s, t. Then*

$$|x_1 - z_1|^2_\Sigma = |x_0 - z_0|^2_\Sigma + (s - t)^2.$$

Given $x \in \Sigma$, let $t \to \delta^x_a(t)$ be the parametrization of $[x, \lambda(a)]$ by arc-length starting from x. Since $\measuredangle(\mathrm{U}, \mathrm{W}^a) \to 0$ as $a \to \infty$, we get that

$$\delta^x_a(t) \to \zeta(b + t) \quad \text{as} \quad a \to \infty$$

for any fixed $t \geqslant 0$.

Analogously, since $\measuredangle(\mathrm{U}, \mathrm{W}^a) \to \pi$ as $a \to -\infty$, and therefore

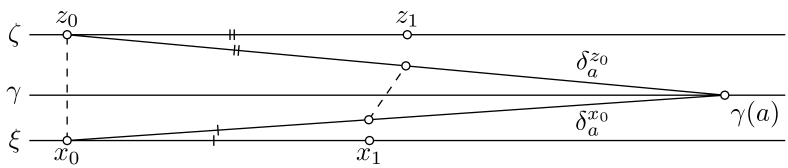
$$\delta^x_a(t) \to \zeta(b - t) \quad \text{as} \quad a \to -\infty$$

for any fixed $t \geqslant 0$.

Assume that $s, t \geqslant 0$, then

$$x_1 = \lim_{a\to\infty} \delta^{x_0}_a(s), \qquad\qquad z_1 = \lim_{a\to\infty} \delta^{z_0}_a(t).$$

Recall that the triangle $[x_0\,\gamma(a)\,z_0]$ is fat (16.11c). Therefore we get a lower bound for the distance $|\delta^{x_0}_a(s) - \delta^{z_0}_a(t)|_\Sigma$. Note that if a is large,



then the two long sides of the triangle are close to a. Straightforward computations show that passing to the limit as $a \to \infty$, we get

$$|x_1 - z_1|^2_\Sigma \geqslant |x_0 - z_0|^2_\Sigma + (s-t)^2.$$

Now let us swap x_0 with x_1 and z_0 with z_1, and repeat the argument above for $a \to -\infty$. Note that the two longer sides of the triangle are close to $s - a$ and $t - a$. Therefore we get the opposite inequality

$$|x_1 - z_1|^2_\Sigma \leqslant |x_0 - z_0|^2_\Sigma + (s-t)^2.$$

Whence

❸ $$|x_1 - z_1|^2_\Sigma = |x_0 - z_0|^2_\Sigma + (s-t)^2.$$

if $s, t \geqslant 0$. The same argument proves ❸ if $s, t \leqslant 0$. Applying it a couple of times for $s_1 = t_1 \geqslant 0$ and $s_2, t_2 \leqslant 0$, we get that ❸ holds for $s = s_1 + s_2$ and $t = t_1 + t_2$. Whence ❷ follows for any pair s and t.

Final step. Note that since bus_- is affine, the level set

$$L = \{\, x \in \Sigma \,:\, \mathrm{bus}_-(x) = 0 \,\}$$

is closed *totally convex* and *geodesic*; that is, if a geodesic α has two common points with L, then α lies in L. It follows that L is either closed or both-sides-infinite geodesic.

Choose an arc-length parametrization $v \mapsto \gamma(v)$ of L by a circle (or a line). Denote by λ^v the line thru $\gamma(v)$ in the direction of U with the parametrization as above: that is, $\mathrm{bus}_- \circ \lambda^v(u) = u$ for any u and v. According to ❷, $(u, v) \mapsto \lambda^v(u)$ is an intrinsic isometry from a circular cylinder (or, respectively, the Euclidean plane) to Σ. $\square$

16.20. Exercise. *Let Σ be an open smooth surface with nonnegative Gauss curvature. Suppose that Σ has a line and a half-line that meet at exactly one point. Show that Σ admits an intrinsic isometry to the Euclidean plane.*

Semisolutions

0.1. Check all the conditions in the definition of metric, page 7.

0.2; *(a)*. Observe that $|p-q|_{\mathcal{X}} \leqslant 1$. Apply the triangle inequality to show that $|p-x|_{\mathcal{X}} \leqslant 2$ for any $x \in B[q,1]$. Make a conclusion.

(b). Take $\mathcal{X}$ to be the half-line $[0,\infty)$ with the standard metric; $p = 0$ and $q = \frac{4}{5}$.

0.5. Show that the conditions in 0.4 hold for $\delta = \varepsilon$.

0.8. Suppose the complement $\Omega = \mathcal{X}\backslash Q$ is open. Then for each point $p \in \Omega$ there is $\varepsilon > 0$ such that $|p-q|_{\mathcal{X}} > \varepsilon$ for any $q \in Q$. It follows that p is not a limit point of any sequence $q_n \in Q$. That is, any limit of a sequence in Q lies in Q; that is, by the definition, Q is closed.

Now suppose $\Omega = \mathcal{X}\backslash Q$ is not open. Show that there is a point $p \in \Omega$ and a sequence $q_n \in Q$ such that such that $|p - q_n|_{\mathcal{X}} < \frac{1}{n}$ for any n. Conclude that $q_n \to p$ and $n \to \infty$; therefore Q is not closed.

1.2; *(a)*. Use that a continuous injection defined on a compact domain is an embedding (0.27).

parit*(b)*. The image of γ might have the shape of the digit 8 or 9.

1.3. Let α be a path, connecting p to q.

Passing to an arc of α if necessary, we can assume that $\alpha(t) \neq p, q$ for $t \neq 0, 1$.

An open set Ω in $(0,1)$ will be called *suitable* if for any connected component (a,b) of Ω we have $\alpha(a) = \alpha(b)$. Show that the union of nested suitable sets is suitable. Therefore we can find a maximal suitable set $\hat{\Omega}$.

Define $\beta(t) = \alpha(a)$ for any t in a connected component $(a,b) \subset \hat{\Omega}$, and $\beta(t) = \alpha(t)$ for t not in $\hat{\Omega}$. Note that for any $x \in [0,1]$ the set $\beta^{-1}\{\beta(x)\}$ is connected.

It remains to reparametrize β to make it injective. In order to do that, we need to construct a non-decreasing surjective function $\tau\colon [0,1] \to [0,1]$ such that $\tau(t_1) = \tau(t_2)$ if and only if there is a connected component $(a,b) \subset \hat{\Omega}$ such that $t_1, t_2 \in [a,b]$.

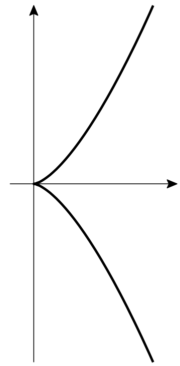
The required function τ can be constructed in a similar way to the so-called *devil's staircase* — learn this construction and modify it.

The simple arc we are looking for is $\beta \circ \sigma$, where $\sigma : [0,1] \to [0,1]$ is any right inverse of τ.

1.4. Denote the union of two half-axis by L.

Recall that the function f is smooth; see Section 0E. Therefore $t \mapsto \alpha(t) = (f(t), f(-t))$ is smooth as well.

Observe that $f(t) \to \infty$ as $t \to \infty$. Since $f(0) = 0$, the intermediate value theorem implies that $f(t)$ takes all nonnegative values for $t \geqslant 0$. Use it to show that L is the range of α.



Further, show that the function f is strictly increasing for $t > 0$. Use this to show that the map $t \mapsto \alpha(t)$ is injective.

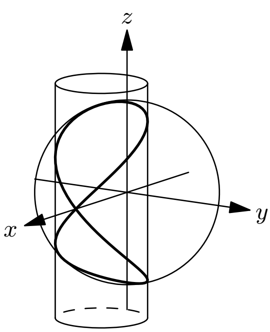
Summarizing, we get that α is a smooth parametrization of L.

Now suppose $\beta\colon t \mapsto (x(t), y(t))$ is a smooth parametrization of L. Without loss of generality we may assume that $x(0) = y(0) = 0$. Note that $x(t) \geqslant 0$ for any t therefore $x'(0) = 0$. The same way we get that $y'(0) = 0$. That is, $\beta'(0) = 0$; so L does not admit a smooth regular parametrization.

1.5. Apply the definitions. For *(b)* you need to check that $\gamma'_\ell \neq 0$. For *(c)* you need to check that $\gamma_\ell(t_0) = \gamma(t_1)$ only if $t_0 = t_1$.

1.6. This is the so-called *semicubical parabola*; it is shown on the diagram. Try to argue similarly to 1.4.

1.7. For $\ell = 0$ the system describes a pair of points $(0, 0, \pm 1)$, so we can assume that $\ell \neq 0$. Note that the first equation describes the unit sphere centered at the origin and the second equation describes a cylinder over the circle in the (x, y)-plane with center $(-\frac{\ell}{2}, 0)$ and radius $|\frac{\ell}{2}|$.



For $\ell \neq 0$, find the gradients ∇f and ∇h for the functions

$$f(x, y, z) = x^2 + y^2 + z^2 - 1$$
$$h(x, y, z) = x^2 + \ell{\cdot}x + y^2$$

and show that they are linearly dependent only on the x-axis. Conclude that for $\ell \neq \pm 1$ each connected component of the set of solutions is a smooth regular curve.

Show that

- ◇ if $|\ell| < 1$, then the set has two connected components with $z > 0$ and $z < 0$.
- ◇ if $|\ell| \geqslant 1$, then the set is connected.

Note that the linear independence of the gradients provides only a sufficient condition. Therefore the case $\ell = \pm 1$ has to be checked by hand. In this case a neighborhood of $(\pm 1, 0, 0)$ does not admit a smooth regular parametrization — try to prove it. The case $\ell = 1$ is shown on the diagram.

Remark. The case $\ell = \pm 1$ is called *Viviani's curve.* It admits the following smooth regular parametrization with a self-intersection:

$$t \mapsto (\pm(\cos t)^2, \cos t \cdot \sin t, \sin t).$$

1.8. Assume contrary, then there is a sequence of real numbers $t_n \to \pm\infty$ such that $\gamma(t_n)$ converges; denote its limit by p. Let K be a closed ball centered at p. Observe that $\gamma^{-1}(K)$ is not compact. Conclude that γ is not proper.

1.9. Show and use that a set $C \subset \mathbb{R}^3$ is closed if and only if the intersection $K \cap C$ is compact for any compact $K \subset \mathbb{R}^3$.

1.10. Without loss of generality we may assume that the origin does not lie on the curve.

Show that inversion of the plane $(x,y) \mapsto (\frac{x}{x^2+y^2}, \frac{y}{x^2+y^2})$ maps our curve to a closed curve with the origin removed. Apply Jordan's theorem for the obtained curve and use the inversion again.

2.2. Show that the least upper bound in 2.1 can be taken for all sequences $a = t_0 \leqslant t_1 \leqslant \dots$ $\dots \leqslant t_k = b$.

Suppose that γ_2 is reparametrization of γ_1 by $\tau\colon [a_1, b_1] \to [a_2, b_2]$; without loss of generality we may assume that τ is nondecreasing. Set $\theta_i = \tau(t_i)$. Observe that $a_2 = \theta_0 \leqslant$ $\leqslant \theta_1 \leqslant \dots \leqslant \theta_k = b_2$ if and only if $a_1 = t_0 \leqslant t_1 \leqslant \dots \leqslant t_k = b_1$. Make a conclusion.

2.3. Choose a partition $0 = t_0 < \dots < t_n = 1$ of $[0,1]$. Set $\tau_0 = 0$ and

$$\tau_i = \max\{\, \tau \in [0,1] \,:\, \beta(\tau_i) = \alpha(t_i)\,\}$$

for $i > 0$. Show that (τ_i) is a partition of $[0,1]$; that is, $0 = \tau_0 < \tau_1 < \dots < \tau_n = 1$.

By construction

$$\begin{aligned}|\alpha(t_0) - \alpha(t_1)| + |\alpha(t_1) - \alpha(t_2)| + \dots + |\alpha(t_{n-1}) - \alpha(t_n)| &= \\ = |\beta(\tau_0) - \beta(\tau_1)| + |\beta(\tau_1) - \beta(\tau_2)| + \dots + |\beta(\tau_{n-1}) - \beta(\tau_n)|.\end{aligned}$$

Since the partition (t_i) is arbitrary, we get

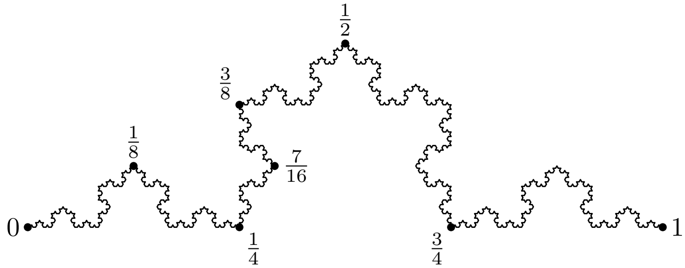
$$\operatorname{length} \beta \geqslant \operatorname{length} \alpha.$$

Remark. Note that the partition (τ_i) is not arbitrary, therefore the inequality might be strict; it might happen if β runs back and forth along α.

2.4. For *(a)*, apply the fundamental theorem of calculus for each segment in a given partition. For *(b)* consider a partition such that the velocity vector $\alpha'(t)$ is nearly constant on each of its segments.

2.5. Use the theorems of Rademacher and Lusin (0.16 and 0.17).

2.6; *(a)*. Look at the diagram and guess the parametrization of an arc of the snow flake



by $[0,1]$. Extend it to the whole snow flake. Show that it indeed describes an embedding of the circle in the plane.

(b). Suppose that $\gamma\colon[0,1]\to\mathbb{R}^2$ is a rectifiable curve and let γ_k be a scaled copy of γ with factor $k>0$; that is, $\gamma_k(t)=k\cdot\gamma(t)$ for any t. Show that

$$\operatorname{length}\gamma_k=k\cdot\operatorname{length}\gamma.$$

Now suppose the arc γ of the Koch snowflake shown on the diagram is rectifiable; denote its length by ℓ. Observe that γ can be divided into 4 arcs such that each one is a scaled copy of γ with factor $\frac13$. It follows that $\ell=\frac43\cdot\ell$. Evidently $\ell>0$ — a contradiction.

2.8. We have to assume that $a\neq0$ or $b\neq0$; otherwise we get a constant curve.

Show that the curve has constant velocity $|\gamma'(t)|\equiv\sqrt{a^2+b^2}$. Therefore $s=t/\sqrt{a^2+b^2}$ is an arc-length parameter. It remains to substitute $s\cdot\sqrt{a^2+b^2}$ for t.

2.12. Choose a closed polygonal line $p_1\dots p_n$ inscribed in β. By 2.11, we can assume that its length is arbitrary close to the length of β; that is, given $\varepsilon>0$

$$\operatorname{length}(p_1\dots p_n)>\operatorname{length}\beta-\varepsilon.$$

Show that we may assume in addition that each point p_i lies on α.

Observe that since α is simple, the points $p_1,\dots,p_n$ appear on α in the same cyclic order; that is, the polygonal line $p_1\dots p_n$ is also inscribed in α. In particular

$$\operatorname{length}\alpha\geqslant\operatorname{length}(p_1\dots p_n).$$

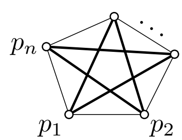
It follows that

$$\operatorname{length}\alpha>\operatorname{length}\beta-\varepsilon.$$

for any $\varepsilon>0$. Whence

$$\operatorname{length}\alpha\geqslant\operatorname{length}\beta.$$

If α has self-intersections, then the points $p_1,\dots,p_n$ might appear on α in a different order, say $p_{i_1},\dots,p_{i_n}$. Apply the triangle inequality to show that



$$\operatorname{length}(p_{i_1}\dots p_{i_n})\geqslant\operatorname{length}(p_1\dots p_n)$$

and use it to modify the proof above.

2.14. Denote by ℓ_{U} the line segment obtained by orthogonal projection of γ to the line in the direction U. Since γ_{U} runs along ℓ_{U} back and forth, we get

$$\operatorname{length}\gamma_{\mathrm{U}}\geqslant2\cdot\operatorname{length}\ell_{\mathrm{U}}.$$

Applying the Crofton formula, we get that

$$\operatorname{length}\gamma\geqslant\pi\cdot\overline{\operatorname{length}\ell_{\mathrm{U}}}.$$

In the case of equality, the curve γ_{U} runs exactly back and forth along ℓ_{U} without additional zigzags for almost all (and therefore for all) U.

Let K be a closed set bounded by γ. Observe that the last statement implies that every line may intersect K only along a closed segment or a single point. It follows that K is convex.

2.15. The proof is identical to the proof of the standard Crofton formula. To find the coefficient it is sufficient to check it on a unit interval. The latter can be done by integration:

$$\frac{1}{k_a}=\frac{1}{\operatorname{area}\mathbb{S}^2}\cdot\iint\limits_{\mathbb{S}^2}|x|; \qquad\qquad \frac{1}{k_b}=\frac{1}{\operatorname{area}\mathbb{S}^2}\cdot\iint\limits_{\mathbb{S}^2}\sqrt{1-x^2}.$$

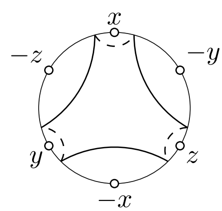
The answers are $k_a=2$ and $k_b=\frac{4}{\pi}$.

2.17. The "only-if" part is trivial. To show the "if" part, assume A is not convex; that is, there are points $x,y\in A$ and a point $z\notin A$ that lies between x and y.

Since A is closed, its complement is open. That is, the ball $B(z,\varepsilon)$ does not intersect A for some $\varepsilon>0$.

Show that there is $\delta>0$ such that any path of length at most $|x-y|_{\mathbb{R}^3}+\delta$ passes thru $B(z,\varepsilon)$. It follows that $|x-y|_A\geqslant|x-y|_{\mathbb{R}^3}+\delta$, in particular $|x-y|_A\neq|x-y|_{\mathbb{R}^3}$.

2.20. The spherical curve shown on the diagram does not have antipodal pairs of points. However it has three points x,y,z on one of its sides and their antipodal points $-x,-y,-z$ on the other. Show that this property is sufficient to conclude that the curve does not lie in any hemisphere.



2.21. Assume contrary, then by the hemisphere lemma (2.19) γ lies in an open hemisphere. In particular it cannot divide $\mathbb{S}^2$ into two regions of equal area — a contradiction.

2.22. The very first sentence is wrong — it is *not* sufficient to show that diameter is at most 2. For example an equilateral triangle with circumradius slightly above 1 may have diameter (which is defined as the maximal distance between its points) slightly bigger than $\sqrt{3}$, so it can be made smaller that 2.

On the other hand, it is easy to modify the proof of the hemisphere lemma (2.19) to get a correct solution. That is, (1) choose two points p and q on γ that divide it into two arcs of the same length; (2) set z to be a midpoint of p and q, and (3) show that γ lies in the unit disc centered at z.

2.23. For *(a)*, modify the proof of the original Crofton formula (Section 2E).

(b). Assume $\operatorname{length}\gamma<2\cdot\pi$. By *(a)*,

$$\overline{\operatorname{length}\gamma_{\textsc{u}}}<2\cdot\pi.$$

Therefore we can choose $\textsc{u}$ so that

$$\operatorname{length}\gamma_{\textsc{u}}<2\cdot\pi.$$

Observe that $\gamma_{\textsc{u}}$ runs in a semicircle h and therefore γ lies in a hemisphere with h as a diameter.

3.2. For a unit-speed parametrization γ, we have $\gamma''\equiv0$. Therefore the velocity $\textsc{v}=\gamma'$ is constant. Hence $\gamma(t)=p+(t-t_0)\cdot\textsc{v}$, where $p=\gamma(t_0)$.

3.3. Observe that $\alpha(t):=\gamma_\lambda(t/\lambda)$ is a unit-speed parametrization of the curve γ_λ. Apply the chain rule twice.

3.4. Differentiate the identity $\langle\gamma(s),\gamma(s)\rangle=1$ a couple of times.

3.5. Set $\mathrm{T}=\frac{\gamma'}{|\gamma|}$. Prove and use the following identities:

$$\gamma''-(\gamma'')^{\perp}=\mathrm{T}\cdot\langle\gamma'',\mathrm{T}\rangle, \qquad\qquad |\gamma'|=\sqrt{\langle\gamma',\gamma'\rangle}.$$

3.6. Apply 3.5a for the parametrization $t\mapsto(t,f(t))$.

3.7. Without loss of generality we may assume that γ has a unit-speed parametrization.

Consider the tangent indicatrix $\mathrm{T}(s)=\gamma'(s)$. Note that T is a spherical curve and $|\mathrm{T}'|\leqslant 1$. Use it to construct a sequence of unit-speed spherical curves $\mathrm{T}_n\colon\mathbb{I}\to\mathbb{S}^2$ such that $\mathrm{T}_n(s)\to\mathrm{T}(s)$ as $n\to\infty$ for any s.

It remains to show that the following sequence of curves solves the problem:

$$\gamma_n(s)=\gamma(a)+\int\limits_a^s\mathrm{T}_n(t)\cdot dt.$$

3.8. Show that $\gamma''_{a,b}\perp\gamma'_{a,b}$ and apply 3.5a.

3.11. Apply Fenchel's theorem.

3.12. We can assume that γ is a unit-speed curve. Set $\theta(s)=\measuredangle(\gamma(s),\gamma'(s))$. Since $\langle\mathrm{T},\mathrm{T}\rangle=1$, we have $\mathrm{T}'\perp\mathrm{T}$. Observe that $\langle\mathrm{T},\sigma\rangle=\cos\theta$; therefore

$$\kappa\cdot\sin\theta=|\mathrm{T}'|\cdot\sin\theta\geqslant-\langle\mathrm{T}',\sigma\rangle=\langle\mathrm{T},\sigma'\rangle-\langle\mathrm{T},\sigma\rangle'=(|\sigma'|+\theta')\cdot\sin\theta.$$

Whence $\kappa\geqslant|\sigma'|+\theta'$ if $\theta\ne 0,\pi$. It remains to integrate this inequality and show that the set with $\theta\ne 0$ or π does not create a problem.

An alternative solution can be built on 3.19.

3.14; *(a).* Observe that

$$\langle\gamma'(s),\gamma'(s)\rangle=1,\quad\text{and}\quad\langle\gamma(s),\gamma(s)\rangle\leqslant 1.$$

Therefore

$$\langle\gamma''(s),\gamma(s)\rangle\geqslant-|\gamma''(s)|\cdot|\gamma(s)|\geqslant-\kappa(s) \tag{❶}$$

for all s.

(b). Since γ is unit-speed, we have $\ell=\operatorname{length}\gamma$. Therefore

$$\int\limits_0^\ell\langle\gamma(s),\gamma'(s)\rangle'\cdot ds=\int\limits_0^\ell\langle\gamma'(s),\gamma'(s)\rangle\cdot ds+\int\limits_0^\ell\langle\gamma(s),\gamma''(s)\rangle\cdot ds\geqslant\operatorname{length}\gamma-\Phi(\gamma).$$

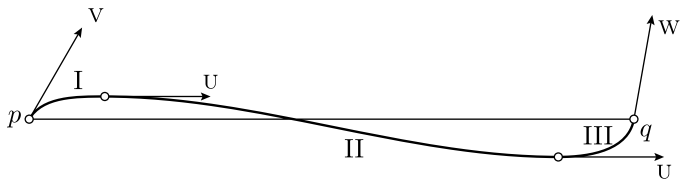
(c). By the fundamental theorem of calculus, we have

$$\int\limits_0^\ell\langle\gamma(s),\gamma'(s)\rangle'\cdot ds=\langle\gamma(\ell),\gamma'(\ell)\rangle-\langle\gamma(0),\gamma'(0)\rangle.$$

Since $\gamma(0) = \gamma(\ell)$ and $\gamma'(0) = \gamma'(\ell)$, the right hand side vanishes.

Note that without loss of generality we can assume the curve in 3.13 is described by a loop $\gamma\colon [0, \ell] \to \mathbb{R}^3$ parametrized by arc-length. We can also assume that the origin is the center of the ball; that is $|\gamma| \leqslant 1$. Since γ is a smooth closed curve, we have $\gamma'(0) = \gamma'(\ell)$ and $\gamma(0) = \gamma(\ell)$. Therefore *(b)* and *(c)* imply 3.13.

3.17. Set $\alpha = \measuredangle(\mathrm{W}, \mathrm{U})$ and $\beta = \measuredangle(\mathrm{V}, \mathrm{U})$. Try to guess the example from the diagram.



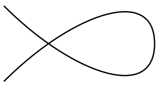
The shown curve is divided into three arcs: I, II, and III. Arc I turns from V to U; it has total curvature α. Similarly, the arc III turns from U to W and has total curvature β. Arc II goes very close and almost parallel to the chord pq and its total curvature can be made arbitrarily small.

3.18. Use the fact that an exterior angle of a triangle equals the sum of the two remote interior angles; for the second part apply induction on the number of vertices.

3.20. By 3.19, $\Phi(\gamma) \geqslant \Phi(\beta)$; it remains to show that $\Phi(\gamma) \leqslant \sup\{\Phi(\beta)\}$. In other words, given $\varepsilon > 0$ and any polygonal line $\sigma = \mathrm{U}_0 \dots \mathrm{U}_k$ inscribed in the tangent indicatrix T of γ, we need to construct a polygonal line β inscribed in γ such that

$$\text{❷} \qquad \operatorname{length} \sigma < \Phi(\beta) + \varepsilon.$$

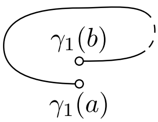
Suppose $\mathrm{U}_i = \mathrm{T}(s_i)$. Choose an inscribed polygonal line $\beta = p_0 \dots p_{2\cdot k+1}$ such that $p_{2\cdot i}$ and $p_{2\cdot i+1}$ lie sufficiency close to $\gamma(s_i)$; so we can assume that the direction of $p_{2\cdot i+1} - p_{2\cdot i}$ is sufficiency close to U_i for each i. Conclude that ❷ holds true for the constructed polygonal line β.



3.21. An example for *(a)* is shown on the diagram.

(b). Assume x is a point of self-intersection. Show that we may choose two points y and z on γ so that the triangle xyz is nondegenerate. In particular, $\measuredangle[y\,{}^{x}_{z}] + \measuredangle[z\,{}^{y}_{x}] < \pi$, or, equivalently, for a nonclosed polygonal line $xyzx$ we have $\Phi(xyzx) > \pi$. It remains to apply 3.19.

3.23. Consider the closed polygonal line $acbd$. Observe that $\Phi(acbd) = 4\cdot\pi$. It remains to apply 3.19.



3.25. Start with the curve γ_1 shown on the diagram. To obtain γ_2, slightly unbend (that is, decrease the curvature of) the dashed arc of γ_1.

3.26. Choose a value $s_0 \in [a, b]$ that splits the total curvature of γ into two equal parts. Observe that $\measuredangle(\gamma'(s_0), \gamma'(s)) \leqslant \theta$ for any s. Use this inequality in the same way as in the proof of the bow lemma.

3.27. Let $\ell = \operatorname{length} \gamma$. Suppose $\ell_1 < \ell < \ell_2$. Let γ_1 be an arc of unit circle with length ℓ.

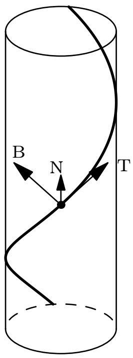
Show that the distance between the ends of γ_1 is smaller than $|p-q|$ and apply the bow lemma (3.24).

3.28. If $\operatorname{length}\gamma < 2{\cdot}\pi$, apply the bow lemma (3.24) to γ and an arc of the unit circle of the same length.

4.1. The arc-length parameter s is already found in 2.8. It remains to find the Frenet frame and calculate curvature and torsion. The latter can be done by straightforward calculations; the answers are

$$\mathrm{T}(t) = \tfrac{1}{\sqrt{a^2+b^2}}{\cdot}(-a{\cdot}\sin t, a{\cdot}\cos t, b), \qquad \mathrm{N}(t) = (-\cos t, -\sin t, 0),$$
$$\mathrm{B}(t) = \tfrac{1}{\sqrt{a^2+b^2}}{\cdot}(b{\cdot}\sin t, -b{\cdot}\cos t, a), \qquad \kappa \equiv \tfrac{a}{a^2+b^2}, \qquad \tau \equiv \tfrac{b}{a^2+b^2}.$$

It remains to show that the map $(a,b) \mapsto (\tfrac{a}{a^2+b^2}, \tfrac{b}{a^2+b^2})$ sends bijectively the half plane $a > 0$ onto itself.



4.2. By the product rule, we get

$$\mathrm{B}' = (\mathrm{T}\times\mathrm{N})' = \mathrm{T}'\times\mathrm{N} + \mathrm{T}\times\mathrm{N}'.$$

It remains to substitute the values from ❶ and ❷ and simplify.

4.3. This is a consequence of the equation $\mathrm{B}' = -\tau{\cdot}\mathrm{N}$.

4.4. Observe that $\frac{\gamma'\times\gamma''}{|\gamma'\times\gamma''|}$ is a unit vector perpendicular to the plane spanned by γ' and γ'', so, up to sign, it has to be equal to B. It remains to check that the sign is right.

4.6; *(a)*. Observe that $\langle \mathrm{W},\mathrm{T}\rangle' = 0$. Show that it implies that $\langle \mathrm{W},\mathrm{N}\rangle = 0$.
By Frenet formulas, $\langle \mathrm{W},\mathrm{N}\rangle' = 0$ implies that $\langle \mathrm{W}, -\kappa{\cdot}\mathrm{T} + \tau{\cdot}\mathrm{B}\rangle = 0$.

(b). Show that $\mathrm{W}' = 0$; it implies that $\langle \mathrm{W},\mathrm{T}\rangle = \frac{\tau}{\kappa}$. In particular, the velocity vector of γ makes a constant angle with W; that is, γ has constant slope.

4.7. Suppose α is an evolvent of γ and W is a fixed vector. Show that $\langle \mathrm{W},\alpha\rangle$ is constant if γ makes constant angle with W.

4.9. Part *(a)* follows from the fact that $(\mathrm{T},\mathrm{N},\mathrm{B})$ is an orthonormal basis. For *(b)* take the first and second derivatives of the identity $\langle\gamma,\gamma\rangle = 1$ and simplify using the Frenet formulas. Part *(c)* follows from *(b)* and the Frenet formulas. By *(c)*, $\int \frac{\tau}{\kappa} = 0$, hence *(d)* follows. Part *(e)* is proved by algebraic manipulations.

(f). Use the Frenet formulas to show that $(\gamma + \frac{1}{\kappa}{\cdot}\mathrm{N} + \frac{\kappa'}{\kappa^2{\cdot}\tau}{\cdot}\mathrm{B})' = 0$.

4.11. Use the second statement in 4.1.

4.12. Note that the function

$$\rho(\ell) = |\gamma(t+\ell) - \gamma(t)|^2 = \langle \gamma(t+\ell) - \gamma(t), \gamma(t+\ell) - \gamma(t)\rangle$$

is smooth and does not depend on t. Express speed, curvature and torsion of γ in terms of the derivatives $\rho^{(n)}(0)$. Be patient, you will need two derivatives for the speed, four for the curvature and six for the torsion. Once it is done, apply 4.11.

5.1. Without loss of generality, we may assume that γ_0 is parametrized by its arc-length. Then

$$|\gamma_1'| = |\gamma_0' + \mathrm{T}'| = |\mathrm{T} + \kappa\cdot\mathrm{N}| = \sqrt{1+\kappa^2} \geqslant 1 = |\gamma_0'|;$$

that is, $|\gamma_1'(t)| \geqslant |\gamma_0'(t)|$ for any $t \in [a, b]$. It remains to integrate this inequality and apply 2.4.

5.4. Observe that $\gamma_a'(t) = (1 + a\cdot\cos t, -a\cdot\sin t)$; that is, γ_a' runs clockwise along a circle with center at $(1, 0)$ and radius $|a|$.

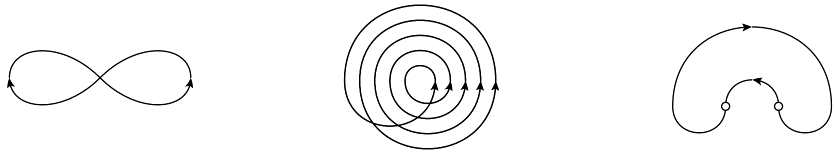
Case $|a| > 1$, Note that $\mathrm{T}_a(t) = \gamma_a'/|\gamma_a'|$ runs clockwise and makes a full turn in time $2\cdot\pi$. Therefore $\Psi(\gamma_a) = -2\cdot\pi$ and $\Phi(\gamma)_a = |\Psi(\gamma_a)| = 2\cdot\pi$.

Case $|a| < 1$. Set $\theta_a = \arcsin a$. Show that $\mathrm{T}_a(t) = \gamma_a'/|\gamma_a'|$ starts with the horizontal direction $\mathrm{T}_a(0) = (1, 0)$, turns monotonically to angle θ_a, then monotonically to $-\theta_a$ and then monotonically to back to $\mathrm{T}_a(2\cdot\pi) = (1, 0)$. It follows that if $|a| > 1$, then $\Psi(\gamma_a) = 0$ and $\Phi(\gamma_a) = 4\cdot\theta_a$.

Case $a = -1$. The velocity $\gamma_{-1}'(t)$ vanishes at $t = 0$ and $2\cdot\pi$. Nevertheless, the curve admits a smooth regular parametrization — find it. In this case $\Psi(\gamma_{-1}) = -\pi$ and $\Phi(\gamma_{-1}) = \pi$.

Case $a = 1$. The velocity $\gamma_1'(t)$ vanishes at $t = \pi$. At $t = \pi$ the curve has a cusp; that is, γ_1 turns exactly back at the time π. So $\gamma_1(t)$ has undefined total signed curvature. The curve is a joint of two smooth arcs with external angle π, and the total curvature of each arc is $\frac{\pi}{2}$, so $\Phi(\gamma_1) = \frac{\pi}{2} + \pi + \frac{\pi}{2} = 2\cdot\pi$.

5.6. Look at the drawings; we assume that the two marked points in the last example have parallel tangent lines.



5.7; *(a)*. Show that

$$\gamma_\ell'(t) = (1 - \ell\cdot k(t))\cdot\gamma'(t).$$

By regularity of γ, $\gamma' \neq 0$. So if $\gamma_\ell'(t) = 0$, then $\ell\cdot k(t) = 1$.

(b). Observe that we can assume that γ is parametrized by its arc-length, so $\gamma'(t) = \mathrm{T}(t)$. Suppose $|\ell| < \frac{1}{\kappa(t)}$ for any t. Then

$$|\gamma_\ell'(t)| = (1 - \ell\cdot k(t)).$$

Therefore

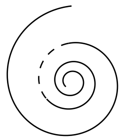
$$L(\ell) = \int_a^b (1 - \ell\cdot k(t))\cdot dt = \int_a^b 1\cdot dt - \ell\cdot \int_a^b k(t)\cdot dt = L(0) + \ell\cdot\Psi(\gamma).$$

(c). Consider the unit circle $\gamma(t) = (\cos t, \sin t)$ for $t \in [0, 2{\cdot}\pi]$ and γ_ℓ for $\ell = 2$.

5.10. Use the definition of osculating circle via order of contact and use the fact that inversions send circles to circlines.

5.12. Find $\mathrm{T}(t)$ and $\mathrm{N}(t)$. Use the formula in 3.5*b* to calculate curvature $\kappa(t)$. Apply the formula given right before the exercise.

5.14. Start with a plane spiral curve as shown on the diagram. Increase the torsion of the dashed arc without changing the curvature until a self-intersection appears.



5.15. Observe that if a line or circle is tangent to γ, then it is tangent to the osculating circle at the same point. Then apply the spiral lemma (5.11).

6.2. Apply the spiral lemma (5.11).

Direct solution. We will assume that the curvature does not vanish at p, the remaining case is simpler. We may assume that γ is a unit-speed curve and $p = \gamma(0)$. Further, we may assume that be the center of curvature of γ at p is the origin; in other words, $\kappa(0){\cdot}\gamma(0) + \mathrm{N}(0) = 0$.

Since the osculating circle supports γ at p, we get that the function $f\colon t \mapsto \langle\gamma(t), \gamma(t)\rangle$ has a local minimum or maximum at 0.

Direct calculations show that

$$\begin{aligned} f' &= \langle\gamma,\gamma\rangle' = 2{\cdot}\langle\mathrm{T},\gamma\rangle, \\ f'' &= 2{\cdot}\langle\mathrm{T},\gamma\rangle' = 2{\cdot}\langle\mathrm{T}',\gamma\rangle + 2{\cdot}\langle\mathrm{T},\mathrm{T}\rangle = 2{\cdot}\kappa{\cdot}\langle\mathrm{N},\gamma\rangle + 2, \\ f''' &= 2{\cdot}\kappa'{\cdot}\langle\mathrm{N},\gamma\rangle + 2{\cdot}\kappa{\cdot}\langle\mathrm{N}',\gamma\rangle + \cancel{2{\cdot}\kappa{\cdot}\langle\mathrm{N},\mathrm{T}\rangle}. \end{aligned}$$

Observe that $\mathrm{N}'(0) \perp \gamma(0)$. Therefore $f'(0) = 0$, $f''(0) = 0$, and $f'''(0) = -2{\cdot}\kappa'/\kappa$. Since f is a maximum or minimum at 0, we get that $f'''(0) = 0$ and therefore $\kappa' = 0$.

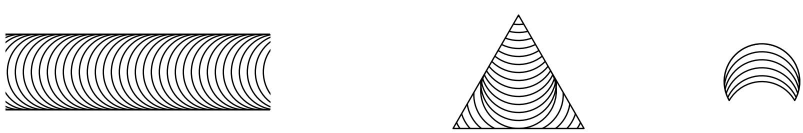
6.4. Consider the coordinate system with p as the origin and x-axis as the common tangent line to γ_1 and γ_2. We may assume that γ_i are defined in $(-\varepsilon, \varepsilon)$ for some small $\varepsilon > 0$, so that they run almost horizontally to the right.

Given $t \in [0,1]$ consider the curve γ_t that is tangent and cooriented to the x axis at $\gamma_t(0)$ and has signed curvature defined by $k_t(s) = (1-t){\cdot}k_0(s) + t{\cdot}k_1(s)$. It exists by 5.2.

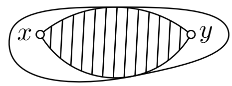
Choose $s \approx 0$. Consider the curve $\alpha_s\colon t \mapsto \gamma_t(s)$. Show that α_s moves almost vertically up. Use the fact that γ_t moves almost horizontally to the right to show that in a small neighborhood of p, the curve γ_1 lies above γ_0, whence the statement follows.

6.5. Reduce the radius of the circle until it touches γ. Observe that the circle supports γ and apply 6.3.

6.6 + 6.7 + 6.8. Observe that one of the arcs of curvature 1 in the families shown on the diagram supports γ and apply 6.3. To do the second part in 6.6, use the shown family



and another family of arcs curved in the opposite direction. (Compare to 6.15.)



6.11. Note that we can assume that γ bounds a convex figure F, otherwise by 6.9 its curvature changes sign and therefore it has zero curvature at some point.

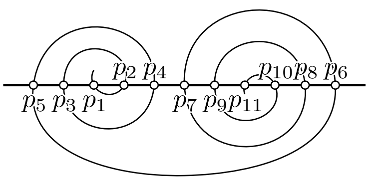
Choose two points x and y surrounded by γ such that $|x - y| > 2$. Look at the maximal lens bounded by two arcs with a common chord xy that lies in F. Apply the supporting test (6.3).

6.12; *(a).* Apply the lens lemma to show that if p_2 lies between p_1 and p_3, then the curvature of γ switches its sign.

(b). Show that p_4 lies between p_2 and p_3; further p_5 lies between p_3 and p_4; and so on.

(c). According to *(a)*, the point p_3 might lie between p_1 and p_2 and then further order is determined uniquely or p_1 lies between p_2 and p_3. In the latter case we have two choices, either p_4 lies between p_2 and p_3 and then further order is determined uniquely or p_2 lies between p_3 and p_4. In the latter case we get a choice again.

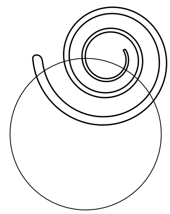
Assume we make the first choice on the step number k. Without loss of generality we may assume that p_k lies to the right from p_{k-2}. Then we have the following order:



$$p_{k-2}, p_{p-4}, \dots, p_{p-5}, p_{k-3}, p_k, p_{k+2}, \dots, p_{k+1}, p_{k-1}.$$

The case $k = 7$ is shown on the diagram.

6.15. Note that γ contains a simple loop; apply to it 6.13.



6.18. Repeat the proof of the theorem for each cyclic concatenation of an arc of γ from p_i to p_{i+1} with a large arc of the circle.

An example for the second part can be guessed from the diagram.

7.3. Denote the set of solutions by Σ_ℓ.

Show that $\nabla_p f = 0$ if and only if $p = (0,0,0)$. Use 7.2 to conclude that if $\ell \neq 0$, then Σ_ℓ is a union of disjoint smooth regular surfaces.

Show that Σ_ℓ is connected if and only if $\ell \leqslant 0$. It follows that if $\ell < 0$, then Σ_ℓ is a smooth regular surface, and if $\ell > 0$ — it is not.

The case $\ell = 0$ has to be done by hand — it does not satisfy the sufficient condition in 7.2, but that does not solely imply that Σ_0 is not a smooth surface.

Show that any neighborhood of origin in Σ_0 cannot be described by a graph in any coordinate system; so by the definition (Section 7B) Σ_0 is not a smooth surface.

7.6. First check that the image of s lies in the unit sphere centered at $(0,0,1)$; that is, show that

$$\left(\tfrac{2\cdot u}{1+u^2+v^2}\right)^2 + \left(\tfrac{2\cdot v}{1+u^2+v^2}\right)^2 + \left(\tfrac{2}{1+u^2+v^2} - 1\right)^2 = 1.$$

for any u and v.

Furthermore, show that the map $(x,y,z) \mapsto (\tfrac{2\cdot x}{x^2+y^2+z^2}, \tfrac{2\cdot y}{x^2+y^2+z^2})$ describes the inverse of s, which is continuous away from the origin. In particular, s is an embedding that covers the entire sphere except for the origin.

It remains to show that s is regular; that is, s_u and s_v are linearly independent at all points of the (u,v)-plane.

7.7. Set $s\colon (t,\theta)\mapsto (x(t), y(t)\cdot\cos\theta, y(t)\cdot\sin\theta)$. Show that s is regular; that is, s_t and s_θ are linearly independent. (It might help to observe that $s_t\perp s_\theta$).

Show that s is a embedding; that is, any (t_0,θ_0) admits a neighborhood U in the (t,θ)-plane such that the restriction $s|_U$ has a continuous inverse. It remains to apply 7.5.

7.8. The solutions of these exercises are built on the following general construction known as the *Moser trick.*

Suppose that u_t is a smooth time-dependent vector field on a plane. Consider the ordinary differential equation $x'(t)=\mathrm{u}_t(x(t))$. Consider the map $\iota\colon x(0)\mapsto x(1)$ where $t\mapsto x(t)$ is a solution of the equation. The map ι is called the *flow* of the vector field u_t for the time interval $[0,1]$. Observe that according to 0.25 the map ι is smooth in its domain of definition. Moreover the same holds for its inverse ι^{-1}; indeed, ι^{-1} is the flow of the vector field $-\mathrm{u}_{1-t}$. That is, ι is a diffeomorphism from its domain of definition to its image.

Therefore, in order to construct a diffeomorphism from one open subset of the plane to another it is sufficient to construct a smooth vector field such that its flow maps one set to the other; such a map is automatically a diffeomorphism.

(a). Suppose $\Sigma=\mathbb{R}^2\backslash p_1,\dots,p_n$ and $\Theta=\mathbb{R}^2\backslash q_1,\dots,q_n$. Choose smooth paths $\gamma_i\colon[0,1]\to\mathbb{R}^2$ such that $\gamma_i(0)=p_i$, $\gamma_i(1)=q_i$, and $\gamma_i(t)\ne\gamma_j(t)$ if $i\ne j$.

Choose a smooth vector field v_t such that $\mathrm{v}_t(\gamma_i(t))=\gamma_i'(t)$ for any i and t. We can assume in addition that v_t vanishes outside of a sufficiently large disc; this can be arranged by multiplying the vector field by a function $\sigma_1(R-|x|)$; see page 15.

It remains to apply the Moser trick to the constructed vector field.

(b)–(d). Without loss of generality we can assume that the origin belongs to both Σ and Θ.

In each case show that there is a vector field v defined on $\mathbb{R}^2\backslash\{0\}$ that flows Σ to Θ. In fact one can choose radial fields of that type, but be careful with the cases *(c)* and *(d)* — they are not as easy as one might think.

8.3. Let γ be a smooth curve in Σ. Observe that $f\circ\gamma(t)\equiv 0$. Differentiate this identity and apply the definition of tangent vector (8.1).

8.4. Assume a neighborhood of p in Σ is a graph $z=f(x,y)$. In this case $s\colon(u,v)\mapsto(u,v,f(u,v))$ is a smooth chart at p. Show that the plane spanned by s_u and s_v is not vertical; together with 8.2, this proves the if part.

For the only-if part, fix a chart $s\colon(u,v)\mapsto(x(u,v),y(u,v),z(u,v))$, and apply the inverse function theorem to the map $(u,v)\mapsto(x(u,v),y(u,v))$.

8.5. Choose (x,y,z)-coordinates so that Π is the (x,y)-plane and p is the origin. Let $(u,v)\mapsto s(u,v)$ be a small chart of Σ such that $p=s(0,0)$. Denote by κ the unit vector in the direction of the z axis.

Show that we can assume that $\langle s(u,v),\kappa\rangle$ has constant sign in a punctured neighborhood of 0. Conclude that $s_u\perp\kappa$ and $s_v\perp\kappa$ at 0, hence the result.

8.8. By 8.3, $\nu=\frac{\nabla h}{|\nabla h|}$ defines a unit normal field on Σ.

8.10. Use cutoffs and mollifiers from the analysis section (0E) to construct a smooth nonnegative function f on the (x,y)-plane such that $f(x,y)=0$ if and only if $(x,y)\in A$. Observe that the graph $z=f(x,y)$ describes the required surface.

9.1. Fix a point p on γ. Since Σ is mirror symmetric with respect to Π, so is the tangent plane T_p.

Choose (x, y)-coordinates on T_p so that the x-axis is the intersection $\Pi \cap \mathrm{T}_p$. Suppose that the osculating paraboloid is described by the graph $z = \frac{1}{2}\cdot(\ell\cdot x^2 + 2\cdot m\cdot x\cdot y + n\cdot y^2)$. Since Σ is mirror symmetric, so is the paraboloid; that is, changing y to $(-y)$ does not change the value $\ell\cdot x^2 + 2\cdot m\cdot x\cdot y + n\cdot y^2$. In other words $m = 0$, or equivalently, the x-axis points in a direction of curvature.

9.2. Note that the principal curvatures have the same sign at each point. Therefore we can choose a unit normal ν at each point so that both principal curvatures are positive. Show that it defines a global field on the surface.

9.4. Apply 9.3 to a map s such that $s_u(0,0) = \mathrm{u}$ and $s_v(0,0) = \mathrm{v}$.

9.7; *(a).* Observe that Σ has unit Hessian matrix at each point and apply the definition of shape operator.

(b). Choose a chart s in Σ. Show that $\frac{\partial}{\partial u}(s + \nu) = \frac{\partial}{\partial v}(s + \nu) = 0$. Make a conclusion.

9.8. We can assume that γ is parametrized by arc-length. Denote by $\nu_1(s)$ and $\nu_2(s)$ the unit normal vectors to Σ_1 and Σ_2 at $\gamma(s)$. Since γ is a curvature line in Σ_1, ν_1' is proportional to γ'; in particular $\langle \nu_1', \nu_2 \rangle = 0$.

By the assumption, $\langle \nu_1(t), \nu_2(t) \rangle$ is constant. By taking its derivative and applying the above identity, show that $\langle \nu_1, \nu_2' \rangle = 0$. Conclude that ν_2' is proportional to γ' and therefore γ is a curvature line in Σ_2.

9.9; *(a).* Fix t. Set $f_t \colon p \mapsto p + t\cdot\nu(p)$; it maps Σ to Σ_t.

Apply the definition of shape operator to show that $d_p f_t(\mathrm{v}) = \mathrm{v} - t\cdot \mathrm{Shape}_p(\mathrm{v})$. Since Σ is closed, the norm of Shape is bounded. Whence f_t is regular for t sufficiently close to 0.

(b). By the area formula (8.7), we have

$$a(t) = \int\limits_{p \in \Sigma} \mathrm{jac}_p\, f_t.$$

Show and use that for fixed p, the function $t \mapsto \mathrm{jac}_p\, f_t$ has derivative $-H(p)$ at zero.

9.12. Apply 9.11 and the definition of mean curvature.

9.14. Use Meusnier's theorem (9.13), to find the center and radius of curvature of γ in terms of its normal curvature at p; make a conclusion.

9.15. Use 9.1 and Meusnier's theorem (9.13).

9.16. Use 9.15.

9.17. Apply Meusnier's theorem (9.13) to show that the coordinate curves $\alpha_v \colon u \mapsto s(u, v)$ and $\beta_u \colon v \mapsto s(u, v)$ are asymptotic; that is, they have vanishing normal curvature.

Observe that these two families are orthogonal to each other. Therefore the Hessian matrix in the frame $s_u/|s_u|$ and $s_v/|s_v|$ will have zeros on the diagonal. Use the fact that the mean curvature is the trace of the Hessian matrix.

9.18. Use 9.1 and 6.13.

9.20. Drill an extra hole or combine two examples together.

9.21. Denote by L the *cut locus* of V; that is, L is a closure of the set of points $x \in V$ such that there are at least two points on ∂V that minimize the distance to x.

Choose a connected component Σ of the boundary ∂V. Show that L is a smooth surface and the closest point projection $L \to \Sigma$ is a smooth regular parametrization of Σ. In particular there is a unique point on Σ that minimizes the distance to a given point $x \in L$. It follows that there is another component Σ' with the same property.

Finally show that ∂V cannot be more than two components.

9.22. Read about Bing's two-room house. Try to thicken it to construct the needed example.

Assume V does not contain a ball of radius r_3. Show that its cut locus L is formed by a few smooth surfaces that meet by three along their boundary curves. Use this description to show that L is not *simply-connected*; that is, there is a loop in L that cannot be deformed continuously to a trivial loop. Conclude that V is not simply-connected.

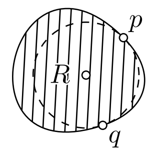
Finally, arrive at a contradiction by showing that if V is bounded by a smooth sphere, then V is simply-connected.

10.2. Choose curvatures such that $k_2(p)_{\Sigma_1} > k_2(p)_{\Sigma_2} > k_1(p)_{\Sigma_1} > k_1(p)_{\Sigma_2}$ and suppose that the first principal direction of Σ_1 coincides with the second principal direction of Σ_2 and vice-versa.

10.4 + 10.5. Apply the same reasoning as in the problems 6.6–6.8, but use families of spheres instead.

10.6. Show and use that any tangent plane T_p supports Σ at p.

10.7. Assume a maximal ball in V has radius r and touches the boundary of R at the points p and q. Consider the projection R of V to a plane thru p, q and the center of the ball. Show that R is bounded by a smooth closed convex curve with curvature at most 1. Argue as in 6.6 to show that $r \geqslant 1$.



10.9. Suppose that a point p lies in the intersection $\Pi \cap \Sigma$.

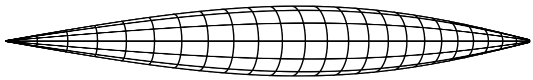
Show that if the tangent plane $\mathrm{T}_p\Sigma$ is parallel to Π, then p is an isolated point of the intersection $\Pi \cap \Sigma$.

It follows that if γ is a connected component of the intersection $\Pi \cap \Sigma$ that is not an isolated point, then Π *intersects* Σ *transversally* along γ; that is, at each point $p \in \gamma$ the tangent plane $\mathrm{T}_p\Sigma$ is not parallel to Π. Apply the implicit function theorem to show that γ is a smooth regular curve.

Finally, observe that the curvature of γ cannot be smaller than the normal curvature of Σ in the same direction. Whence γ has no points with vanishing curvature, and therefore its signed curvature has constant sign.

10.11. Look for a supporting spherical dome with the unit circle as the boundary.

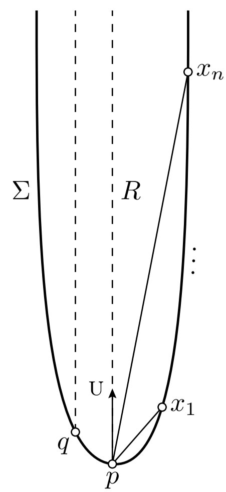
10.12. Note that we can assume that the surface has positive Gauss curvature, otherwise the statement is evident. By 10.10, the surface bounds a convex region that contains a line segment of length π.



Use 3.6 and 9.13 to show that the Gauss curvature of the surface of revolution of the graph $y = a \cdot \sin x$ for $x \in (0, \pi)$ cannot exceed 1. Try to support the surface Σ from inside by a surface of revolution of the described type. (Compare to 6.11.)

Remark. The exercise can be reduced to the following deeper result: *if the Gauss curvature of Σ is at least 1, then the intrinsic diameter of Σ cannot exceed π.* The latter means that any two points in Σ can be connected by a path that lies in Σ and has length at most π. This theorem was proved by Heinz Hopf and Willi Rinow [40] and named after Sumner Myers who generalized it [62].

10.14. To prove *(a)* use the convexity of R. To prove *(b)* observe that the map $\Sigma \to \mathbb{S}^2$ is smooth and regular, then apply the inverse function theorem to show that its inverse is smooth as well.



10.16; *(a)* We can assume that the origin lies on Σ. Consider a sequence of points $x_n \in \Sigma$ such that $|x_n| \to \infty$ as $n \to \infty$. Denote by u_n the unit vector in the direction of x_n; that is $\mathrm{u}_n = \frac{x_n}{|x_n|}$.

Since the unit sphere is compact, we can pass to a subsequence of (x_n) such that u_n converges to a unit vector u. Show that the half-line from the origin in the direction of u can be taken as ℓ.

(b) + (c) + (d). Since R is convex, so is its projection Ω.

Note and use that for any $q \in \Sigma$, the directions $\mathrm{v}_n = \frac{x_n - q}{|x_n - q|}$ converge to u as well.

Show that Σ has no vertical tangent planes. Conclude that the projection map from Σ to the (x, y)-plane is regular. Use the inverse function theorem to show that Ω is open.

(e). Arguing by contradiction, suppose that for some sequence $(x_n, y_n) \to (x_\infty, y_\infty) \in \partial\Omega$ the sequence $f(x_n, y_n)$ stays bounded above. We can pass to a subsequence such that either $f(x_n, y_n)$ converges to some finite value z_∞ or it diverges to $-\infty$.

In the first case, show that the point $(x_\infty, y_\infty, z_\infty)$ does not lie on Σ, but it has arbitrarily close points on Σ. That is Σ is not proper — a contradiction.

If $f(x_n, y_n) \to -\infty$, use the convexity of f to show that $f(\frac{x_n}{2}, \frac{y_n}{2}) \to -\infty$. Note that origin belongs to Ω; use it to show that $(\frac{x_\infty}{2}, \frac{y_\infty}{2}) \in \Omega$. Arrive to a contradiction.

10.17. By 10.16d, Σ is parametrized by an open convex plane domain Ω. It remains to show that Ω can parameterize the whole plane.

We may assume that the origin of the plane lies in Ω. Show that in this case the boundary of Ω can be written in polar coordinates as $(\theta, f(\theta))$ where $f\colon \mathbb{S}^1 \to \mathbb{R}$ is a positive continuous function. Then homeomorphism to the plane can be described in polar coordinates by changing only the radial coordinate; for example as $(\theta, r) \mapsto (\theta, \frac{r}{1 - r/f(\theta)})$.

To do the second part, one may apply 7.8c.

10.18. Choose a coordinate system so that the (x, y)-plane supports Σ at the origin, so Σ lies in the upper half-space.

Show that there is $\varepsilon > 0$ such that any line starting from the origin with slope at most ε may intersect Σ only in the unit ball centered at the origin; we may assume that ε is small, say $\varepsilon < 1$. Consider the cone formed by the half-lines from the origin with slope ε shifted down by 1 and observe that the entire surface lies inside this cone.

10.19. Choose distinct points $p, q \in \Sigma$. Apply 10.10 to show that the angle $\measuredangle(\nu(p), p - q)$ is acute and $\measuredangle(\nu(q), p - q)$ is obtuse. Conclude that $\nu(p) \neq \nu(q)$; that is, $\nu\colon \Sigma \to \mathbb{S}^2$ is injective.

(a). Given a unit vector u, consider a point $p \in \Sigma$ that maximizes the scalar product $\langle p, \mathrm{u}\rangle$. Show that $\nu(p) = \mathrm{u}$. Conclude that the spherical map $\nu\colon \Sigma \to \mathbb{S}^2$ is onto, and therefore it is a bijection.

Applying 9.6, we get that the integral is $4{\cdot}\pi = \operatorname{area}\mathbb{S}^2$.

(b). Choose an (x, y, z)-coordinate system provided by 10.16d. Observe that for any p the normal vector $\nu(p)$ forms an obtuse angle with the z-axis. It follows that the image $\nu(\Sigma)$ lies in the south hemisphere.

Applying 9.6, we get that the integral is $2{\cdot}\pi = \frac{1}{2}\cdot \operatorname{area}\mathbb{S}^2$.

10.20. Apply 9.15.

10.21. Prove and use that each point $p \in \Sigma$ has a direction with vanishing normal curvature.

10.22. Suppose $p \in \Sigma$ is a point of strict local maximum of f. Arrive to a contradiction by showing that Σ is supported by its tangent plane at p.

10.23. Denote by Π_t the tangent plane to Σ at $\gamma(t)$ and by ℓ_t the tangent line to γ at time t.

Note that Π_t is the graph of a linear function, say h_t, defined on the (x, y)-plane. Denote by $\bar\ell_t$ the projection of ℓ_t to the (x, y)-plane. Show that the derivative $\frac{d}{dt}h_t(w)$ vanishes at the point w if and only if $w \in \bar\ell_t$ and the derivative changes sign if w goes from one side of $\bar\ell_t$ to the other.

Denote by $\bar\gamma$ the projection of γ to the (x, y)-plane. If $\bar\gamma$ is star shaped with respect to a point w, then w cannot cross $\bar\ell_t$. Therefore the function $t \mapsto h_t(w)$ is monotonic on $\mathbb{S}^1$. Observe that this function cannot be constant, and arrive at a contradiction.

10.24. This problem follows easily from the so-called *Morse lemma.* The following sketch is a slightly stripped version of it. A more conceptual proof can be built on the Moser's trick [65].

Choose tangent-normal coordinates at p so that the coordinate axes point in the principal directions of Σ at p; let $z = f(x, y)$ be the local graph representation of Σ. We need to show that the solution of the equation $f(x, y) = 0$ is a union of two smooth curves that intersect transversely at p.

Prove the following claim:

◇ Suppose that $x \mapsto h(x)$ is a smooth function defined in an open interval $\mathbb{I} \ni 0$ such that $h(0) = h'(0) = 0$ and $h''(0) > 0$. Then, for a smaller interval $\mathbb{J} \ni 0$ there is a unique smooth function $a\colon \mathbb{J} \to \mathbb{R}$ such that $h = a^2$, $a(0) = 0$ and $a'(0) > 0$.

Note that by passing to a small rectangular domain $|x|, |y| < \varepsilon$, we can assume that $f_{xx} > \varepsilon$ and $f_{yy} < -\varepsilon$. Show that if ε is small, then for every x there is unique $y(x)$ such that $f_x(x, y(x)) = 0$; moreover the function $x \mapsto y(x)$ is smooth.

Set $h(x) = f(x, y(x))$. Note that $h(0) = h'(0) = 0$ and $h'' > 0$. Applying the claim, we get a function a such that $h = a^2$, where $a(0) = 0$ and $a'(0) > 0$.

Observe that $g(x, y) = h(x) - f(x, y) \geqslant 0$ and $g_y(x, y(x)) = g(x, y(x)) = 0$ and $g_{yy} > 0$. Applying the claim to each function $y \mapsto g(x, y)$ with fixed x, we get that $g(x, y) = b(x, y)^2$ for a smooth function b such that $b(x, y(x)) = 0$ and $b_y(x, y(x)) > 0$.

It follows that

$$f(x, y) = a(x)^2 - b(x, y)^2 = (a(x) - b(x, y))\cdot(a(x) + b(x, y)).$$

That is $f(x, y) = 0$ if $a(x) \pm b(x, y) = 0$.

It remains to observe that the two functions $g_\pm(x, y) = a(x) \pm b(x, y)$ have distinct non-zero gradients at 0. Therefore each equation $a(x) \pm b(x, y) = 0$ defines a smooth regular curve in a neighborhood of p; see Section 1E.

10.26. Use 10.25 and the hemisphere lemma (2.19).

10.27. Assume Σ is an open saddle surface that lies in a cone K. Show that there is a plane Π that cuts Σ and cuts from K a compact region. Conclude that Π cuts from Σ a compact region as well.

By 8.11 one can move the plane Π slightly so that it cuts from Σ a compact surface with boundary. Apply 10.25.

10.29. Consider the radial projection of F_ε to the sphere Σ with center at $p = (0, 0, \varepsilon)$; that is, a point $q \in F_\varepsilon$ is mapped to a point $s(q)$ on the sphere that lies on the ray pq.

Show that s is a diffeomorphism from F_ε to the south hemisphere of Σ. It remains to observe that the unit disc is diffeomorphic to the hemisphere.

10.30. Apply 10.28.

10.32. Find an example among the surfaces of revolution. Use 9.15 in the proof.

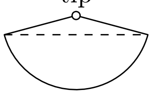
10.33. Look at the sections of the graph by planes parallel to the (x, y)-plane and to the (x, z)-plane, then apply Meusnier's theorem (9.13).

10.34. Suppose that orthogonal projection of Σ to the (x, y)-plane is not injective. Show that there is a point $p \in \Sigma$ with a vertical tangent plane; that is, $\mathrm{T}_p\Sigma$ is perpendicular to the (x, y)-plane.

Let Γ_p be the connected component of p in the intersection of Σ and the affine tangent plane $\mathrm{T}_p\Sigma$. Use 10.24 to show that Γ_p is a union of smooth regular curves that can cross each other transversely. Moreover two of these curves pass thru p and Γ_p does not bound a compact region on Σ.

Show that Γ_p must have at least 4 ways to escape to infinity. On the other hand, since Σ is a graph outside of compact set K, we have that $\Gamma_p \backslash K$ is a graph of a real-to-real function that has only two ways to escape to infinity — a contradiction.

11.1. Cut the lateral surface of the ice-mountain by a line from the cowboy to the tip; unfold it on the plane (see the picture) and try to figure out what is the image of the strained lasso.



11.6. Note that by 10.10, Σ bounds a strictly convex region. Therefore we can assume that $\nu(p) \neq \nu(q)$, otherwise $p = q$ and the inequality is evident.

Further, we can assume that $\nu(p)+\nu(q) \neq 0$, otherwise the right hand side is undefined.

In the remaining case the tangent planes T_p and T_q intersect along a line, say ℓ. Set $\alpha = \frac{1}{2}\cdot\measuredangle(\nu(p), \nu(q))$; observe that $2\cdot \cos\alpha = |\nu(p) + \nu(q)|$. Let $x \in \ell$ be the point that minimizes the sum $|p - x| + |x - q|$. Show that $\measuredangle[x\,{}^p_q] \geqslant \pi - 2\cdot\alpha$. Conclude that

$$|p - x| + |x - q| \leqslant \frac{|p-q|}{\cos\alpha}.$$

Finally, apply 11.5 to show that

$$|p - q|_\Sigma \leqslant |p - x| + |x - q|.$$

11.7. Suppose there is a minimizing curve $\gamma \not\subset \Delta$ with endpoints p and q in Δ.

Without loss of generality, we may assume that only one arc of γ lies outside of Δ. Reflection of this arc with respect to Π together with the remaining part of γ forms another curve $\hat\gamma$ from p to q; it runs partly along Σ and partly outside Σ, but does not get inside Σ. Note that

$$\operatorname{length}\hat\gamma = \operatorname{length}\gamma.$$

Denote by $\bar\gamma$ the closest point projection of $\hat\gamma$ on Σ. Note that the curve $\bar\gamma$ lies in Σ, it has the same ends as γ, and by 11.5

$$\operatorname{length}\bar\gamma < \operatorname{length}\gamma.$$

This means that γ is not length minimizing, a contradiction.

11.8. Show that the closest point projection $\partial B \to \Sigma$ is surjective and apply 11.4.

12.2. We can assume that the origin lies on the axis of revolution and I points in the direction of such axis. Use 12.1 to show that it is sufficient to prove that $\langle\gamma' \times \gamma, \mathrm{I}\rangle$ is constant.

Since $\gamma''(t) \perp \mathrm{T}_{\gamma(t)}$, the three vectors I, γ, and γ'' lie in the same plane. In particular $\langle\gamma'' \times \gamma, \mathrm{I}\rangle = 0$. Therefore $\langle\gamma' \times \gamma, \mathrm{I}\rangle' = \langle\gamma' \times \gamma', \mathrm{I}\rangle + \langle\gamma'' \times \gamma, \mathrm{I}\rangle = 0$.

12.3. By Lemma 12.1, we can assume that γ is parametrized by arc-length. By the definition of geodesic, we have that $\gamma''(s) \perp \mathrm{T}_{\gamma(s)}$. Therefore

$$\gamma''(s) = k_n(s)\cdot\nu(\gamma(s)),$$

where $k_n(s)$ is the normal curvature of γ at time s. Since γ is asymptotic, $k_n(s) \equiv 0$; that is, $\gamma''(s) \equiv 0$, therefore γ' is constant and γ runs along a line segment; see 3.2.

12.4. Denote by μ a unit vector perpendicular to Π. Since γ lies in Π, we have that γ'' is parallel to Π, or equivalently $\gamma'' \perp \mu$. Since γ is unit speed, 3.1 implies that $\gamma'' \perp \gamma'$.

Since Σ is mirror symmetric with respect to the plane Π, the tangent plane $\mathrm{T}_{\gamma(t)}\Sigma$ is also mirror symmetric with respect to Π as well. It follows that $\mathrm{T}_{\gamma(t)}\Sigma$ is spanned by μ and $\gamma'(t)$. Hence $\gamma'' \perp \mu$ and $\gamma'' \perp \gamma'$ imply $\gamma'' \perp \mathrm{T}_{\gamma(t)}\Sigma$; that is, γ is a geodesic.

12.7. By 12.4, any meridian of Σ is a closed geodesic. Consider an arbitrary closed geodesic γ.

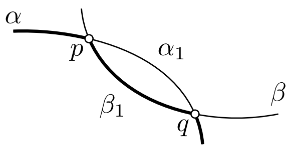
In the case when γ is tangent to a meridian at some point, then by the uniqueness part of Proposition 12.5, γ runs along that meridian; in particular it is non-contractible.

In the remaining case, γ can intersect meridians only transversely. Therefore the longitude of γ is monotonic. Whence again γ is non-contractible.

12.12. Show that $\nu(\gamma(t)) = (\cos t, \sin t, 0)$. Calculate $\gamma''(t)$ and show that it is proportional to $\nu(\gamma(t))$. The latter is equivalent to $\gamma''(t) \perp \mathrm{T}_{\gamma(t)}$. Note that the line segment from $\gamma(0)$ to $\gamma(2\pi)$ is contained in Σ.

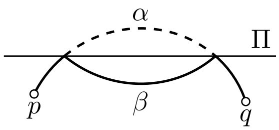
16I. Assume that two shortest paths α and β have two common points p and q. Denote by α_1 and β_1 the arcs of α and β from p to q. Suppose that α_1 is distinct from β_1.

Note that α_1 and β_1 are shortest paths with the same endpoints; in particular they have the same length. Exchanging α_1 in α to β_1 produces a shortest path, say $\hat{\alpha}$, that is distinct from α. By 12.11, $\hat{\alpha}$ is a geodesic.



Suppose α_1 is a proper subarc of α; that is, $\alpha_1 \neq \alpha$, or equivalently, either p or q is not an endpoint of α. Then α and $\hat{\alpha}$ share one point and velocity vector at this point. By 12.5 α coincides with $\hat{\alpha}$ — a contradiction.

It follows that p and q are the endpoints of α. Analogously, p and q are the endpoints of β.



For the second part, one could consider two distinct curves of the form

$$\gamma_b(t) = (\cos t, \sin t, bt), t \in \mathbb{R}$$

in the cylinder $x^2 + y^2 = 1$.

12.15. Assume a shortest path α crosses Π at least twice. In this case there is an arc α_1 of α that lies entirely on one side on Π, has ends on Π, and these ends are distinct from the ends of α.

Let us remove the arc α_1 from α and replace it with the reflection of α_1 across Π. Note that the obtained curve, say β, lies in the surface; it has the same length as α, and it connects the same pair of points, say p and q. Therefore β is another shortest path from p to q in Σ that is distinct from α.

By 12.11, α and β are geodesics. Since α and β have a common subarc, they share one point and velocity vector at this point; by 12.5 α coincides with β — a contradiction.

12.16. Let W be the closed region outside of Σ. Show that the distance $|p^t - q|_W$ is constant with respect to t. This implies that the concatenation of the line segment $[p^t, \gamma(t)]$ and the arc $\gamma|_{[t,\ell]}$ is a shortest path from p^t to q in W. Since Σ is strictly convex,

$$|\gamma(t) - q|_W > |\gamma(t) - q|_{\mathbb{R}^3}$$

for all $t < \ell$. Hence

$$\begin{aligned}|p^t - q|_W &= |p^t - \gamma(t)|_W + |\gamma(t) - q|_W\\ &> |p^t - \gamma(t)|_{\mathbb{R}^3} + |\gamma(t) - q|_{\mathbb{R}^3}\\ &\geqslant |p^t - q|_{\mathbb{R}^3}.\end{aligned}$$

If the segment $[p^t, q]$ was entirely contained in W for some $t < \ell$, then $|p^t - q|_W = |p^t - q|_{\mathbb{R}^3}$, which would be a contradiction.

12.18. Equip Σ with unit normal field ν that points inside. Denote by $k(t)$ the normal curvature of Σ at $\gamma(t)$ in the direction of $\gamma'(t)$. Since Σ is convex, $k(t) \geqslant 0$ for any t.

Since γ is a geodesic, we have $\gamma''(t) = k(t)\cdot\nu(\gamma(t))$.

Since γ has unit speed, $\langle\gamma'(t), \gamma'(t)\rangle = 1$ for any t.

Without loss of generality, we can assume that p is the origin of $\mathbb{R}^3$. Since p is inside Σ, we have that $\langle\gamma(t), \nu(\gamma(t))\rangle \leqslant 0$ for any t. It follows that

$$\langle\gamma''(t), \gamma(t)\rangle = k(t)\cdot\langle\gamma(t), \nu(\gamma(t))\rangle \leqslant 0$$

for any t.

Applying the above estimates, we get

$$\rho''(t) = \langle\gamma(t), \gamma(t)\rangle'' = 2\cdot\langle\gamma''(t), \gamma(t)\rangle + 2\cdot\langle\gamma'(t), \gamma'(t)\rangle \leqslant 2.$$

12.19. We can assume γ is parametrized by arc-length. If we write $\nu(t) = \nu(\gamma(t))$, then $\langle\gamma'(t), \nu(t)\rangle \equiv 0$. Also, since γ is a geodesic, we have that $\gamma''(t) \parallel \nu(t)$. Then

$$\begin{aligned}|\gamma''(t)| &= |\langle\gamma''(t), \nu(t)\rangle|\\ &= |\langle\gamma'(t), \nu(t)\rangle' - \langle\gamma'(t), \nu'(t)\rangle|\\ &= |\langle\gamma'(t), \nu'(t)\rangle|\\ &\leqslant |\gamma'(t)||\nu'(t)|\\ &= |\nu'(t)|.\end{aligned}$$

Integrating both sides with respect to t, we get

$$\operatorname{length}(\nu\circ\gamma) \geqslant \Phi(\gamma).$$

12.21. Suppose $\gamma(t) = (x(t), y(t), z(t))$. Show that

❶ $$|\gamma''(t)| = z''(t)\cdot\sqrt{1 + \ell^2}$$

for any t.

Observe that $z'(t) \to \pm\frac{\ell}{\sqrt{1+\ell^2}}$ as $t \to \pm\infty$. Conclude that

❷ $$\int\limits_{-\infty}^{+\infty} z''(t)\cdot dt = \frac{2\cdot\ell}{\sqrt{1 + \ell^2}}.$$

By ❶ and ❷, we have

$$\Phi(\gamma) = \int\limits_{-\infty}^{+\infty} |\gamma''(t)|\cdot dt = \sqrt{1+\ell^2}\cdot \int\limits_{-\infty}^{+\infty} z''(t)\cdot dt = 2\cdot\ell.$$

12.22. Use 12.20 and 3.21. For the second part, consider a geodesic on a cone with slope 2 and mollify its tip.

Remark. The statement still holds for $\sqrt{3}$-Lipschitz functions, and $\sqrt{3} = \operatorname{tg}\frac{\pi}{3}$ is the optimal bound; see 14.6. It is the same slope as in the problem about the cowboy and the lasso (11.1).

13.1; *(a)*. Show and use that $\langle \mathrm{v}(t), \mathrm{v}'(t)\rangle = 0$.

(b) Show that $|\mathrm{v}(t)|$, $|\mathrm{w}(t)|$, and $\langle \mathrm{v}(t), \mathrm{w}(t)\rangle = 0$, are constants; it can be done the same way as *(a)*. Then use the formula $\langle \mathrm{v}(t), \mathrm{w}(t)\rangle = |\mathrm{v}(t)|\cdot|\mathrm{w}(t)|\cdot\cos\theta$.

13.3. Observe that Σ_1 supports Σ_2 at any point of γ. Conclude that γ has identical spherical images as a curve in Σ_1 and in Σ_2. Apply 13.2.

13.4. Consider the right angled spherical triangle that an octant of $\mathbb{R}^3$ cuts form the sphere. Argue that parallel transport around it rotates the tangent plane by the angle $\frac{\pi}{2}$.

14.2. By 10.14, Σ is a smooth sphere. By Jordan theorem (0.28) the curve γ divides Σ in two discs. Let us denote by Δ the disc that lies on the left from γ.

Observe that $\Psi(\gamma) = \operatorname{length}\gamma$ and apply the Gauss–Bonnet formula (14.1) for Δ.

14.3. Apply 9.6, 10.19a, and the Gauss–Bonnet formula (14.1).

For the last part, apply 2.21.

14.4. For the first part apply the Gauss–Bonnet formula (14.1).

For the second part, arguing by contradiction, assume that two closed geodesics γ_1 and γ_2 do not intersect. This would imply that γ_2 lies in one of the regions, say R_1, that γ_1 cuts from Σ. Similarly γ_1 lies in one of the regions, say R_2, that γ_2 cuts from Σ.

Observe that R_1 and R_2 cover Σ with overlap. Therefore the first part implies that the integral of the Gauss curvature over Σ is less than $4\cdot\pi$. The latter contradicts 10.19a.

14.5; *(easy)*. Consider the 4 regions bounded by loops. Apply Gauss–Bonnet formula (14.1) to show that the integral of the Gauss curvature on each of these regions exceeds π. The latter contradicts 10.19a.

(tricky). Denote by α, β, and γ the angles of the triangle. Apply the Gauss–Bonnet formula (14.1) to show that the loops surround regions over which the integral of the Gauss curvature is $\pi+\alpha$, $\pi+\beta$, and $\pi+\gamma$ respectively.

Apply the Gauss–Bonnet formula to the triangular region to show that $\alpha+\beta+\gamma>\pi$. Use 10.19a to get at a contradiction.

14.6. Note that it is sufficient to show that the surface has no geodesic loops. Estimate the integral of the Gauss curvature over the entire surface and over a disc in it surrounded by a geodesic loop.

14.7; *(a).* By 11.3 and 12.11, any two points in Σ can be connected by a geodesic. Suppose that points p and q can be connected by two distinct geodesics γ_1 and γ_2. By Exercise , γ_1 and γ_2 share only their endpoints. Since the surface is simply-connected, γ_1 and γ_2 together bound a disc, say $\Delta \subset \Sigma$. It remains to apply Gauss–Bonnet formula to Δ and make a conclusion.

(b). Use part *(a)* and 12.10.

14.10. Apply 13.5 and 14.9.

14.12. Repeat the end of the proof of 14.11 for a one-parameter family of geodesics γ_τ defined by $\gamma_\tau(0) = \alpha(\tau)$ and $\gamma'(0) = \alpha'(\tau)$.

15.4. By the Gauss lemma (15.2), polar coordinates with respect to q produce a semi-geodesic chart at any nearby point. Therefore, it is sufficient to find a point $q \neq p$ such that polar coordinates on Σ with respect to q cover p. By 12.9, any q sufficiently close to p does the job.

15.5; *(a).* Show that we can choose an orientation on Σ so that $b_r(0, \theta) = 1$ for any θ. Conclude that we can assume that $b(r, \theta) > 0$ for small positive values r.

Suppose $b(r_1, \theta_1) < 0$ at some pair (r_1, θ_1) with $0 < r_1 < r_0$. Observe that if θ_2 is sufficiency close to θ_1 then the radial curves $r \mapsto b(r, \theta_1)$ and $r \mapsto b(r, \theta_2)$ defined on the interval $(0, r_1)$ intersect. Therefore $\exp_p$ is not injective in B — a contradiction.

(b). Suppose that $b(r_1, \theta_1) = 0$. Apply *(a)* to show that $b_r(r_1, \theta_1) = 0$.

Apply 15.3 to conclude $b(r, \theta_1) = 0$ for any r. The latter contradicts that $b_r(0, \theta_1) = 1$.

(c). We need to show that $\exp_p$ is regular in B. Suppose that vector $\mathrm{v} \in B$ has polar coordinates (r, θ) for some $r > 0$. Show that $\exp_p$ is regular at v if $b(r, \theta) \neq 0$. Conclude that $\exp_p$ is regular in $B \backslash \{0\}$.

By 12.8, $\exp_p$ is regular at 0. Whence $\exp_p|_B$ is a regular injective smooth map. Use the inverse function theorem (0.21) to show that the restriction $\exp_p|_B$ is a diffeomorphism to its image.

15.6; *(a).* Since the frame $\mathrm{u}, \mathrm{v}, \nu$ is orthonormal, the first two vector identities are equivalent to the following six real identities:

❶
$$\begin{aligned} &\langle \mathrm{u}_u, \mathrm{u} \rangle = 0, \quad \langle \mathrm{u}_u, \mathrm{v} \rangle = -\tfrac{1}{b} \cdot a_v, \quad \langle \mathrm{u}_u, \nu \rangle = a \cdot \ell, \\ &\langle \mathrm{v}_u, \mathrm{v} \rangle = 0, \quad \langle \mathrm{v}_u, \mathrm{u} \rangle = \tfrac{1}{b} \cdot a_v, \qquad \langle \mathrm{v}_u, \nu \rangle = a \cdot m. \end{aligned}$$

Taking the partial derivatives of the identities $\langle \mathrm{u}, \mathrm{u} \rangle = 1$ and $\langle \mathrm{v}, \mathrm{v} \rangle = 1$ with respect to u we get the first two identities in ❶.

Further, observe that

❷
$$\mathrm{v}_u = \tfrac{\partial}{\partial v}(\tfrac{1}{b} \cdot s_v) = \tfrac{1}{b} \cdot s_{uv} + \tfrac{\partial}{\partial u}(\tfrac{1}{b}) \cdot s_v.$$

Since $s_u \perp s_v$, it follows that

$$\langle \mathrm{v}_u, \mathrm{u} \rangle = \tfrac{1}{a \cdot b} \cdot \langle s_{vu}, s_u \rangle = \tfrac{1}{2 \cdot a \cdot b} \cdot \tfrac{\partial}{\partial v} \langle s_u, s_u \rangle = \tfrac{1}{2 \cdot a \cdot b} \cdot \tfrac{\partial a^2}{\partial v} = \tfrac{1}{b} \cdot a_v.$$

Taking the partial derivative of $\langle \mathrm{u}, \mathrm{v} \rangle = 0$ with respect to u we get

$$\langle \mathrm{v}_u, \mathrm{u} \rangle + \langle \mathrm{v}, \mathrm{u}_u \rangle = 0.$$

Hence we get two more identities in ❶.

Since u, v is an orthonormal frame, by 9.3 we have

❸ $$\begin{aligned} \tfrac{1}{a^2}\cdot\langle s_{uu},\nu\rangle &= \ell, & \tfrac{1}{a\cdot b}\cdot\langle s_{uv},\nu\rangle &= m,\\ \tfrac{1}{a\cdot b}\cdot\langle s_{vu},\nu\rangle &= m, & \tfrac{1}{b^2}\cdot\langle s_{vv},\nu\rangle &= n. \end{aligned}$$

Applying ❷, ❸, and $s_v \perp \nu$ we get

$$\langle \mathrm{u}_u,\nu\rangle = \tfrac{1}{a}\cdot\langle s_{uu},\nu\rangle = a\cdot\ell, \qquad \langle \mathrm{v}_u,\nu\rangle = \tfrac{1}{a}\cdot\langle s_{uv},\nu\rangle = a\cdot m,$$

that imply the last two equalities in ❶.

Therefore the first two identities in *(a)* are proved; the remaining two identities can be proved along the same lines.

(b). Recall that the Gauss curvature equals the determinant of the matrix $\left(\begin{smallmatrix}\ell & m\\ m & n\end{smallmatrix}\right)$; that is, $K = \ell\cdot n - m^2$. Therefore

$$\begin{aligned} a\cdot b\cdot K &= a\cdot b\cdot(\ell\cdot n - m^2) = \langle \mathrm{u}_u, \mathrm{v}_v\rangle - \langle \mathrm{u}_v, \mathrm{v}_u\rangle\\ &= \left(\tfrac{\partial}{\partial v}\langle \mathrm{u}_u,\mathrm{v}\rangle - \cancel{\langle \mathrm{u}_{uv},\mathrm{v}\rangle}\right) - \left(\tfrac{\partial}{\partial u}\langle \mathrm{u}_v,\mathrm{v}\rangle - \cancel{\langle \mathrm{u}_{uv},\mathrm{v}\rangle}\right) = \tfrac{\partial}{\partial v}\left(-\tfrac{1}{b}\cdot a_v\right) - \tfrac{\partial}{\partial u}\left(\tfrac{1}{a}\cdot b_u\right). \end{aligned}$$

15.7. Apply 15.6 assuming that $a = b$ and simplify.

15.11. Note and use that by 15.10, $\exp_p$ is length-preserving.

15.12. Modify the proof of 15.10 to show that $K \equiv 1$ implies that $b(\theta, r) = \sin r$ for all small $r \geqslant 0$.

16.1. Apply the triangle inequality $c_n \leqslant a_n + b_n$ and the bound $a_n, b_n > \varepsilon$ to show that the sequence

$$\frac{a_n + b_n + c_n}{2\cdot a_n\cdot b_n}$$

is bounded. Conclude that

$$\frac{(a_n+b_n)^2 - c_n^2}{2\cdot a_n\cdot b_n} = \frac{(a_n+b_n+c_n)\cdot(a_n+b_n-c_n)}{2\cdot a_n\cdot b_n} \to 0$$

as $n \to \infty$.

Use the last statement together with the cosine rule

$$a_n^2 + b_n^2 - 2\cdot a_n\cdot b_n\cdot\cos\tilde\theta_n - c_n^2 = 0$$

to show that $\cos\tilde\theta_n \to -1$ as $n \to \infty$; conclude that $\theta_n \to \pi$.

16.3. Observe that $|p - x|_\Sigma \leqslant |p - q|_\Sigma$ and $|q - x|_\Sigma \leqslant |p - q|_\Sigma$. Conclude that $\tilde\measuredangle(x\,{}^p_q) \geqslant \frac{\pi}{3}$. Apply 16.2*b*.

16.4. Observe that $\measuredangle[p\,{}^x_y] + \measuredangle[p\,{}^y_z] + \measuredangle[p\,{}^z_x] \leqslant 2\cdot\pi$. Apply 16.2*b*.

16.8; *(a).* Apply Rauch comparison 15.10 and the property of exponential map in 12.9.

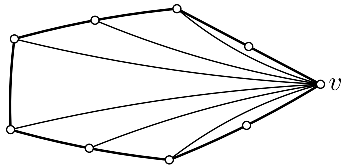
(b). Argue by contradiction; if the statement does not hold then for any $p \in \Sigma$ there is a point $q = q(p) \in \Sigma$ such that $|q - p|_\Sigma < 100\cdot R_p$ and $R_q < (1 - \frac{1}{100})\cdot R_p$.

Start with any point p_0 and consider a sequence p_n defined by $p_{n+1} = q(p_n)$. Show that p_n converges and $R_{p_n} \to 0$ as $n \to \infty$. Arrive to a contradiction using *(a)*.

(c). Repeat the proof assuming that p is provided by *(b)*.

16.9; *(a).* Apply the Gauss–Bonnet formula for each region that γ cuts from Σ and simplify the obtained inequalities.

(b). Consider the pentagon Δ with induced length-metric. Note that all its angles cannot exceed π. Repeat the proof of 16.2*b* to show that the comparison holds in Δ.



Choose a vertex v of Δ, subdivide its sides so that each arc of the subdivision is a minimizing geodesic, so the boundary of Δ is a broken geodesic. Divide Δ into triangles by joining v to other vertices of the broken geodesic. Take a model triangle for each so that they shared sides as in Δ. Use the comparison to show that these plane triangles form a required polygon.

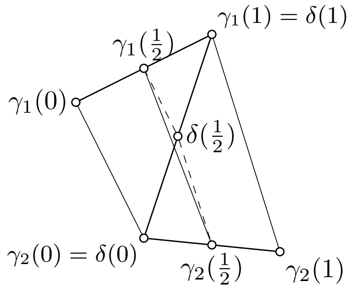
(c). It remains to show that the plane pentagon provided by *(b)* cannot exist. Imagine that there is air pressure in the pentagon, it should not move it; so the total force should vanish. Use part *(a)* to show that this is not the case. (Equivalently, show that the sum of its side vectors oriented counterclocwise does not vanish.)

16.13. Apply 16.11*b* and 16.11*B*.

16.14. Use 16.11*b* or 16.11*B* twice: first for the triangle $[pxy]$ and $\bar{x} \in [p, x]$ and second for the triangle $[p\bar{x}y]$ and $\bar{y} \in [p, y]$. Then apply the angle monotonicity (0.11).

16.15. It is sufficient to prove the Jensen inequality; that is,

$$|\gamma_1(\tfrac{1}{2}) - \gamma_2(\tfrac{1}{2})| \leqslant \tfrac{1}{2}\cdot|\gamma_1(0) - \gamma_2(0)| + \tfrac{1}{2}\cdot|\gamma_1(1) - \gamma_2(1)|.$$



Let δ be the geodesic path from $\gamma_2(0)$ to $\gamma_1(1)$. From 16.14, we have

$$|\gamma_1(\tfrac{1}{2}) - \delta(\tfrac{1}{2})| \leqslant \tfrac{1}{2}\cdot|\gamma_1(0) - \delta(0)|,$$
$$|\delta(\tfrac{1}{2}) - \gamma_2(\tfrac{1}{2})| \leqslant \tfrac{1}{2}\cdot|\delta(1) - \gamma_2(1)|,$$

Sum it up and apply the triangle inequality.

Remark. Note that modulo the comparison theorem, the case of the Euclidean plane is just as hard.

16.16; *(a).* Choose two points $x, z \in \Sigma$; let y be the midpoint of $[x, z]$. Denote by $\bar{x}$ and $\bar{z}$ the points on δ that minimize the distances to x and z respectively; so we have $f(x) = |x - \bar{x}|_\Sigma$, and $f(z) = |z - \bar{z}|_\Sigma$.

Since Δ is convex, the midpoint of $[\bar{x}, \bar{z}]$, say $\bar{y}$, lies in Δ. By 16.15, we have

$$f(y) \leqslant |y - \bar{y}|_\Sigma \leqslant \tfrac{1}{2}\cdot(|x - \bar{x}|_\Sigma + |z - \bar{z}|_\Sigma) = \tfrac{1}{2}\cdot(f(x) + f(z)).$$

Observe that this inequality implies convexity of f.

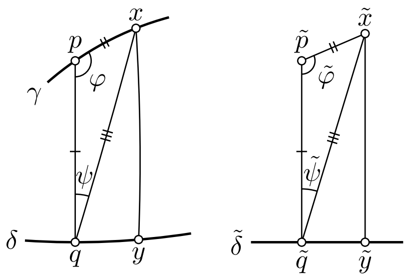
(b). Suppose γ is a unit-speed geodesic in Δ and $\gamma(0) = p$. Let q be a point on δ that minimize the distance to p. Denote by φ the angle between γ and $[p, q]$ at p. Note that it is sufficient to show that

$$f\circ\gamma(t) \leqslant |p - q|_\Sigma + t\cdot\cos\varphi \qquad ❶$$

for t sufficiently close to 0.

Chose small $t > 0$ and set $x = \gamma(t)$. Consider a model triangle $[\tilde p\tilde q\tilde x] = \tilde\triangle(pqx)$. Let $\tilde\delta$ be the line thru $\tilde q$ perpendicular to $[\tilde p, \tilde q]$. Denote by $\tilde y$ the orthogonal projection of $\tilde x$ to $\tilde\delta$. By comparison we have that

$$\begin{aligned}\varphi = \measuredangle[q\,{}^{p}_{x}] \geqslant \tilde\measuredangle(q\,{}^{p}_{x}) =: \tilde\varphi,\\ \psi := \measuredangle[p\,{}^{q}_{x}] \geqslant \tilde\measuredangle(p\,{}^{q}_{x}) =: \tilde\psi\end{aligned}$$



Since $t = |\tilde p - \tilde x|_{\mathbb{R}^2}$ is small, so is the distances $|\tilde q - \tilde y|_{\mathbb{R}^2}$. Therefore we can choose a point y on δ such that $|q - y|_\Sigma = |\tilde q - \tilde y|_{\mathbb{R}^2}$. Moreover, if we choose in on the *right side*, then $\frac{\pi}{2} - \psi \geqslant \measuredangle[q\,{}^{x}_{y}]$. By construction and comparison, we have

$$\measuredangle[\tilde q\,{}^{\tilde x}_{\tilde y}] = \tfrac{\pi}{2} - \tilde\psi \geqslant \tfrac{\pi}{2} - \psi \geqslant \measuredangle[q\,{}^{x}_{y}] \geqslant \tilde\measuredangle(q\,{}^{x}_{y}).$$

Therefore $|x - y|_\Sigma \leqslant |\tilde x - \tilde y|_{\mathbb{R}}$. Since $\tilde\varphi \leqslant \varphi$, we get that

$$\begin{aligned}f\circ\gamma(t) \leqslant |x - y|_\Sigma \leqslant |\tilde x - \tilde y|_{\mathbb{R}} =\\ = |p - q|_\Sigma + t\cdot\cos\tilde\varphi \leqslant |p - q|_\Sigma + t\cdot\cos\varphi.\end{aligned}$$

That is, we proved ❶ for small $t > 0$. The case $t < 0$ can be done along the same lines.

16.18; *(a).* Choose a sequence of points $x_n \in K$ that escapes to infinity; that is, such that $|p - x_n|_\Sigma \to \infty$ as $n \to \infty$. Denote by $\mathrm{u}_n \in \mathrm{T}_p$ the unit vector in the direction of $[p, x_n]_\Sigma$. Let u_∞ be a partial limit of u_n.

Show that the geodesic that starts from p in the direction of u_∞ is a half-line that runs in K.

(b). The concavity of f follows from 16.17*c*. Whence the convexity of S_c follows.

Suppose that S_c is not compact. Note that we can assume that $p \in S_c$. Then by *(a)* there is a half-line λ that starts at p and runs in S_c for some c. It remains to observe that $f\circ\lambda(t) + t = 0$ and use it to arrive at a contradiction.

(c). Show that $f(x) - c = \operatorname{dist}_{\partial S_c} x$ for any $x \in S_c$. Conclude that S_s has no interior points. Since S_s is convex, the first part follows.

In the Euclidean plane S_s is a single point. In the cylinder, S_s is a circle. The surface of r-neighborhood of an infinite rectangle provides an example with a line segment as S_s; this example is not smooth, but it can be smoothed keeping this property. One can use the cutoffs and mollifiers to do this (Section 0E).

16.20. Show that a cylinder does not contain a line and half-line that meet at exactly one point. It remains to apply the line splitting theorem (16.19).

Afterword

For further study, you need tensor calculus; the book of Richard Bishop and Samuel Goldberg [10] is one of our favorites. Once it is done, you are ready to do Riemannian geometry.

Further you may go with “Comparison geometry” [20] — the good old classics from Jeff Cheeger and David Ebin. In this case you might be surprised to see that almost all of it is already known altho it is written in a different language. For example proofs of the comparisons theorem require only cosmetic modifications.

A safer option would be another classical book “Riemannsche Geometrie im Großen” [35] by Detlef Gromoll, Wilhelm Klingenberg, and Werner Meyer (it is available in German and Russian).

Mikhael Gromov’s “Sign and geometric meaning of curvature” [37] is more challenging, but worth trying.

Good luck.

Anton Petrunin and
Sergio Zamora Barrera.

Index

Bibliography

[1] A. Akopyan, A. Petrunin. “Long geodesics on convex surfaces”. *Math. Intell.* 40.3 (2018), 26–31.

[2] А. Д. Александров. *Внутренняя геометрия выпуклых поверхностей.* 1948.

[3] A. D. Aleksandrov, Yu. G. Reshetnyak. “Rotation of a curve in an n-dimensional Euclidean space”. *Siberian Math. J.* 29.1 (1988), 1–16.

[4] S. Alexander, R. Bishop. “The Hadamard–Cartan theorem in locally convex metric spaces”. *Enseign. Math. (2)* 36.3-4 (1990), 309–320.

[5] S. Alexander, V. Kapovitch, A. Petrunin. *An invitation to Alexandrov geometry: CAT(0) spaces.* 2019.

[6] S. Alexander, V. Kapovitch, A. Petrunin. *Alexandrov geometry: preliminary version no. 1.* 2019. arXiv: `1903.08539 [math.DG]`.

[7] V. Arnold. *Ordinary differential equations.* 2006.

[8] I. D. Berg. “An estimate on the total curvature of a geodesic in Euclidean 3-space-with-boundary.” *Geom. Dedicata* 13 (1982), 1–6.

[9] S. N. Bernstein. “Sur un théoréme de géométrie et son application aux équations aux dérivées partielles du type elliptique”. *Сообщения Харьковского математического общества* 15.1 (1915), 38–45. [German translation in Math. Zeit, 26; Russian translation in УМН, вып. VIII (1941), 75—81 and С. Н. Бернштейн, Собрание сочинений. Т. 3. (1960) с. 251—258].

[10] R. Bishop, S. Goldberg. *Tensor analysis on manifolds.* 1980.

[11] W. Blaschke. *Kreis und Kugel.* 1916.

[12] O. Bonnet. “Mémoire sur les surfaces dont les lignes de courbure sont planes ou sphériques”. *Journal de l’École Polytechnique. Paris* 20 (1853), 117–306.

[13] D. Burago, Yu. Burago, S. Ivanov. *A course in metric geometry.* 2001.

[14] Yu. Burago, M. Gromov, G. Perelman. “A. D. Aleksandrov spaces with curvatures bounded below”. *Russian Math. Surveys* 47.2 (1992), 1–58.

[15] H. Busemann. “Spaces with non-positive curvature”. *Acta Math.* 80 (1948), 259–310.

[16] É. Cartan. *Leçons sur la Géométrie des Espaces de Riemann.* 1928.

[17] G. D. Chakerian. “An inequality for closed space curves”. *Pacific J. Math.* 12 (1962), 53–57.

[18] G. D. Chakerian. “On some geometric inequalities”. *Proc. Amer. Math. Soc.* 15 (1964), 886–888.

[19] G. Chambers, Y. Liokumovich. “Converting homotopies to isotopies and dividing homotopies in half in an effective way”. *Geom. Funct. Anal.* 24.4 (2014), 1080–1100.

[20] J. Cheeger, D. Ebin. *Comparison theorems in Riemannian geometry.* 2008.

[21] J. Cheeger, D. Gromoll. “The splitting theorem for manifolds of nonnegative Ricci curvature”. *J. Differential Geometry* 6 (1971/72), 119–128.

[22] А. В. Чернавский. *Дифференциальная геометрия 2 курс.* 2012.

[23] S. Cohn-Vossen. “Totalkrümmung und geodätische Linien auf einfachzusammenhängenden offenen vollständigen Flächenstücken”. *Матем. сб.* 1(43).2 (1936), 139–164.

[24] E. Denne. *Alternating quadrisecants of knots.* Ph.D. thesis–UIUC. 2004.

[25] P. H. Doyle. "Plane separation". *Proc. Cambridge Philos. Soc.* 64 (1968), 291.

[26] V. Dubrovsky. "In Search of a Definition of Surface Area". *Quantum* 1.4 (1991), 6–9.

[27] J.-H. Eschenburg. "The splitting theorem for space-times with strong energy condition". *J. Differential Geom.* 27.3 (1988), 477–491.

[28] I. Fáry. "Sur certaines inégalites géométriques". *Acta Sci. Math. Szeged* 12.Leopoldo Fejér et Frederico Riesz LXX annos natis dedicatus, Pars A (1950), 117–124.

[29] А. Ф. Филиппов. "Элементарное доказательство теоремы Жордана". *УМН* 5.5(39) (1950), 173–176.

[30] J. Fine. *One-step problems in geometry.* MathOverflow. eprint: `https://mathoverflow.net/q/8378`.

[31] R. Foote, M. Levi, S. Tabachnikov. "Tractrices, bicycle tire tracks, hatchet planimeters, and a 100-year-old conjecture". *Amer. Math. Monthly* 120.3 (2013), 199–216.

[32] D. Fuchs, S. Tabachnikov. *Mathematical omnibus: Thirty lectures on classic mathematics.* 2007.

[33] D. Gale. "The Teaching of Mathematics: The Classification of 1-Manifolds: A Take-Home Exam". *Amer. Math. Monthly* 94.2 (1987), 170–175.

[34] K. F. Gauss. "Disquisitiones generales circa superficies curvas". *Commentationes Societatis Regiae Scientiarum Gottingensis recentiores* 6 (classis mathematicae) (1828), 99–146. [English translation in *General investigations of curved surfaces* 1902.]

[35] D. Gromoll, W. Klingenberg, W. Meyer. *Riemannsche Geometrie im Großen.* 1975.

[36] M. Gromov. "Hyperbolic groups". *Essays in group theory.* Vol. 8. Math. Sci. Res. Inst. Publ. 1987, 75–263.

[37] M. Gromov. "Sign and geometric meaning of curvature". *Rend. Sem. Mat. Fis. Milano* 61 (1991), 9–123 (1994).

[38] J. Hadamard. "Sur certaines propriétés des trajectoires en dynamique". *J. math. pures appl.* 3 (1897), 331–387.

[39] A. Hatcher. *Algebraic topology.* 2002.

[40] H. Hopf, W. Rinow. "Ueber den Begriff der vollständigen differentialgeometrischen Fläche". *Comment. Math. Helv.* 3.1 (1931), 209–225.

[41] H. Hopf. "Über die Drehung der Tangenten und Sehnen ebener Kurven". *Compositio Math.* 2 (1935), 50–62.

[42] H. Hopf. *Differential geometry in the large.* 1989.

[43] В. К. Ионин, Г. Г. Пестов. "О наибольшем круге, вложенном в замкнутую кривую". *Доклады АН СССР* 127 (1959), 1170–1172.

[44] F. Joachimsthal. "Demonstrationes theorematum ad superficies curvas spectantium". *J. Reine Angew. Math.* 30 (1846), 347–350.

[45] A. P. Kiselev. *Kiselev's Geometry: Stereometry.* 2008.

[46] A. Kneser. "Bemerkungen über die Anzahl der Extreme der Krümmung auf geschlossenen Kurven und über verwandte Fragen in einer nichteuklidischen Geometrie." *Heinrich Weber Festschrift.* 1912.

[47] C. Kosniowski. *A first course in algebraic topology.* 1980.

[48] В. Н. Лагунов. "О наибольшем шаре, вложенном в замкнутую поверхность". *Сиб. матем. журн.* 1 (1960), 205–232.

[49] В. Н. Лагунов. "О наибольшем шаре, вложенном в замкнутую поверхность. II". *Сиб. матем. журн.* 2 (1961), 874–883.

[50] В. Н. Лагунов, А. И. Фет. "Экстремальные задачи для поверхностей заданного топологического типа. I". *Сиб. матем. журн.* 4 (1963), 145–176. [English translation in `arXiv:1903.01224`].

[51] В. Н. Лагунов, А. И. Фет. "Экстремальные задачи для поверхностей заданного топологического типа. II". *Сиб. матем. журн.* 6 (1965), 1026–1036.

[52] M. A. Lancret. "Mémoire sur les courbes à double courbure". *Mémoires présentés à l'Institut des Sciences, Lettres et Arts, par divers savants, et lus dans ses assemblées. Sciences mathématiques et physiques.* 1 (1802), 416–454.

[53] U. Lang, V. Schroeder. "On Toponogov's comparison theorem for Alexandrov spaces". *Enseign. Math.* 59.3-4 (2013), 325–336.

[54] N. Lebedeva, A. Petrunin. "On the total curvature of minimizing geodesics on convex surfaces". *St. Petersburg Math. J.* 29.1 (2018), 139–153.

[55] M. Levi. "A "bicycle wheel" proof of the Gauss–Bonnet theorem". *Exposition. Math.* 12.2 (1994), 145–164.

[56] J. Liberman. "Geodesic lines on convex surfaces". *C. R. (Doklady) Acad. Sci. URSS (N.S.)* 32 (1941), 310–313.

[57] H. von Mangoldt. "Ueber diejenigen Punkte auf positiv gekrümmten Flächen, welche die Eigenschaft haben, dass die von ihnen ausgehenden geodätischen Linien nie aufhören, kürzeste Linien zu sein". *J. Reine Angew. Math.* 91 (1881), 23–53.

[58] G. H. Meisters. "Polygons have ears". *Amer. Math. Monthly* 82 (1975), 648–651.

[59] J. B. Meusnier. "Mémoire sur la courbure des surfaces". *Mem des savan etrangers* 10.1776 (1785), 477–510.

[60] А. Д. Милка. "Кратчайшая с неспрямляемым сферическим изображением". *Укр. геометрический сб.* 16 (1974), 35–52.

[61] S. Mukhopadhyaya. "New methods in the geometry of a plane arc". *Bull. Calcutta Math. Soc.* 1 (1909), 31–37.

[62] S. B. Myers. "Riemannian manifolds with positive mean curvature". *Duke Math. J.* 8 (1941), 401–404.

[63] R. Osserman. "The four-or-more vertex theorem". *Amer. Math. Monthly* 92.5 (1985), 332–337.

[64] J. Pach. "Folding and turning along geodesics in a convex surface". *Geombinatorics* 7.2 (1997), 61–65.

[65] R. Palais. "The Morse lemma for Banach spaces". *Bull. Amer. Math. Soc.* 75 (1969), 968–971.

[66] D. Panov. "Parabolic curves and gradient mappings". *Proc. Steklov Inst. Math.* 2(221) (1998), 261–278.

[67] A. Petrunin, A. Yashinski. "Piecewise isometric mappings". *St. Petersburg Math. J.* 27.1 (2016), 155–175.

[68] P. Pizzetti. "Paragone fra due triangoli a lati uguali". *Atti della Reale Accademia dei Lincei, Rendiconti (5) Classe di Scienze Fisiche, Matematiche e Naturali* 16 (1) (1907), 6–11.

[69] A. V. Pogorelov. *Extrinsic geometry of convex surfaces.* 1973.

[70] W. Rinow. *Die innere Geometrie der metrischen Räume.* 1961.

[71] W. Rudin. *Principles of mathematical analysis.* 1976.

[72] A. Schur. "Über die Schwarzsche Extremaleigenschaft des Kreises unter den Kurven konstanter Krümmung". *Math. Ann.* 83.1-2 (1921), 143–148.

[73] J. J. Stoker. "Über die Gestalt der positiv gekrümmten offenen Flächen im dreidimensionalen Raume". *Compositio Mathematica* 3 (1936), 55–88.

[74] J. Sullivan. "Curves of finite total curvature". *Discrete differential geometry.* Oberwolfach Semin. 2008, 137–161.

[75] S. Tabachnikov. "The tale of a geometric inequality". *MASS selecta.* 2003, 257–262.

[76] P. G. Tait. "Note on the circles of curvature of a plane curve." *Proc. Edinb. Math. Soc.* 14 (1896), 26.

[77] V. Toponogov. *Differential geometry of curves and surfaces:A concise guide.* 2006.

[78] В. А. Топоногов. "Римановы пространства кривизны, ограниченной снизу". *УМН* 14.1 (85) (1959), 87–130.

[79] В. А. Топоногов. "Свойство выпуклости римановых пространств положительной кривизны". *Докл. АН СССР* 115.4 (1957), 674–676.

[80] S. Treil. *Linear algebra done wrong.* 2016.

[81] V. V. Usov. "The length of the spherical image of a geodesic on a convex surface". *Siberian Math. J.* 17.1 (1976), 185–188.

[82] R. Webster. *Convexity.* 1994.

[83] В. А. Залгаллер. "Вопрос о сферическом изображении кратчайшей". *Укр. геометрический сб.* 10 (1971), 12.