Differential geometry of curves and surfaces: a working approach

Anton Petrunin and Sergio Zamora Barrera

Contents

Preface

These notes contain a small part of lectures at MASS program (Mathematics Advanced Study Semesters at Pennsylvania State University) Fall semester 2018.

The notes are designed for those who plan to do differential geometry in the future, or at least who want to have a solid ground to decide not to do it. It gives an idea about subject and at the same time it is elementary rigorous and nontechnical.

The highest point is the theorem about the Moon in a puddle proved by Ionin and Pestov. This is the first nontrivial example of the so called *local to global theorems* which is the hart of differential geometry.

Part I

Curves

Chapter 1

Definitions

Simple curves

Recall that that a bijective continuous map $f\colon X\to Y$ between subsets of some metric spaces is called *homeomorphism* if its inverse $f^{-1}\colon Y\to$ $\to X$ 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; that is, any point $p\in\gamma$ has a neighborhood $U\ni p$ such that the intersection $U\cap\gamma$ is homeomorphic to an open real interval.*

It turns out that any curve can be *parameterized* by an open real interval or a circle. That is, for any curve γ there is a homeomorphism $(a,b)\to\gamma$ or $\mathbb{S}^1\to\gamma$ where $\mathbb{S}^1$ denotes the unit circle; that is

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

We omit a proof of this statement; it is not hard, but would take us away from the subject. We hope however that this is intuitively obvious.

Curves that admit a parametrization by a circle are called *closed.* The subsets of curves bounded from one or two sides by points are called *curves with endpoints.* If it has two endpoints, then it is called *arc*; note that any arc can be parameterized by a closed interval. A curve as well as a curve with endpoint(s) can be regarded as a curve; if we need to emphasize that we work with a genuine curve we may say a *curve without endpoints.*

A parametrization describes a curve completely. Often 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 curve admits many different parametrization.

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

Parameterized curves

A *parameterized curve* is defined as a continuous map γ from a circle or a real interval (open, closed or semi-open) to a metric space. For a parameterized curve we do not assume that γ is injective; in other words the parameterized curve might have *self-intersections.*

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

If the domain of a parameterized curve is the closed 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; the point p in this case is called *base* of the loop.

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.*

Smooth curves

A curve in the Euclidean space or plane, is called *space* or *plane curve* correspondingly.

A parameterized space curve can be described by its coordinate functions

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

Plane curves can be considered as a partial 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 $t \mapsto x(t), t \mapsto y(t)$ and $t \mapsto z(t)$ of the space curve γ is a smooth, then the parameterized curve is called *smooth.*

If the *velocity vector*

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

does not vanish at all points, then the parameterized curve γ is called *regular.*

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

1.4. Exercise. *Note that the function*

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

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 any n.

Show that $\gamma(t) = (f(t), f(-t))$ 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 vanishing velocity vector at the origin. Conclude that this curve is not regular and smooth; that is it does not admit a regular smooth parametrization.

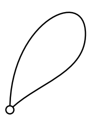
1.5. Exercise. *Describe the set of real numbers ℓ such that the plane curve $\gamma_\ell(t) = (t + \ell\cdot \sin t, \ell\cdot \cos t)$, $t \in \mathbb{R}$ is*

(a) regular;

(b) simple.

Periodic parametrization

Note that any closed simple curve can be described as an image of a loop. However, it is more natural to present it as a *periodic* parameterized curve $\gamma\colon \mathbb{R} \to \mathcal{X}$; that is, such that $\gamma(t + \ell) = \gamma(t)$ for a fixed period ℓ and any t. For example the unit circle in the plane can be described by $2\cdot\pi$-periodic parametrization $\gamma(t) = (\cos t, \sin t)$.



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

Implicitly defined curves

Let $f\colon \mathbb{R}^2 \to \mathbb{R}$ be a smooth function; that is, all its partial derivatives defined in its domain of definition. Consider the set γ of solution of equation $f(x, y) = 0$ in the plane.

Assume γ is connected. According to implicit function theorem (A.10), the set γ is a smooth regular simple curve if 0 is a *regular* *value* of f. In this case it means that the gradient ∇f does not vanish at any point $p \in \gamma$. In other words, if $f(p) = 0$, then $\frac{\partial f}{\partial x}(p) \neq 0$ or $\frac{\partial f}{\partial y}(p) \neq 0$.

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 of solutions is connected and 0 is a regular value of the map $F\colon \mathbb{R}^3 \to \mathbb{R}^2$ defined as

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

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

$$\begin{pmatrix} \frac{\partial f}{\partial x} & \frac{\partial f}{\partial y} & \frac{\partial f}{\partial z} \\ \frac{\partial h}{\partial x} & \frac{\partial h}{\partial y} & \frac{\partial f}{\partial z} \end{pmatrix}$$

for the map $F\colon \mathbb{R}^3 \to \mathbb{R}^2$ has rank 2 at any p such that $f(p) = h(p) = 0$.

The described way to define a curve is called *implicit*; if a curve is defined by its parametrization, we say that it is *explicitly defined.* While implicit function theorem guarantees the existence of regular smooth parametrizations, do not expect it to be in a closed form. When it comes to calculations, usually it is easier to work directly with implicit representation.

1.6. Exercise. *Consider the set in the plane described by the equation*

$$y^2 = x^3.$$

Is it a simple curve? and if "yes", 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.

Proper curves

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

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

Note that any closed curve as well arc are proper curves since its parameter set is compact.

A simple curve is called proper it it admits a proper parametrization. It turns out that simple curve is proper if and only if its image is a closed set. In particular any implicitly defined plane or space curve is proper. We omit the proof of this statement, but it is not hard.

1.8. Exercise. *Use the Jordan's theorem (A.19) to show that any proper plane curve divides the plane in two connected components.*

Chapter 2

Length

Recall that a sequence

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

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*

$$\begin{aligned}\operatorname{length}\gamma = \sup\{|\gamma(t_0) - \gamma(t_1)| + |\gamma(t_1) - \gamma(t_2)| + \dots \\ \dots + |\gamma(t_{k-1}) - \gamma(t_k)|\},\end{aligned}$$

where the exact upper bound is taken over all partitions

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

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 a corresponding loop. If a curve is defined on a open or semi-open interval, then its length is defined as the exact upper bound for lengths of all its closed arcs.

If γ is a space curve, then the above definition says that it length is the exact upper bound of the lengths of polygonal lines $p_0 \dots p_k$ *inscribed* in the curve, where $p_i = \gamma(t_i)$ for a partition $a = t_0 < t_1 < \dots < t_k = b$. If γ is closed, then $p_0 = p_k$ and therefore the inscribed polygonal line is also closed.

2.2. Exercise. *Let $\alpha\colon [0,1] \to \mathbb{R}^3$ be a parametrization of a simple closed arc. Suppose a path $\beta\colon [0,1] \to \mathbb{R}^3$ has that same image as α;*

that is, $\beta([0,1]) = \alpha([0,1])$. *Show that*

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

2.3. 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) $\operatorname{length}\gamma \leqslant \int_a^b |\gamma'(t)|\cdot dt$.

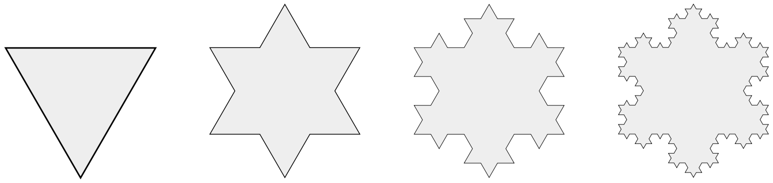
Conclude that

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

Nonrectifiable curves

A classical example of a nonrectifiable curve is the so-called *Koch snowflake*; it is a fractal curve that can be constructed the following way:

Start with an equilateral triangle, divide each of its side into three segments of equal length and add an equilateral triangle with base at the middle segment. Repeat this construction recursively to the obtained polygons. Few first iterations of the construction are shown



on the diagram. The Koch snowflake is the boundary of the union of all the polygons.

2.4. Exercise.

(a) Show that Koch snowflake is a closed simple curve; that is, it can be parameterized by a circle.

(b) Show that Koch snowflake is not rectifiable.

Arc length parametrization

We say that a curve γ has an *arc-length parametrization* (also called *natural parametrization*) if for any two values of parameters $t_1 < t_2$, the value $t_2 - t_1$ is the length of $\gamma|_{[t_1,t_2]}$; that is, the closed arc of γ from t_1 to t_2.

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

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

by that reason smooth curves equipped with arc-length parametrization also called *unit-speed curves.* Note that smooth unit-speed curves are automatically regular.

Any rectifiable curve can be parameterized by arc length. For a parametrized smooth curve γ, the arc-length parameter s can be written as an integral

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

Note that $s(t)$ is a smooth increasing function. Further by fundamental theorem of calculus, $s'(t) = |\gamma'(t)|$. Therefore if γ is regular, then $s'(t) \neq 0$ for any parameter value t. By inverse function theorem (A.9) the inverse function $s^{-1}(t)$ is also smooth. Therefore $\gamma \circ s^{-1}$ — the reparametrization of γ by arc length s — remains smooth and regular.

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

2.5. 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 the given smooth regular curve comes with an arc-length parametrization. A good property of arc-length parametrizations is that it is almost canonical — these parametrizations differ only by a sign and additive constant. By that reason, it is easier to express parametrization-independent quatities using arc-length parametrizations; this will be usefull in the definition of curvature and torsion.

On the other hand, often it is impossible to find an arc-length parametrization in a closed form which makes it hard to use it calculations; usually it is more convenient to use the original parametrization.

Convex curves

A simple plane curve is called *convex* if it bounds a convex region.

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

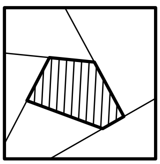
$$\operatorname{length} \alpha \leqslant \operatorname{length} \beta.$$

Note that it is sufficient to show that for any polygon P inscribed in α there is a polygon Q inscribed in β with $\operatorname{perim} P \leqslant \operatorname{perim} Q$, where $\operatorname{perim} P$ denotes the perimeter of P.

Therefore it is sufficient to prove the following lemma.

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

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



Proof. Note that by the triangle inequality, the inequality

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

holds if P can be obtained from Q by cutting it along a chord; that is, a line segment with ends on the boundary of Q that lies in Q.

Note that there is an increasing sequence of polygons

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

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. □

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

Proof. Any closed curve is bounded; that is, it lies in a sufficiently large square. 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 continous functions defined on $[0,1]$. Therefore the absolute values of both of these functions are bounded by some constant C. That is, α lies in the square defined by the inequalities $|x| \leqslant C$ and $|y| \leqslant C$.

By Proposition 2.6, the length of the curve cannot exceed the perimeter of the square $8{\cdot}C$, whence the result. □

Recall that convex hull of a set X is the smallest convex set that contains X; in other words convex hull is the intersection of all convex sets containing X.

2.9. Exercise. *Let α be a closed simple 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 curve α, assuming only that X has nonempty interior.

Crofton formulas*

Consider a smooth plane curve $\gamma\colon [a,b]\to\mathbb{R}^2$. Given a unit vector u, denote by γ_u the curve that follows orthogonal projections of γ to the line in the direction u; that is,

$$\gamma_u(t) = \langle u, \gamma(t)\rangle{\cdot}u.$$

Note that

$$|\gamma_u'(t)| = |\langle u, \gamma'(t)\rangle|$$

for any t. Note that for any plane vector the magnitude of its average projection is proportional to its magnitude with coefficient; that is,

$$|w| = k{\cdot}\overline{|w_u|},$$

where $\overline{|w_u|}$ denotes the average value of $|w_u|$ for all unit vectors u. (The value k is the average value of $|\cos\varphi|$ for $\varphi\in[0,2{\cdot}\pi]$; it can be found by integration, but soon we will show another way to find it.)

If the curve γ is smooth, then according to Exercise 2.3

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

This formula and its relatives are called *Crofton formulas.* To find the coefficient k one can apply it for the unit circle: the left hand side is $2{\cdot}\pi$ — this is the length of unit circle. Note that for any unit vector u, the curve γ_u runs back and forth along an interval of length 2. Therefore $\operatorname{length}\gamma_u = 4$ and hence its average value is also 4. It follows that the coefficient k has to satisfy the equation $2{\cdot}\pi = k{\cdot}4$; whence

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

The Crofton's formula holds for arbitrary rectifiable curves, not necessary smooth; it can be proved using Exercises 2.12.

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

Use this statement to give another solution of Exercise 2.9.

2.11. Exercise. *Show that the length of space curve is proportional to the average length of its projections to all lines and to planes. Find the coefficients in each case.*

2.12. 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 simple curve $\gamma\colon [a,b]\to\mathbb{R}^3$ such that the velocity vector $\gamma'(t)$ is defined and bounded for almost all $t\in[a,b]$, but the formula ❶ does not hold.*

Semicontinuity of length

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

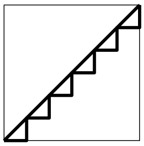
$$\varliminf_{n\to\infty} x_n.$$

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 exteded real line $[-\infty,\infty]$

2.13. Theorem. *Length is a lower semi-continuous 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, $\gamma_n(t)\to\gamma_\infty(t)$ for any fixed $t\in[a,b]$ as $n\to\infty$. Then

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



Note that the inequality ❷ might be strict. For example the diagonal γ_∞ of the unit square can be approximated by a sequence of stairs-like polygonal curves γ_n with sides parallel to the sides of the square (γ_6 is on the picture). In this case

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

for any n.

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

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

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

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

If γ_∞ is rectifiable, we can assume that

$$\operatorname{length}\gamma_\infty < \Sigma_\infty + \varepsilon.$$

for any given $\varepsilon > 0$. By ❸ it follows that

$$\varliminf_{n\to\infty} \operatorname{length}\gamma_n > \operatorname{length}\gamma_\infty - \varepsilon$$

for any $\varepsilon > 0$; whence ❷ follows.

It remains to consider the case when γ_∞ is not rectifiable; that is, $\operatorname{length}\gamma_\infty = \infty$. In this case we can choose a partition so that $\Sigma_\infty > L$ for any real number L. By ❸ it follows that

$$\varliminf_{n\to\infty} \operatorname{length}\gamma_n > L$$

for any given L; whence

$$\varliminf_{n\to\infty} \operatorname{length}\gamma_n = \infty$$

and ❷ follows. $\square$

Length metric

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

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

Evidently $d(x, y) \geqslant |x - y|$ for any pair of points $x, y \in \mathcal{X}$. If the equality holds for any pair, then then the metric is called *length metric* and the space is called *length-metric space*.

Most of the time we consider length-metric spaces. In particular the Euclidean space is a length-metric space. A subspaces A of length-metric space $\mathcal{X}$ might be not 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 exact lower bound for the length of paths in A.

2.14. 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 $x, y \in A$

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.

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

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

Proof. The short arc γ of a grt circle from u to v in $\mathbb{S}^2$ has length $\measuredangle(u, v)$; that is, $|u - v|_{\mathbb{S}^2} \leqslant \measuredangle(u, v)$.

It remains to prove the opposite inequality. In other words, we need to sho that given an 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

❹ $$\operatorname{length}\beta_1 \geqslant \operatorname{length}\beta.$$

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 the triangle inequality for angles (A.24), 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 ❹ follows. □

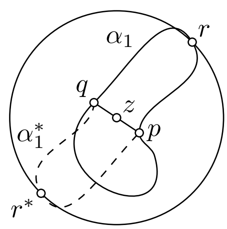
2.16. Hemisphere lemma. *Any closed spherical curve of length less than $2{\cdot}\pi$ lies in an open hemisphere.*

This lemma is a keystone in the proof of Fenchel's theorem given below (see 3.7). The lemma is not as simple as you might think — try to prove it yourself. The following proof is due to Stephanie Alexander.

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

Let us divide γ into two arcs γ_1 and γ_2 of length ℓ, with endpoints p and q. Note that

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



The north hemisphere corresponds to the disc and the south hemisphere to the complement of the disc.

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, say r. Without loss of generality we may assume that r lies on γ_1.

Rotate the arc γ_1 by 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 $2{\cdot}\ell$ that passes thru r and its antipodal point r^*. Therefore

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

a contradiction. □

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

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

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

2.19. Exercise. *Consider the following problem, find a flaw in the given solution. 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,

$$|p-q| \leqslant 2 \tag{❺}$$

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 ❺ follows. □

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

1. *Show that for any spherical curve γ we have that*

   $$\operatorname{length}\gamma = \overline{\operatorname{length}\gamma_v},$$

   where $\overline{\operatorname{length}\gamma_v}$ denotes the average length for all $v \in \mathbb{S}^2$. (This is a spherical analog of Crofton's formula.)
2. *Give another proof of hemisphere lemme using part (1).*

Chapter 3

Curvature

Acceleration of a unit-speed curve

Recall that any regular smooth curve can be parametrized by its arc length. 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 $\langle \gamma', \gamma' \rangle = 1$. Therefore

$$2 \cdot \langle \gamma'', \gamma' \rangle = \langle \gamma', \gamma' \rangle' = 0,$$

whence $\gamma'' \perp \gamma'$. $\square$

Curvature

For a unit-speed smooth space curve γ the magnitude of its acceleration $|\gamma''(s)|$ is called its *curvature* at the 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 latter case 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, curvature tells how much to turn the steering wheel at the given point (note that it does not depend on your speed).

In general, the term *curvature* is used for different types of geometric objects, and it always measures how much it deviates from being *straight*; for curves, it measures how fast it deviates from a straight line.

3.2. Exercise. *Show that any regular smooth spherical curve has curvature at least 1 at each time.*

Tangent indicatrix

Let γ be a regular smooth space curve. Let us consider another curve

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

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

If γ has a unit-speed parametrization, then $\mathrm{T}(t) = \gamma'(t)$. In this case we have the following expression for curvature: $\kappa(t) = |\mathrm{T}'(t)| = |\gamma''(t)|$.

In general case we have

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

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

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

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

3.3. 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)

$$\kappa(t) = \frac{|\mathrm{W}(t)|}{|\gamma'(t)|^2},$$

where $\mathrm{W}(t)$ *denotes the projection of* $\gamma''(t)$ *to the normal plane of* $\gamma'(t)$*;*

(b)

$$\kappa(t) = \frac{|\gamma''(t) \times \gamma'(t)|}{|\gamma'(t)|^3},$$

where $\times$ *denotes the* vector product *(also known as* cross product*).*

3.4. 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 $p = (x, f(x))$*.*

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 *tangent line* at t. It 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 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 coinside with the tangent line of γ_2 at s_2; if both of the 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.

Total curvature

Let $\gamma\colon \mathbb{I} \to \mathbb{R}^3$ be a regular smooth curve and T its tangent indicatrix. Recall that without loss of generality we can assume that γ has a unit-speed parametrization; in this case $\mathrm{T}(s) = \gamma'(s)$ and hence

$$\begin{aligned}\kappa(s) &:= |\gamma''(s)| = \\ &= |\mathrm{T}'(s)|;\end{aligned}$$

that is, the curvature of γ at time s is the speed of the tangent indicatrix T at the same time moment.

The integral

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

is called *total curvature of* γ.

3.5. 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 $t \in [0, 2{\cdot}\pi]$.

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

Proof. It is sufficient to prove the observation for a unit-speed space curve $\gamma\colon \mathbb{I} \to \mathbb{R}^3$. Denote by T its tangent indicatrix. Then

$$\begin{aligned}\Phi(\gamma) &:= \int\limits_{\mathbb{I}} \kappa(s){\cdot}ds = \\ &= \int\limits_{\mathbb{I}} |\gamma''(s)|{\cdot}ds = \\ &= \int\limits_{\mathbb{I}} |\mathrm{T}'(s)|{\cdot}ds = \\ &= \operatorname{length} \mathrm{T}.\end{aligned}$$

□

3.7. 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 it is described by a 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 T lies in the north hemisphere defined by the inequality $z > 0$ in (x, y, z)-coordinates. It means that $z'(t) > 0$ at any time, where $\gamma(t) = (x(t), y(t), z(t))$. 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 (3.6) and the hemisphere lemma (2.16), we get that

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

3.8. Exercise. *Show that a closed space curve γ with curvature at most* 1 *cannot be shorter than the unit circle; that is,* $\operatorname{length}\gamma\geqslant 2\cdot\pi$*.*

3.9. 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*

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

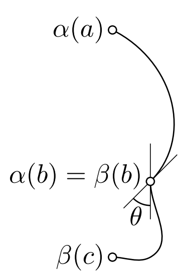
Moreover, if γ is closed, then

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

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

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

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 $\alpha(b)=\beta(b)$. Then one can combine these two curves 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 cancatination can be applied if α and/or β are defied on semiopen intervals $(a,b]$ and/or $[b,c)$.

The concatenation can be also defined if the end point of the first curve coincides with the starting point of the second curve; 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 concatination 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 *external angle* of γ at time b.

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

If γ is a concatination 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.

3.10. 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

$$\Phi(\gamma_1) + \dots + \Phi(\gamma_n) + \theta_1 + \dots + \theta_n \geqslant 2\cdot\pi. \qquad ❸$$

Consider the tangent indicatrix $\mathrm{T}_1, \dots, \mathrm{T}_n$ for each arc $\gamma_1, \dots, \gamma_n$; 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 end point 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 $\gamma_{n+1} = \gamma_1$). Therefore if we connect the end point of T_i to the starting point of T_{i+1} by a short arc of a great circle in the sphere, then we add $\theta_1 + \dots + \theta_n$ to the total length of $\mathrm{T}_1, \dots, \mathrm{T}_n$.

Applying the hemisphee lemma (2.16) to the obtained closed curve, we get that

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

Applying the observation (3.6), we get ❸. □

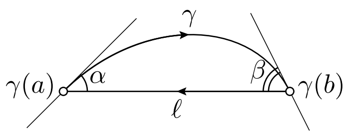
3.11. Chord lemma. *Let ℓ be the chord to a smooth regular arc $\gamma\colon [a,b] \to \mathbb{R}^3$. Assume γ meets ℓ at angles α and β at its ends; that is*

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

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

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

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 curvature of ℓ vanish, we get that



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

According to the generalized Fenechel's theorem (3.10),

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

hence the result. □

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

Polygonal lines

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

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

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

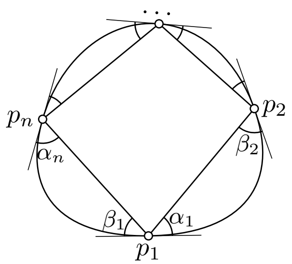
Use this statement to show that any closed polygonal line has curvature at least $2{\cdot}\pi$.

3.14. Proposition. *Assume a polygonal line $\beta = p_1 \dots p_n$ is inscribed in a smooth regular curve γ. Then*

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

Moreover if γ is closed we can assume that the inscribed polygonal line β is also closed.

Proof. Since the curvature of line segments vanishes, the total curvature of polygonal line is the sum of external angles $\theta_i = \pi - \measuredangle p_{i-1}p_ip_{i+1}$.



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 case of closed curve we use indexes modulo n, in particular $p_{n+1} = p_1$.

Note that $\theta_i = \measuredangle(\mathrm{w}_{i-1}, \mathrm{w}_i)$. By triangle inequality for angles A.24, 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:

$$\begin{aligned} \Phi(\beta) &= \theta_2 + \dots + \theta_{n-1} \leqslant \\ &\leqslant \beta_2 + \alpha_2 + \dots + \beta_{n-1} + \alpha_{n-1} \leqslant \\ &\leqslant (\alpha_1 + \beta_2) + \dots + (\alpha_{n-1} + \beta_n) \leqslant \\ &\leqslant \Phi(\gamma). \end{aligned}$$

□

3.15. Exercise.

(a) *Draw a smooth regular plane curve γ which has a self-intersection, 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 $\Phi(\gamma) > \pi$.*

3.16. Proposition. *The equality case in the 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 3.14.

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

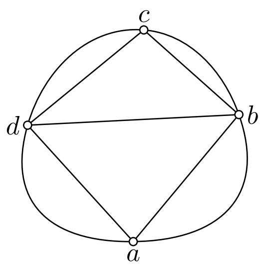
$$\begin{aligned}\Phi(abcd) &= (\pi - \measuredangle dab) + (\pi - \measuredangle abc) + (\pi - \measuredangle bcd) + (\pi - \measuredangle cda) = \\ &= 4\cdot\pi - (\measuredangle dab + \measuredangle abc + \measuredangle bcd + \measuredangle cda))\end{aligned}$$

Note that

❹ $$\measuredangle abc \leqslant \measuredangle abd + \measuredangle dbc \quad \text{and} \quad \measuredangle cda \leqslant \measuredangle cdb + \measuredangle bda.$$

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

$$\begin{aligned}\Phi(abcd) &\geqslant 4\cdot\pi - (\measuredangle dab + \measuredangle abd + \measuredangle bda) - (\measuredangle bcd + \measuredangle cdb + \measuredangle dbc) = \\ &= 2\cdot\pi.\end{aligned}$$



By 3.14,

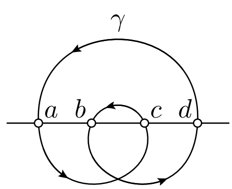
$$\Phi(abcd) \leqslant \Phi(\gamma) \leqslant 2\cdot\pi.$$

Therefore we have equalities in ❹. 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 quadraliteral.

It follows that any quadraliteral inscribed in γ is convex plane quadraliteral. Therefore all points of γ lie in one plane and the domain bounded by γ is convex; that is, γ is a convex plane curve. □

3.17. 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*

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

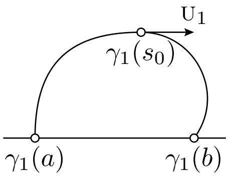


Lines crossing a curve at four points as in the exercise are called *alternating quadrisecants.* It turns out that any *nontrivial knot* admits an alternating quadrisecant [1]; according to the exercise the latter implies the so-called *Fáry–Milnor theorem* — the total curvature any knot exceeds $4\cdot\pi$.

Bow lemma

3.18. 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; denote by $\kappa_1(s)$ and $\kappa_2(s)$ their curvatures at s. Suppose that $\kappa_1(s)\geqslant\kappa_2(s)$ for any s and the curve γ_1 is a simple arc of a convex curve; that is, it runs in the boundary of a covex plane figure. Then the distance between the ends of γ_1 cannot exceed the distance between the ends of γ_2; that is,*

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



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

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

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

By construction the vector U_1 is parallel to $\gamma(b)-\gamma(a)$ in particular

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

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

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

That is,

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

if $s \geqslant s_0$. The same argument shows that

$$❺ \qquad \measuredangle(\gamma_1'(s), \mathrm{u}_1) \geqslant \measuredangle(\gamma_2'(s), \mathrm{u}_2)$$

for $s \geqslant s_0$; therefore the inequality holds for any s.

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 ❺, we get that

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

Hence the result. □

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

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

(a) *Show that*

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

(b) *Use part (a) to give another solution of 3.15b.*

(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 $\cos\theta$.*

3.20. Exercise. *Suppose that two points p and q lie on a unit circle dividing it in two arcs with lengths $\ell_1 < \ell_2$. Show that if a curve γ runs from p to q and has curvature at most 1, then either*

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

The following exercise generalizes 3.8.

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

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

DNA inequality*

Recall that curvature of a spherical curve is at least 1 (Exercise 3.2). In particular the length of spherical curve cannot exceed its total curvature. The following theorem shows that the same inequality holds for *closed* curves in a unit ball.

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

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

This theorem was proved by Don Chakerian [2]; for plane curves it was proved earlier by István Fáry [3]. We present the proof given by Don Chakerian in [4]; few other proofs of this theorem are discussed by Serge Tabachnikov [5].

Proof. Without loss of generality we can assume the curve is described by a loop $\gamma\colon [0,\ell]\to\mathbb{R}^3$ parameterized by its arc length, so $\ell=\operatorname{length}\gamma$. We can also assume that the origin is the center of the ball. It follows that

$$\langle\gamma'(s),\gamma'(s)\rangle=1,\qquad |\gamma(s)|\leqslant 1$$

and in particular

$$\text{❻}\qquad \begin{aligned}\langle\gamma''(s),\gamma(s)\rangle&\geqslant -|\gamma''(s)|\cdot|\gamma(s)|\geqslant\\ &\geqslant -\kappa(s)\end{aligned}$$

for any s, where $\kappa(s)=|\gamma''(s)|$ is the curvature of γ at s.

Since γ is a smooth closed curve, we have that $\gamma'(0)=\gamma'(\ell)$ and

$\gamma(0)=\gamma(\ell)$. Applying ❻, we get that

$$\begin{aligned}0&=\langle\gamma(\ell),\gamma'(\ell)\rangle-\langle\gamma(0),\gamma'(0)\rangle=\\&=\int\limits_0^\ell\langle\gamma(s),\gamma'(s)\rangle'\cdot ds=\\&=\int\limits_0^\ell\langle\gamma'(s),\gamma'(s)\rangle\cdot ds+\int\limits_0^\ell\langle\gamma(s),\gamma''(s)\rangle\cdot ds\geqslant\\&\geqslant\ell-\Phi(\gamma),\end{aligned}$$

whence the result. □

Nonsmooth curves*

3.23. Theorem. *For any regular smooth space curve γ we have that*

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

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

Proof. Note that the inequality

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

follows from 3.14; it remains to show

❼ $$\Phi(\gamma)\leqslant\sup\{\Phi(\beta)\}.$$

Let $\gamma\colon[a,b]\to\mathbb{R}^3$ be a smooth curve. Fix a partition $a=t_0<\dots<t_k=b$ and consider the corresponding inscribed polygonal line $\beta=p_0\dots p_k$. (If γ is closed, then $p_0=p_k$ and β is closed as well.)

Let $\tau=\xi_1\dots\xi_k$ be a spherical polygonal line with the vertexes $\xi_i=\frac{p_i-p_{i-1}}{|p_i-p_{i-1}|}$. We can assume that τ has constant speed on each arc and $\tau(t_i)=\xi_i$ for each i. The spherical polygonal line τ will be called tangent indicatrix for β.

Consider a sequence of finer and finer partitions, denote by β_n and τ_n the corresponding inscribed polygonal lines and their tangent indicatrixes. Note that since γ is smooth, the idicatrixes τ_n converge pointwise to T — the tangent indicatrix of γ. By semi-continuity of the length (2.13), we get that

$$\begin{aligned}\Phi(\gamma) &= \operatorname{length} \mathrm{T} \leqslant \\ &\leqslant \varliminf_{n\to\infty} \operatorname{length} \tau_n = \\ &= \varliminf_{n\to\infty} \Phi(\beta_n) \leqslant \\ &\leqslant \sup\{\Phi(\beta)\},\end{aligned}$$

where the least upper bound is taken over all partitions and their corresponding inscribed polygonal lines β; whence ❼ follows. □

The theorem above can be used to define total curvature for arbitrary curves, not necessary (piecewise) smooth and regular. We say that a parameterized curve is trivial if it is constant; that is, it stays at one point.

3.24. Definition. *The total curvature of a nontrivial parameterized space curve γ is the exact upper bound on the total curvatures of inscribed nondegenerate polygonal lines; if γ is closed, then we assume that the inscribed polygonal lines are closed as well.*

3.25. Exercise. *Show that the total curvature is lower semi-continuous with respect to pointwise convergence of curves. That is, if a sequence of curves $\gamma_n\colon [a,b] \to \mathbb{R}^3$ converges pointwise to a nontrivial curve $\gamma_\infty\colon [a,b] \to \mathbb{R}^3$, then*

$$\varliminf_{n\to\infty} \Phi(\gamma_n) \geqslant \Phi(\gamma_\infty).$$

3.26. Exercise. *Generalize Fenchel's theorem all nontrivial closed space curves. That is, show that*

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

for any closed space curve γ(not necessary piecewise smooth and regular).

3.27. Exercise. *Assume that a curve $\gamma\colon [a,b] \to \mathbb{R}^3$ has finite total curvature. Show that γ is rectifiable.*

Construct a rectifiable curve $\gamma\colon [a,b] \to \mathbb{R}^3$ that has infinite total curvature.

A good survey on curves of finite total curvature is written by John Sullivan [6].

DNA inequality revisited*

In this section we will give an alternative proof of 3.22 that works for arbitrary, not necessarily smooth, curves. In the proof we use 3.24 to define for the total curvature; according to 3.23, it is more general than the smooth definition given on page 24.

Alternative proof of 3.22. We will show that

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

for any closed polygonal line $\gamma = p_1 \dots p_n$ in a unit ball. It implies the theorem since in any nontrivial closed curve we can inscribe a closed polygonal line with arbitrary close total curvature and length.

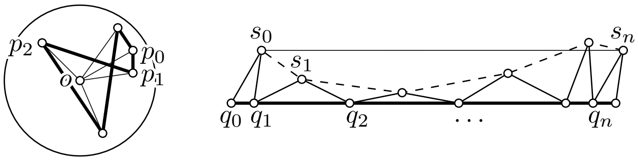
The indexes are taken modulo n, in particular $p_n = p_0$, $p_{n+1} = p_1$ and so on. Denote by θ_i the external angle of γ at p_i; that is,

$$\theta_i = \pi - \measuredangle p_{i-1}p_ip_{i+1}.$$

Denote by o the center of the ball. Consider a sequence of $n+1$ plane triangles

$$\begin{aligned}\triangle q_0s_0q_1 &\cong \triangle p_0op_1,\\ \triangle q_1s_1q_2 &\cong \triangle p_1op_2,\\ &\dots\\ \triangle q_ns_nq_{n+1} &\cong \triangle p_nop_{n+1},\end{aligned}$$

such that the points $q_0, q_1 \dots$ lie on one line in that order and all the points $s_0, \dots, s_n$ lie on one side from this line.



Since $p_0 = p_n$ and $p_1 = p_{n+1}$, we have that

$$\triangle q_ns_nq_{n+1} \cong \triangle p_nop_{n+1} = \triangle p_0op_1 \cong \triangle q_0s_0q_1,$$

so $s_0q_0q_ns_n$ is a parallelogram. Therefore

$$\begin{aligned}|s_0 - s_1| + \dots + |s_{n-1} - s_n| &\geqslant |s_n - s_0| =\\ &= |q_0 - q_n| =\\ &= |p_0 - p_1| + \dots + |p_{n-1} - p_n|\\ &= \operatorname{length}\gamma.\end{aligned}$$

Since $|q_i - s_{i-1}| = |q_i - s_i| = |p_i - o| \leqslant 1$, we have that

$$\measuredangle s_{i-1} q_i s_i > |s_{i-1} - s_i|$$

for each i. Therefore

$$\begin{aligned}
\theta_i &= \pi - \measuredangle p_{i-1} p_i p_{i+1} \geqslant \\
&\geqslant \pi - \measuredangle p_{i-1} p_i o - \measuredangle o p_i p_{i+1} = \\
&= \pi - \measuredangle q_{i-1} q_i s_{i-1} - \measuredangle s_i q_i q_{i+1} = \\
&= \measuredangle s_{i-1} q_i s_i > \\
&> |s_{i-1} - s_i|.
\end{aligned}$$

That is,

$$\theta_i > |s_{i-1} - s_i|$$

for each i.

It follows that

$$\begin{aligned}
\Phi(\gamma) &= \theta_1 + \dots + \theta_n > \\
&> |s_0 - s_1| + \dots |s_{n-1} - s_n| \geqslant \\
&\geqslant \operatorname{length}\gamma.
\end{aligned}$$

Hence the result. □

Let us mention the following closely related statement:

3.28. Theorem. *Suppose a closed regular smooth curve γ lies in a convex figure with the perimeter $2{\cdot}\pi$. Then*

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

This statement was conjectured by Serge Tabachnikov [5]. Despite the simplicity of the formulation, the proof is annoyingly difficult; it was proved by Jeffrey Lagarias and Thomas Richardson [7]; latter a simpler proof was given by Alexander Nazarov and Fedor Petrov [8].

Chapter 4

Signed curvature

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 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 on page **??**; we will keep the name *Frenet frame* for it.

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

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

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 to the curvature $\kappa(s)$ of γ at s defined on page 21; the sign tells which direction it turns — if γ turns left, then $k>0$. If we want to emphasise that we work with *nonsigned* 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 that

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

Differentiating these idenitites we get that

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

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

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

The equations ❶ and ❷ are Frenet formulas for plane curves. They could be also written in a matrix form:

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

4.1. 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 tracks of idealized bicycle with the distance 1 from rear to front wheel. Thus by the exercise, the front wheel have to have the longer track. For more on geometry of bicycle tracks see [9] and the references there in.

Fundamental theorem of plane curves

4.2. 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 $k(s)$ at every s. Moreover γ is uniquely defined up to a rigid motion of the plane.*

This is the fundamental theorem of plane curves; it is direct analog of **??** and it can be proved along the same lines. We give 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 any 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

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

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

The existence is proved; it remains to prove 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 correspondingly. The triples $\gamma_i, \mathrm{T}_i, \mathrm{N}_i$ satisfy the same system system of ordinary differential equations

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

Motreover, they have the same the initial values at s_0. Therefore $\gamma_1 = \gamma_2$. □

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

4.3. Corollary. *Suppose $\gamma\colon \mathbb{I} \to \mathbb{R}^2$ is 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 on is chosen in such a way that $\gamma(s_0)$ is the origin and $\gamma'(s_0)$ points in the direction of x-axis. Then*

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

where

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

Total signed curvature

Let $\gamma\colon \mathbb{I} \to \mathbb{R}^2$ be a smooth unit-speed plane curve. The integral of its signed curvature is called *total signed curvature* and it denoted by

$\Psi(\gamma)$; so

❸
$$\Psi(\gamma) = \int_{\mathbb{I}} k(s)\cdot ds,$$

where k denotes signed curvature of γ.

If $\mathbb{I} = [a, b]$, then

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

where θ is as in 4.3.

If γ is a piecewise smooth and regular plane curve, then we define its total signed curvature as the sum of total signed curvatures of its arcs plus the sum of signed external angles at the joints; it is positive if γ turns left, negative if γ turns right, 0 if it goes straight and undefined if it turns backward. 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 that

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

for any smooth regular plane curve γ; that is, total signed curvature can not exceed total curvature by absolute value.

4.4. Proposition. *The total signed curvature of any closed simple smooth regular plane curve γ is $\pm 2\cdot\pi$; 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 sum of the internal angles of a polygon (A.23) which we use in the proof. A more conceptual proof was given by Heinz Hopf [10], [11, p. 42].

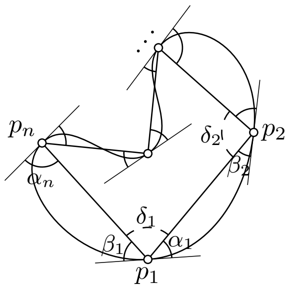
Proof. Without loss of generality we may assume that γ is oriented in such a way that the region bounded by γ lies on the left from it. We can also assume that it parametrized by arc length.

Consider a closed polygonal line $p_1 \dots p_n$ inscribed in γ. We can assume that the arcs between the vertexes are sufficiently small; in this case the polygonal line is simple and each arc γ_i from p_i to p_{i+1} have 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 ❸, the value

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

is a multiple of $2{\cdot}\pi$. Since $\Phi(\gamma_i) < \pi$, by chord lemma (3.11), we also have that $|\alpha_i| + |\beta_i| < \pi$. By ❹, we have that $|\Psi(\gamma_i)| \leqslant \Phi(\gamma_i)$; therefore the value in ❺ vanishes, or equivalently

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

for each i.

Note that

$$❻ \qquad \delta_i = \pi - \alpha_i - \beta_i$$

is the internal angle of $p_1 \dots p_n$ at p_i; $\delta_i \in (0, 2{\cdot}\pi)$ for each i. Recall that the sum of internal angles of an n-gon is $(n-2){\cdot}\pi$ (see A.23); that is,

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

Therefore

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

The piecewise smooth and regular curve 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 ❻ we have

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

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

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

□

4.5. Exercise. *Draw a smooth regular closed plane curve with zero total signed curvature.*

4.6. 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 on the signed distance ℓ.

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

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

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

for all ℓ sufficiently close to 0. Describe an example showing that this formula does not hold for all ℓ.

Osculating circline

4.7. 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 $\sigma(s) = p + s\cdot\mathrm{T}$ which runs along the line; if $k \ne 0$, then σ runs around the circle of radius $\frac{1}{|k|}$ and center $p + \frac{1}{k}\cdot\mathrm{N}$, where T, N is an oriented orthonoral 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 calculation based on 4.2 and 4.3.

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

$$\theta(s) = \int\limits_0^s k\cdot dt = $$
$$= k\cdot s.$$

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 $\sigma(s) = p + s\cdot\mathrm{T}$.

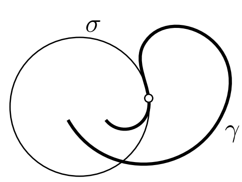
If $k \neq 0$, we get

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

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

4.8. 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 *center of curvature* and *radius of curvature* of the curve at that point.

The osculating circle σ_s can be also defined as the (necessarily 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.

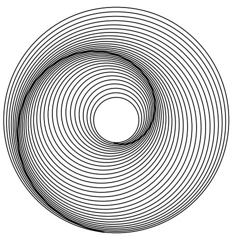
Recall that the *inverse* of a point x with respect to the unit circle centered at the origin is the point $\hat{x} = \frac{x}{|x|^2}$.

4.9. 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 in the unit circle centered at the origin. Show that osculating circline of $\hat{\gamma}$ at s is the inversion of osculating circline of γ at s.*

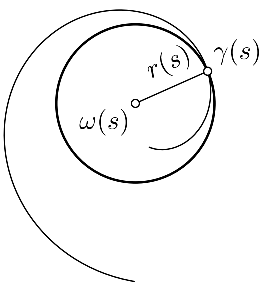
Spiral lemma

The following lemma was proved by Peter Tait [12] and later rediscovered by Adolf Kneser [13].

4.10. 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 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 it must be constant in the annulus [see 14, 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 a circle.



Proof. Let $\mathrm{T}(s), \mathrm{N}(s)$ be the Frenet frame, $\omega(s)$ the curvature center and $r(s)$ the radius of curvature of γ at s. By 4.7,

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

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. It follows that

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

In other words, the distance between the centers of σ_{s_1} and σ_{s_0} is strictly less than the difference between their radiuses. 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 *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 $(\frac{1}{k})'\cdot \mathrm{N}$ is its velocity vector.

4.11. Exercise. *Show that the stretched astroid*

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

is an evolute of the ellipse $\gamma(t) = (a\cdot\cos t, b\cdot\sin t)$.

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

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

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

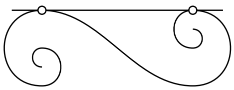
Proof of 4.12. Note that $\gamma(s)$ lies on the osculating circle σ_s of γ at s. If $s_1 \neq s_0$, then by lemma σ_{s_0} does not intersect σ_{s_1}. Therefore $\gamma(s_1)\neq\gamma(s_0)$, hence the result. □

4.13. 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.*

4.14. 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 5

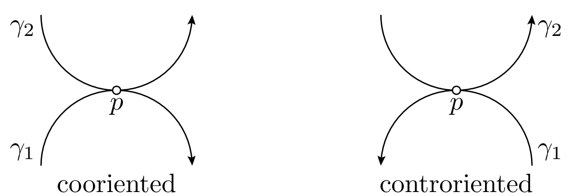
Supporting curves

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; that is, the tangent lines of γ_1 at t_1 coincides with the tangent line γ_2 at t_2.

In this case the point $p = \gamma_1(t_1) = \gamma_2(t_2)$ is called a *point of tangency* of the curves. If one of the curves is simple, then we may say that γ_1 and γ_2 are tangent at the point p without ambiguity.

Note that if γ_1 γ_2 are tangent to 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



$\gamma_2'(t_2)$ point in the same direction we say that the curves a *cooriented*, if these directions are opposite, we say that the curves are *controriented* at the time parameters t_1 and t_2.

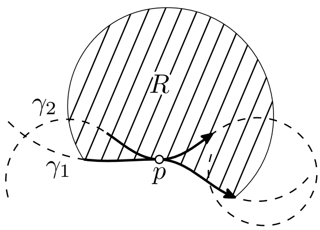
Note that reverting parametrization of one of the curves, cooriented curves become counteroriented and the other way around; so we can always assume that the curves are cooriented at the given point of tangency.

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 supports γ_2 at the time parameters t_1 and t_2. If both curves are simple, then we also could say that γ_1 *locally supports* γ_2 at the point p without ambiguity.



If γ_1 is simple proper curve, so it divides the plane into two closed region that lie on 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.

If γ_2 is a closed curve so it divides the plane into two regions a bounded inside and unbounded outside. In this case we say if γ_1 supports γ_2 *from inside* (*from outside*) if γ_1 supports γ_2 and in the region inside it (correspondingly outside it).

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

5.1. 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 particular, we could say if γ_1 and γ_2 are coorineted or controri-ented at at the time parameters t_1 and t_2. If the curves are coorienated and the region R in the definition of supporting curves lie on the right (left) from the arc of γ_1, then we say that γ_1 supports γ_2 from the left (correspondingly right).

If the curves on the diagram oriented according 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).

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 γ is simple we could say that the point $p = \gamma(s)$ is a vertex of γ without ambiguity.

5.2. Exercise. *Assume that 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.*

Supporting test

The following proposition resembles the second derivative test.

5.3. 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

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

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

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

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

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 $f_1''(0) = k_1(0)$ (see 3.4)

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

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

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

5.4. Advanced exercise. *Let γ_1 and γ_2 be two smooth unit-speed simple plane curves that are tangent and cooriented at the point $p = \gamma_1(0) = \gamma_2(0)$. Assume $k_1(s) \geqslant k_2(s)$ for any s. Show that γ_1 locally supports γ_2 from the left at p.*

Give an example of two proper curves γ_1 and γ_2 satisfying the above condition such that γ_1 does not globally support γ_2 at p.

Note that according to the DNA inequality (3.22) for any closed smooth regular curve that runs in a unit disc, the average of its absolute curvature at lest 1; in particular it has a point with absolute curvature at lest 1. The following exercise says that the last statement holds for loops.

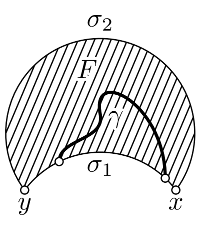
5.5. Exercise. *Assume a closed smooth regular plane loop γ runs in a unit disc. Show that there is a point on γ with absolute curvature at least 1.*

5.6. Exercise. *Assume a closed smooth regular plane curve γ runs between parallel lines on 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.

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

The last exercise above is a baby case of a 5.15.



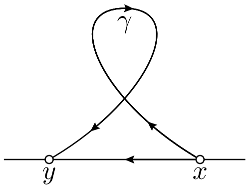
5.8. Exercise. *Let F be a plane figure bounded by two circle arcs σ_1 and σ_2 of signed curvature 1 that run from x to y. Suppose σ_1 is a shorter than σ_2. Assume a simple arc γ runs in F and has the end points on σ_1. Show that the absolute curvature of γ is at least 1 at some parameter value.*

Convex curves

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

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

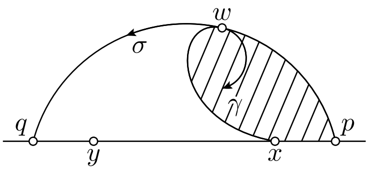
5.10. Lens lemma. *Let γ be a smooth regular simple curve that runs from x to y. Assume that γ runs on the right side (correspondingly left side) of the oriented line xy and only its end points x and y lie on the line. Then γ has a point with positive (correspondingly negative) signed curvature.*

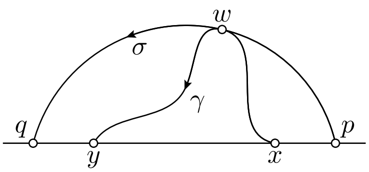


Note that the lemma fails for curves with self-intersections; the curve γ on the diagram always turns right, so it has negative curvature everywhere, but it lies on the right side of the line xy.

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

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





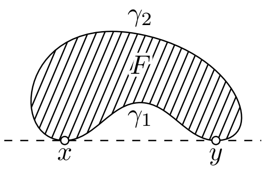
Note that the γ is tangent to σ at w. Let us parameterise σ from p to q. Then γ and σ are cooriented as w. If not, then the arc of γ from w to y would be trapped in the curvelinear triangle xwp bounded by arcs of σ, γ and the line segment $[px]$. But this is impossible since y does not belong to this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

It follows that σ supports γ at t_0 from the right. By 5.3,

$$k(t_0)_\gamma \geqslant k_\sigma,$$

Evidently $k_\sigma > 0$, hence the result. $\square$

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



Proof of 5.9. 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 line has vanishing curvature, the supporting test (5.3) implies that $k \geqslant 0$ at each point.

Now assume F is not convex. Then 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 the one of the arcs 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; so we can apply the lens lemma from which we get that the arcs γ_1 and γ_2 contain points with curvatures of opposite signs.

That is, if F is not convex, then curvature of γ changes sign. Equivalently: if curvature of γ does not change sign, then F is convex. □

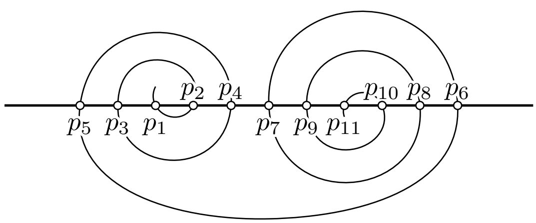
5.11. Exercise. *Suppose γ is a smooth regular simple closed convex plane curve of diameter bigger than 2. Show that γ has a point with absolute curvature less than 1.*

5.12. Exercise. *Suppose γ is a simple smooth regular curve in the plane with positive curvature. Assume γ crosses a line ℓ at the points $p_1, p_2, \ldots p_n$ 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:

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

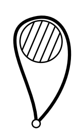


(c) Try to describe all possible orders when p_1 lies between p_2 and p_3 (see the diagram).

Moon in a puddle

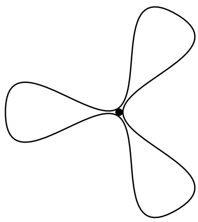
The following theorem is a slight generalization of the theorem proved by Vladimir Ionin and German Pestov in [15].

5.13. 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). For convex curves, this result was known earlier [16, §24].

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 diagram it does not always lead to a solution a closed plane curve of absolute curvature bounded by 1 may surround a disc of radius smaller than 1 that cannot be enlarged continuously.



5.14. 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 support γ from the inside.*

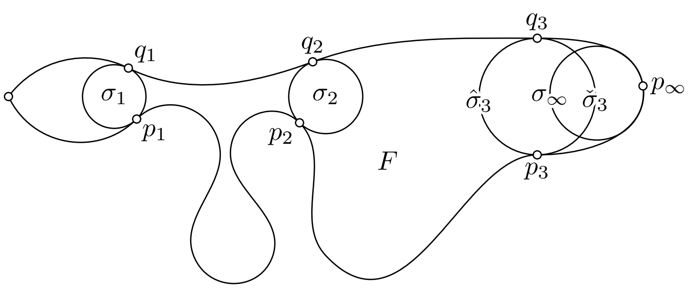
First let us show that the theorem follows from the lemma.

Proof of 5.13 modulo 5.14. 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 support γ from inside. In particular σ lies inside of γ, whence the result. □

Proof of 5.14. Denote by F the closed region surrounded by γ.

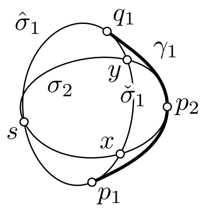
We need to show that one osculating circle lies completely in F. Assume contrary; that is, the osculating circle at each point $p \in \gamma$ does not lie in F.

Given a point $p \in \gamma$ let us consider the maximal circle σ that lies completely in F and tangent to γ at p. Note that σ has to touch γ at another point; otherwise one could increase its radius slightly while keeping the circle in F. The circle σ will be called *incircle* at p.



Fix a point p_1 and 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 and σ_2 be the incircle at p_2.



Two ovals on the diagram pretend to be circles.

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 two more common points with $\check\sigma_1$ — x and y, one for each arc of σ_2 from p_2 to s. Therefore $\sigma_1 = \sigma_2$ as two circles with three common points: s, x, and y. On the other hand, by construction, we have that $p_2 \in \sigma_2$ and $p_2 \notin \sigma_1$ — a contradiction.

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. Note that by construction we have that

❶ $$\operatorname{length}\gamma_2 < \tfrac{1}{2}\cdot\operatorname{length}\gamma_1.$$

Repeat 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

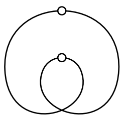
$$\bigcap_n \gamma_n$$

contains a single point; denote it by p_∞.

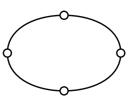
Let σ_∞ be the incircle at p_∞; it has to touch γ at another point, say q_∞. The same argument as above shows that $q_\infty \in \gamma_n$ for any n. It follows that $q_\infty = p_\infty$ — a contradiction. □

5.15. 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 lies in F. Show that absolute curvature of γ is at least $\frac{1}{R}$ at some parameter value.*

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.



5.16. Four-vertex theorem. *Any smooth regular simple plane curve has at least four vertices.*

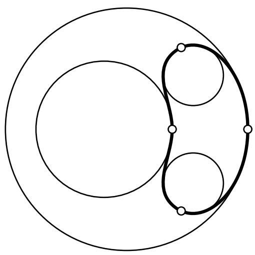
Evidently any closed curve has at least two vertexes — where the minimum and the maximum of the curvature are attained. On the diagram the vertexes are marked; the first curve has one self-intersection and exactly two vertexes; the second curve has exactly four vertexes and no self-intersections.

The four-vertex theorem was first proved by Syamadas Mukhopadhyaya [17] for convex curves. By now it has a large number of different proofs and generalizations. One of my favorite proofs was given by Robert Osserman [18].

We present a proof based on the key lemma in the previous section. In fact we prove the following stronger version of the four-vertex theorem:

5.17. Theorem. *Any smooth regular simple plane curve has is globally supported by its osculating circle at least at 4 distinct points; two from inside and two from outside.*

The diagram shows for supporting osculating circles, two from inside and two outside the curve for the given curve.



Proof of 5.16 and 5.17. First note that if an osculating circline σ at a point p supports γ locally, then p is a vertex. Indeed, if p is not a vertex, then a small arc around p has monotonic curvature. Applying the spiral lemma (4.10) we get that the osculating circles at this arc are nested. In particular the curve γ crosses σ at p and therefore σ is does not locally support γ at p.

We showed that 5.17 implies the four-vertex theorem (5.16); it remains to prove 5.17.

According to key lemma (5.14), 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 existence of another point $q \in \gamma$ with the same property.

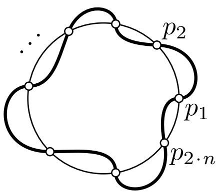
We showed existence of two osculating circles that support γ from inside; it remains to show existence of two osculating circles that support γ from outside.

If one applies to γ an inverse with the center inside γ, then the obtained curve γ_1 also has two osculating circles that support γ_1 from inside. According to 4.9, these osculating circlines are inverses of the circlines osculating to γ. Note that the rigion 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.

(Alternatively, 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.) □

This theorem is a direct corollary of 5.14; indeed, since absolute curvature is bounded by 1, every osculating circle has radius at least 1 and by 5.14 two pf these circles are surrounded by γ.

According to 5.2, if an osculating circle supports the curve at the same point p, then p is a vertex. Therefore 5.14 implies existence of 4 vertexes of γ. That is, we proved the following theorem:



5.18. 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 $2\cdot n$ vertexes.*

Construct an example of a closed simple smooth regular plane curve γ with only 4 vertexes that crosses a given circle at arbitrarily many points.

Part II

Surfaces

Chapter 6

Definitions

Topological surfaces

We will be mostly interested in smooth regular surfaces defined in the following section. However few times we will 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 which is *homeomorphic* to 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.

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* (or more precisely *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 formed by 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 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 surfaces as well since one can choose a coordinate system so that it becomes the (x, y)-plane. We can 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).

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 smooth surface.

Surfaces with boundary

A connected subset in a surface that is bounded by one or more curves is called *surface with boundary*; in this case the collection of curves is called the *boundary line* of the surface.

When we say *surface* we usually mean a *smooth regular surface without boundary*; we may use the terms *surface without boundary* if we need to emphasise it; otherwise we may use the term *surface with possibly nonempty boundary.*

Proper, closed and open surfaces

If the surface Σ is formed by a closed set, then it is called *proper.* For example, paraboloids

$$z = x^2 + y^2, \qquad z = x^2 - y^2$$

or sphere

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

are proper surfaces, while the open disc

$$\{\, (x, y, z) \in \mathbb{R}^3 \,:\, x^2 + y^2 < 1,\ z = 0 \,\}$$

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, any paraboloid is an open surface; sphere is a closed surface.

Note that any proper surface without boundary is either closed or open.

The following claim is a two-dimesional analog of 1.8; hopefully it is intuitively obvious. Its proof is not at all trivial; a standard proof uses the so-called *Alexander's duality* which is a classical technique in algebraic topology. We omit its proof.

6.1. Claim. *The complement of any proper topological surface without boundary (or, equivalently any open or closed topological surface) has exactly two connected components.*

Implicitly defined surfaces

6.2. 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, $\nabla_p f \neq 0$ if $f(p) = 0$. Then any connected component Σ of the set of solutions of the equation $f(x, y, z) = 0$ is a surface.*

Proof. Fix $p \in \Sigma$. Since $\nabla_p f \neq 0$ we have

$$\tfrac{\partial f}{\partial x}(p) \neq 0, \quad \tfrac{\partial f}{\partial y}(p) \neq 0, \quad \text{or} \quad \tfrac{\partial f}{\partial z}(p) \neq 0.$$

We may assume $\frac{\partial f}{\partial z}(p) \neq 0$; otherwise permute the coordinates x, y, z.

The implicit function theorem (A.10) 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 (page 58). $\square$

6.3. Exercise. *Describe the set of real numbers ℓ such that the equation*

$$x^2 + y^2 - z^2 = \ell$$

describes a smooth regular surface.

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 $\frac{\partial s}{\partial u}$ and $\frac{\partial s}{\partial v}$ are linearly independent at any $(u,v) \in U$; equivalently $\frac{\partial s}{\partial u} \times \frac{\partial s}{\partial v} \neq 0$, where $\times$ denotes the vector product.

6.4. 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

$$s(u,v) = (x_s(u,v), y_s(u,v), z_s(u,v)).$$

Since s is regular, its Jacobian matrix

$$\begin{pmatrix} \frac{\partial x_s}{\partial u} & \frac{\partial x_s}{\partial v} \\ \frac{\partial y_s}{\partial u} & \frac{\partial y_s}{\partial v} \\ \frac{\partial z_s}{\partial u} & \frac{\partial z_s}{\partial v} \end{pmatrix}$$

has rank two at any pint $(u,v) \in U$.

Fix a point $p \in \Sigma$; by shifting the coordinate system we may assume that p is the origin. Permuting the coordinates x, y, z if necessary, we may assume that the matrix

$$\begin{pmatrix} \frac{\partial x_s}{\partial u} & \frac{\partial x_s}{\partial v} \\ \frac{\partial y_s}{\partial u} & \frac{\partial y_s}{\partial v} \end{pmatrix},$$

which is the Jacobian matix of the map $(u,v) \mapsto (x_s(u,v), y_s(u,v))$, is invertible.

The inverse function theorem implies that there is a smooth regular function h defined on an open set $W \ni 0$ in the (x,y)-plane such that

$$(x_s \circ h)(x,y) = x \quad \text{and} \quad (y_s \circ h)(x,y) = y$$

for any $(x,y) \in W$. It follows that the graph $z = z_s \circ h(x,y)$ for $(x,y) \in W$ is a subset in Σ. Clearly 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 *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. But 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, $\frac{\partial s}{\partial u} = (1, 0, \frac{\partial h}{\partial u})$ and $\frac{\partial s}{\partial v} = (0, 1, \frac{\partial h}{\partial v})$. Whence $\frac{\partial s}{\partial u}$ and $\frac{\partial s}{\partial v}$ are linearly independent; that is, s is a regular map.

Note that from 6.4, we obtain the following corollary.

6.5. Corollary. *A connected set $\Sigma \subset \mathbb{R}^3$ is a smooth regular surface if Σ has a local parametrization by a smooth regular map at any point $p \in \Sigma$.*

6.6. Exercise. *Consider the following map*

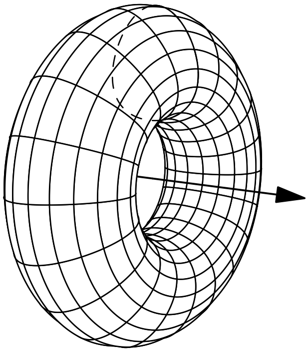
$$s(u, v) = (\tfrac{2\cdot u}{1+u^2+v^2}, \tfrac{2\cdot v}{1+u^2+v^2}, \tfrac{2}{1+u^2+v^2}).$$

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

$$(u, v, 1) \mapsto (\tfrac{2\cdot u}{1+u^2+v^2}, \tfrac{2\cdot v}{1+u^2+v^2}, \tfrac{2}{1+u^2+v^2})$$

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 $(\tfrac{2\cdot u}{1+u^2+v^2}, \tfrac{2\cdot v}{1+u^2+v^2}, \tfrac{2}{1+u^2+v^2})$ lie on one half-line starting at 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

$$(t, \theta) \mapsto (x(t), y(t)\cdot \cos\theta, y(t)\cdot \sin\theta).$$

For fixed t or θ the obtained curves are called *meridian* or correspondingly *parallel* of the surface of revolution; note that parallels are formed by circles in the plane perpendicular to the axis of rotation.

6.7. Exercise. *Assume γ is a closed simple smooth regular plane curve that does not intersect the x-axis. Show that surface of revolution of γ around the x-axis is a smooth regular surface.*

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, c. Note that by 6.2, $\Sigma_{a,b,c}$ is a smooth regular surface. Indeed, set $f(x,y,z) = \frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2}$, 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. Since $\Sigma_{a,b,c}$ it is a smooth regular surface.

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 map $(x,y,z) \mapsto (a{\cdot}x, b{\cdot}y, c{\cdot}z)$ to the unit sphere $\mathbb{S}^2$.

For a surface Σ, a map $s : \Sigma \to \mathbb{R}^3$ is called a *smooth parametrized surface* if 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 following two vectors $\frac{\partial}{\partial u}(s \circ f)$ and $\frac{\partial}{\partial v}(s \circ f)$ are linearly independent.

Evidently the parametric definition includes the embedded surfaces defined previously — as the domain of parameters we can take the surface itself and the identity map as s. But parametrized surfaces are more general, in particular they might have self-intersections.

If Σ is a known surface for example a sphere or a plane, the parametrized surface $s\colon \Sigma \to \mathbb{R}^3$ might be called by the same name. For example, any embedding $s\colon \mathbb{S}^2 \to \mathbb{R}^3$ might be called a topological sphere and if s is smooth and regular, then it might be called smooth sphere. (A smooth regular map $s\colon \mathbb{S}^2 \to \mathbb{R}^3$ which is not necessary an embedding is called a *smooth regular immersion*, so we can say that s describes a smooth immersed sphere.) Similarly an embedding $s\colon \mathbb{R}^2 \to \mathbb{R}^3$ might be called topological plane, and if s is smooth, it might be called smooth plane.

Tangent plane

6.8. 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.*

6.9. 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* 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 $\frac{\partial s}{\partial u}(u_p, v_p)$ and $\frac{\partial s}{\partial v}(u_p, v_p)$.

The tangent plane to Σ at p is usually denoted by T_p or $\mathrm{T}_p\Sigma$. Tangent plane T_p might be considered as a linear subspace of $\mathbb{R}^3$ or as a parallel plane passing thru p. In the latter case it can be interpreted as the best approximation of the surface Σ by a plane at p; 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 function $u(t)$ and $v(t)$ such that

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

Applying chain rule, we get

$$\gamma' = \tfrac{\partial s}{\partial u}\cdot u' + \tfrac{\partial s}{\partial v}\cdot v';$$

that is, γ' is a linear combination of $\frac{\partial s}{\partial u}$ and $\frac{\partial s}{\partial v}$.

Since the functions $u(t)$ and $v(t)$ can be chousen arbitrary, we have that any linear combination of $\frac{\partial s}{\partial u}$ and $\frac{\partial s}{\partial v}$ is a tangent vector at the corresponding point. $\square$

6.10. Exercise. *Let $f : \mathbb{R}^3 \to \mathbb{R}$ be a smooth function with a regular value 0 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 $p \in \Sigma$.*

6.11. Exercise. *Let Σ be a smooth surface and $p \in \Sigma$. Fix an (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 a* vertical plane*; that is, if the projection of T_p to the (x, y)-plane does not degenerate to a line.*

Directional derivative

In this section we extend the definition of directional derivative to smooth functions defined on smooth surfaces.

6.12. Proposition-Definition. *Let Σ be a smooth regular surface and f is a smooth function defined on Σ. Suppose γ is a smooth curve in Σ that starts at p with the 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* directional derivative of f along w at p *and denoted by*

$$D_{\mathrm{w}}f, \quad (D_{\mathrm{w}})f(p), \quad \textit{or} \quad (D_{\mathrm{w}})f(p)_\Sigma$$

— 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 $p = s(u_p, v_p)$, and $\mathrm{w} = a\cdot \frac{\partial s}{\partial u}(u_p, v_p) + b\cdot \frac{\partial s}{\partial v}(u_p, v_p)$*, then*

$$D_{\mathrm{w}}f(p) = a\cdot \tfrac{\partial f\circ s}{\partial u}(u_p, v_p) + b\cdot \tfrac{\partial f\circ s}{\partial v}(u_p, v_p).$$

Note that our definition agrees with 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 the 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 will 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

$$\gamma'(0) = u'(0)\cdot \tfrac{\partial s}{\partial u}(u_p, v_p) + v'(0)\cdot \tfrac{\partial s}{\partial v}(u_p, v_p).$$

Since $\mathrm{w} = \gamma'(0)$ and the vectors $\frac{\partial s}{\partial u}$, $\frac{\partial s}{\partial v}$ are linearly independent, we get that $a = u'(0)$ and $b = v'(0)$.

Applying the chain rule again, we get that

➊ $$(f \circ \gamma)'(0) = a\cdot \tfrac{\partial f\circ s}{\partial u}(u_p, v_p) + b\cdot \tfrac{\partial f\circ s}{\partial v}(u_p, v_p).$$

Notice that the left hand side in ➊ 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 ➊. □

Any smooth map f from a surface Σ to $\mathbb{R}^3$ can be described by its coordinate functions $f(p) = (f_x(p), f_y(p), f_z(p))$. To take a directional derivative of the map we should take the directional derivative of each of its coordinate function.

$$D_{\mathrm{w}}f := (D_{\mathrm{w}}f_x, D_{\mathrm{w}}f_y, D_{\mathrm{w}}f_z).$$

6.13. Exercise. *Assume f is a smooth map from one smooth surface Σ_0 to another Σ_1 and $p \in \Sigma_0$. Show that $D_{\mathrm{w}}f(p) \in \mathrm{T}_{f(p)}\Sigma_1$ for any $\mathrm{w} \in \mathrm{T}_p$.*

6.14. Exercise. *Assume f is a smooth function from one smooth surface Σ. Show that the map $\mathrm{w} \mapsto D_{\mathrm{w}}f(p)$ defined for $\mathrm{w} \in \mathrm{T}_p$ is linear; that is, $D_{c\cdot\mathrm{w}}f = c\cdot D_{\mathrm{w}}f(p)$ and $D_{\mathrm{v}+\mathrm{w}}f = D_{\mathrm{v}}f(p) + D_{\mathrm{w}}f(p)$ for any $c \in \mathbb{R}$ and $\mathrm{v}, \mathrm{w} \in \mathrm{T}_p$.*

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 takes a smooth function φ on Σ and spits the directional derivative $D_{\mathrm{w}}\varphi$. It is straightforward to check the product rule:

❷ $$D_{\mathrm{w}}(\varphi\cdot\psi) = (D_{\mathrm{w}}\varphi)\cdot\psi(p) + \varphi(p)\cdot(D_{\mathrm{w}}\psi).$$

It turns out 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 ❷.

This new definition is less intuitive, but it is more convenient to use since it grabs the key algebraic property of tangent vectors. Many statements admit simpler proofs with this approach, for example linearity of the map $\mathrm{w} \mapsto D_{\mathrm{w}}f$ (6.14) becomes a tautology.

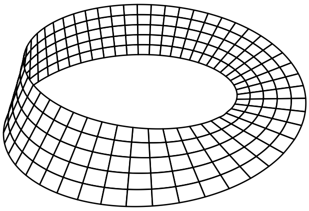
Normal vector and orientation

A unit vector that is normal to T_p is usually denoted by ν_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

[1]Term *functional* is used for functions that take a function as an argument and return a number.

$|\nu_p| = 1$ for any p. The choice of the field ν is called the *orientation* on Σ. A surface Σ is called *orientable* if it can be oriented. Note that each orientable surface admits two orientations ν and $-\nu$.



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, when you go around it changes the sign.

Note that each surface is locally orientable. In fact each chart $f(u,v)$ admits an orientation

$$\nu = \frac{\frac{\partial f}{\partial u} \times \frac{\partial f}{\partial v}}{\left|\frac{\partial f}{\partial u} \times \frac{\partial f}{\partial v}\right|}.$$

Indeed the vectors $\frac{\partial f}{\partial u}$ and $\frac{\partial f}{\partial v}$ 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.

6.15. Exercise. *Let $h : \mathbb{R}^3 \to \mathbb{R}$ be a smooth function with a regular value 0 and Σ is 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 (6.1). Therefore we can choose the unit normal field on any smooth proper surfaces that points into one of the components of the complement. Therefore we obtain the following observation.

6.16. Observation. *Any smooth open or closed surface in Euclidean space is oriented.*

In particular it follows that the Möbius strip cannot be extended to an open or closed smooth surface without boundary.

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 *spherical map* or *Gauss map*.

For surfaces, the spherical map plays essentially the same role as the tangent indicatrix for curves.

Plane sections

6.17. Advanced exercise. *Show that any closed set in the plane can appear as an intersection of this plane with an open smooth regular* *surface.*

As a consequence of the exercise above, the 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.

6.18. Lemma. *Let Σ be a smooth regular surface. Then for any plane Π there is an arbitrarily close parallel plane Π' such that each connected component of the intersection $\Sigma\cap\Pi'$ is a smooth regular curve.*

Proof. Assume Π is described by equation $f(x,y,z)=r_0$, where

$$f(x,y,z)=a{\cdot}x+b{\cdot}y+c{\cdot}z.$$

The surface Σ can be covered by a countable set of charts $s_i\colon U_i\to\Sigma$. Note that the composition $f\circ s_i$ is a smooth function for any i. By Sard's lemma (A.11), almost all real numbers r are regular values of each $f\circ s_i$.

Fix such a value r sufficiently close to r_0 and consider the plane Π' described by the equation $f(x,y,z)=r$. Note that $\Pi'\parallel\Pi$ and is arbitrarily close to it. Any point in the intersection $\Sigma\cap\Pi'$ lies in the image of one of the charts. From above it admits a neighborhood which is a regular smooth curve; hence the result. $\square$

Chapter 7

Curvatures

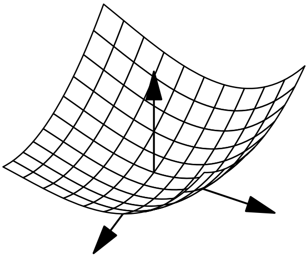
Tangent-normal coordinates

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 in the direction of the normal vector ν_p. By 6.11, we can present Σ locally around p as a graph of a function f. Note that f satisfies the following additional properties:

$$f(0,0) = 0, \qquad (\tfrac{\partial}{\partial x} f)(0,0) = 0, \qquad (\tfrac{\partial}{\partial y} f)(0,0) = 0.$$

The first equality holds since $p = (0,0,0)$ lies on the graph and the last two equalities mean that the tangent plane at p is horizontal.

Set



$$\begin{aligned} \ell &= (\tfrac{\partial^2}{\partial x^2} f)(0,0), \\ m &= (\tfrac{\partial^2}{\partial x \partial y} f)(0,0) = (\tfrac{\partial^2}{\partial y \partial x} f)(0,0), \\ n &= (\tfrac{\partial^2}{\partial y^2} f)(0,0). \end{aligned}$$

The Taylor series for f at $(0,0)$ up to the second oreder 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*

$$z = \tfrac{1}{2}{\cdot}(\ell{\cdot}x^2 + 2{\cdot}m{\cdot}x{\cdot}y + n{\cdot}y^2)$$

gives the best approximation of the surface at p; it 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

$$M_p = \begin{pmatrix} \ell & m \\ m & n \end{pmatrix}; \qquad \text{❶}$$

it is called the Hessian matrix of f at $(0,0)$.

Principle 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 is equivalent to changing its the basis, which results in the rewriting the matrix M_p in this new basis.

Since the Hessian matrix M_p is symmetric, it is diagonalizable by orthogonal matrices. That is, by rotating the (x,y)-plane we can assume that $m = 0$ in ❶. In this case the diagonal components of M_p are called *principle curvatures* of Σ at p; they are uniquely defined up to sign; they are 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 these are the curvatures of 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 principle curvatures can be also defined as the eigenvalues of M_p; the eigendirections of M_p are called *principle directions* of Σ at p. Note that if $k_1(p) \neq k_2(p)$, then p has exactly two principle directions, which are perpendicular to each other; if $k_1(p) = k_2(p)$ then all tangent directions at p are principle.

Note that if we revert the orientation of Σ, then the principle curvatures at each point switch their signs and indexes.

A smooth regular curve on a surface Σ that always runs in the principle directions is called a *line of curvature* of Σ.

7.1. Exercise. *Assume that a smooth surface Σ is mirror symmetric with respect to a plane Π. Suppose that Σ and Π intersect along a curve γ. Show that γ is a line of curvature of Σ.*

Shape operator

Let p be a point on a smooth oriented surface Σ. Suppose Σ is described locally as a graph $z = f(x, y)$ in a tangent-normal coordinates at p and

$$M_p = \begin{pmatrix} \ell & m \\ m & n \end{pmatrix};$$

is the Hessian matrix of f at $(0, 0)$; that is, the components ℓ, m, and n are as on page 69.

The multiplication by the Hessian matrix defies the so called *shape operator*

$$S\colon \left(\begin{smallmatrix} x \\ y \end{smallmatrix}\right) \mapsto M_p \cdot \left(\begin{smallmatrix} x \\ y \end{smallmatrix}\right);$$

it is a linear operator $S\colon \mathrm{T}_p \to \mathrm{T}_p$. For a point $p \in \Sigma$ the shape operator of a tangent vector $\mathrm{w} \in \mathrm{T}_p$ will be denoted by $S(\mathrm{w})$ if it is clear from the context which base point p and which surface we are working with; otherwise we may use notations

$$S_p(\mathrm{w}) \quad \text{or even} \quad S_p(\mathrm{w})_\Sigma.$$

Since M_p is symmetric, S is *self-adjoint*; that is

$$\langle S(\mathrm{v}), \mathrm{w} \rangle = \langle \mathrm{v}, S(\mathrm{w}) \rangle$$

for any $\mathrm{v}, \mathrm{w} \in \mathrm{T}_p$. Note also that principle curvatures of Σ at p are the eigenvalues of S_p and the principle directions are the directions of principle vectors of S_p.

7.2. Proposition. *Let p be a point on a smooth oriented surface Σ. Suppose Σ is described locally as a graph $z = f(x, y)$ in a tangent-normal coordinates at p. Then*

$$\langle S(\mathrm{v}), \mathrm{w} \rangle = D_{\mathrm{w}} D_{\mathrm{v}} f(0, 0)$$

for any $\mathrm{v}, \mathrm{w} \in \mathrm{T}_p$. Moreover S is unique linear operator $\mathrm{T}_p \to \mathrm{T}_p$ that satisfies the above condition.

Here $D_{\mathrm{v}} f$ denoted directional derivative of f along vector v; that is, if $\varphi(t) = f(q + \mathrm{v} \cdot t)$, then $D_{\mathrm{v}} f(q) = \varphi'(0)$.

Proof. Suppose $\mathrm{v} = \left(\begin{smallmatrix} a \\ b \end{smallmatrix}\right)$ and $\mathrm{v} = \left(\begin{smallmatrix} c \\ d \end{smallmatrix}\right)$, then

$$D_{\mathrm{v}} = a \cdot \tfrac{\partial}{\partial x} + b \cdot \tfrac{\partial}{\partial y}, \qquad\qquad D_{\mathrm{w}} = c \cdot \tfrac{\partial}{\partial x} + d \cdot \tfrac{\partial}{\partial y}.$$

Therefore

$$D_{\mathrm{w}} D_{\mathrm{v}} f = a \cdot c \cdot \tfrac{\partial^2 f}{\partial^2 x} + b \cdot c \cdot \tfrac{\partial^2 f}{\partial x \partial y} + a \cdot d \cdot \tfrac{\partial^2 f}{\partial y \partial x} + b \cdot d \cdot \tfrac{\partial^2 f}{\partial^2 y}$$

evaluating this expression at $(0,0)$ we get

$$\begin{aligned} D_{\mathrm{w}}D_{\mathrm{v}}f(0,0) &= a{\cdot}c{\cdot}\ell + b{\cdot}c{\cdot}m + a{\cdot}d{\cdot}m + b{\cdot}d{\cdot}n = \\ &= \langle M_p{\cdot}\mathrm{v}, \mathrm{w}\rangle = \langle \mathrm{v}, M_p{\cdot}\mathrm{w}\rangle = \\ &= \langle S(\mathrm{v}), \mathrm{w}\rangle = \langle \mathrm{v}, S(\mathrm{w})\rangle. \end{aligned}$$ □

7.3. Corollary. *Let p be a point on a smooth oriented surface Σ. Suppose Σ is described locally as a graph $z = f(x,y)$ in a tangent-normal coordinates at p. Denote by I, J and K the standard basis in the (x,y,z)-coordinates. Then*

$$\begin{aligned} &\langle S(\mathrm{I}), \mathrm{I}\rangle = \ell, && \langle S(\mathrm{I}), \mathrm{J}\rangle = m, && \langle S(\mathrm{I}), \mathrm{K}\rangle = 0, \\ &\langle S(\mathrm{J}), \mathrm{I}\rangle = m, && \langle S(\mathrm{J}), \mathrm{J}\rangle = n, && \langle S(\mathrm{J}), \mathrm{K}\rangle = 0, \end{aligned}$$

where ℓ, m, and n are the components of the Hessian matrix of f at $(0,0)$ defined on page 69.

Proof. Note that
$$D_{\mathrm{I}} = \tfrac{\partial}{\partial x} \quad \text{and} \quad D_{\mathrm{J}} = \tfrac{\partial}{\partial y}.$$
It remains to use 7.2 and the expressions for ℓ, m, and n on page 69. □

In the following proposition we use the notion of directional derivative defined in 6.12.

7.4. Proposition. *Let Σ be a smooth surface with unit normal field ν. Suppose $p \in \Sigma$ and $S\colon \mathrm{T}_p \to \mathrm{T}_p$ is the shape operator at p. Then*

❷ $$S(\mathrm{w}) = -D_{\mathrm{w}}\nu$$

for any $\mathrm{w} \in \mathrm{T}_p$.

The reason for minus sign in ❷ is the same as in the formula for curvature of plane curve in its Frenet frame: $\mathrm{N}' = -k{\cdot}\mathrm{T}$. This proposition will be proved in the next section.

7.5. Exercise. *Let Σ be a smooth oriented surface with the unit normal field ν. Suppose that Σ has unit principle curvatures at any point.*

(a) *Show that $S_p(w) = w$ for any $p \in \Sigma$ and $w \in \mathrm{T}_p\Sigma$.*
(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 part of the unit sphere centered at c.*

7.6. Exercise. *Assume that smooth 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 along a smooth curve γ, then γ is a curvature line of Σ.

Proof of 7.4*

Proof of 7.4. Let Σ be a smooth surface with unit normal field ν. Suppose $(u,v)\mapsto s(u,v)$ is a local chart of Σ at p. Since ν is unit we have the identity

$$1=\langle \nu\circ s,\nu\circ s\rangle.$$

Note that the vectors $\frac{\partial s}{\partial u}$ and $\frac{\partial s}{\partial v}$ tangent at their base point; therefore we have two more identities:

$$0=\langle\tfrac{\partial}{\partial u}s,\nu\circ s\rangle, \qquad 0=\langle\tfrac{\partial}{\partial u}s,\nu\rangle.$$

Taking partial derivatives of these there identities and applying the product rule, we get the following six identities:

$$\begin{aligned}
0&=\tfrac{\partial}{\partial u}\langle \nu\circ s,\nu\circ s\rangle=2\cdot\langle\tfrac{\partial}{\partial u}\nu\circ s,\nu\circ s\rangle,\\
0&=\tfrac{\partial}{\partial v}\langle \nu\circ s,\nu\circ s\rangle=2\cdot\langle\tfrac{\partial}{\partial v}\nu\circ s,\nu\circ s\rangle,\\
0&=\tfrac{\partial}{\partial u}\langle\tfrac{\partial}{\partial u}s,\nu\circ s\rangle=\langle\tfrac{\partial^2}{\partial u^2}s,\nu\circ s\rangle+\langle\tfrac{\partial}{\partial u}s,\tfrac{\partial}{\partial u}\nu\circ s\rangle,\\
0&=\tfrac{\partial}{\partial v}\langle\tfrac{\partial}{\partial u}s,\nu\circ s\rangle=\langle\tfrac{\partial^2}{\partial v\partial u}s,\nu\circ s\rangle+\langle\tfrac{\partial}{\partial u}s,\tfrac{\partial}{\partial v}\nu\circ s\rangle\\
0&=\tfrac{\partial}{\partial u}\langle\tfrac{\partial}{\partial v}s,\nu\circ s\rangle=\langle\tfrac{\partial^2}{\partial u\partial v}s,\nu\circ s\rangle+\langle\tfrac{\partial}{\partial v}s,\tfrac{\partial}{\partial u}\nu\circ s\rangle,\\
0&=\tfrac{\partial}{\partial v}\langle\tfrac{\partial}{\partial v}s,\nu\circ s\rangle=\langle\tfrac{\partial^2}{\partial v^2}s,\nu\circ s\rangle+\langle\tfrac{\partial}{\partial v}s,\tfrac{\partial}{\partial v}\nu\circ s\rangle.
\end{aligned}$$

Now suppose $z=f(x,y)$ be a local description of Σ in the tangent-normal coordinates at p. Note that

$$s(u,v)=(u,v,f(u,v))$$

describes a chart of Σ at p.

Denote by I, J and K the standard basis in the (x,y,z)-coordinates. Note that $s(0,0)=p$ and

$$\tfrac{\partial}{\partial u}s(0,0)=\mathrm{I}, \qquad \tfrac{\partial}{\partial v}s(0,0)=\mathrm{J}, \qquad \nu\circ s(0,0)=\mathrm{K},$$

In particular $D_{\mathrm{I}}\nu=\frac{\partial}{\partial u}\nu\circ s(0,0)$ and $D_{\mathrm{J}}\nu=\frac{\partial}{\partial v}\nu\circ s(0,0)$. Further,

$$\tfrac{\partial^2}{\partial u^2}s(0,0)=\ell\cdot\mathrm{K}, \qquad \tfrac{\partial^2}{\partial v\partial u}s(0,0)=m\cdot\mathrm{K}, \qquad \tfrac{\partial^2}{\partial v^2}s(0,0)=n\cdot\mathrm{K},$$

where ℓ, m, and n are the components of the Hessian matrix of f at $(0,0)$ defined on page 69.

Evaluating the above 6 identities at $(u,v)=(0,0)$, we get that

$$\begin{aligned}
&\langle -D_{\mathrm{I}}\nu,\mathrm{I}\rangle=\ell, &&\langle -D_{\mathrm{I}},\mathrm{J}\rangle=m, &&\langle -D_{\mathrm{I}}\nu,\mathrm{K}\rangle=0,\\
&\langle -D_{\mathrm{J}}\nu,\mathrm{I}\rangle=m, &&\langle -D_{\mathrm{J}}\nu,\mathrm{J}\rangle=n, &&\langle -D_{\mathrm{J}}\nu,\mathrm{K}\rangle=0,
\end{aligned}$$

That is, $-D_{\mathrm{I}}\nu$ and $-D_{\mathrm{J}}\nu$ satisfy the same equalities as $S(\mathrm{I})$ in 7.3. Note that these equalities define S completely; therefore ❷ follows. □

More curvatures

Fix an oriented smooth surface Σ and a point $p \in \Sigma$.

The product

$$K(p) = k_1(p)\cdot k_2(p)$$

is called Gauss curvature at p. We may denote it by $K(p)_\Sigma$ if we need to emphasize that this is curvature of Σ. The Gauss curvature can be also interpreted as the determinant of the Hessian matrix M_p, or, equivalently, as the determinant of the shape operator S_p.

The sum

$$H(p) = k_1(p) + k_2(p)$$

is called *mean curvature*[1] at p. We may denote it by $H(p)_\Sigma$ if we need to emphasize that this is curvature of Σ. The mean curvature can be also interpreted as the trace of the Hessian matrix M_p, or, equivalently, as the trace of the shape operator S_p.

Note that the Gauss curvature depends only on Σ and p. The same is true up to sign for the mean curvature — it changes the sign if we revert the orientation of the surface.

7.7. Exercise. *Show that any surface with positive Gauss curvature is orientable.*

Curve 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 can assume that γ is unit-speed; in this case $\mathrm{T}(s) = \gamma'(s)$ is its tangent indicatrix. Let us use a 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. Since $\mathrm{T}(s) \perp \nu(s)$, the vector $\mu(s)$ is tangent at $\gamma(s)$. In fact $\mu(s)$ is a counterclockwise rotation of $\mathrm{T}(s)$ by right angle in $\mathrm{T}_{\gamma(s)}$; it can be also defined as a vector product $\mu(s) = \nu(s) \times \mathrm{T}(s)$.

Recall 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,

$$\gamma''(s) = k_g(s)\cdot\mu(s) + k_n(s)\cdot\nu(s).$$

---

[1] Some authors define mean curvature as $\frac{1}{2}\cdot(k_1(p) + k_2(p))$ — the mean value of the principle curvatures. It is suits the name better, but not as convenient when it comes to calculations.

The values $k_g(s)$ and $k_n(s)$ are called *geodesic* and *normal curvature* of γ at s. 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 7.4, we get the following:

7.8. Proposition. *Assume γ is a smooth unit-speed curve in a smooth surface Σ. Suppose $p = \gamma(s_0)$ and $\mathrm{V} = \gamma'(s_0)$. Then*

$$k_n(s_0) = \langle S_p(\mathrm{V}), \mathrm{V} \rangle,$$

where k_n denotes the normal curvature of γ at s_0 and S_p is the shape operator at p.

Note that according to the proposition, the normal curvature of 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}; that is,

$$k_{\mathrm{V}} = \langle S_p(\mathrm{V}), \mathrm{V} \rangle$$

if V is a unit vector in T_p.

Let p be a point on a smooth surface Σ. Assume we choose the tangent-normal coordinates at p so that the Hessian matrix is diagonalized, we can assume that

$$M_p = \begin{pmatrix} k_1(p) & 0 \\ 0 & k_2(p) \end{pmatrix}.$$

Consider a vector $\mathrm{W} = \left(\begin{smallmatrix} a \\ b \end{smallmatrix}\right)$ in the (x, y)-plane. Then

$$\begin{aligned} \langle S(\mathrm{W}), \mathrm{W} \rangle &= \langle M_p \cdot \left(\begin{smallmatrix} a \\ b \end{smallmatrix}\right), \left(\begin{smallmatrix} a \\ b \end{smallmatrix}\right) \rangle = \\ &= a^2 \cdot k_1(p) + b^2 \cdot k_2(p). \end{aligned}$$

If W is unit, then $a^2 + b^2 = 1$ which implies the following:

7.9. Observation. *For any point p on an oriented smooth surface Σ, the principle curvatures $k_1(p)$ and $k_2(p)$ are correspondingly 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 principle direction at p, then*

$$k_{\mathrm{w}}(p) = k_1(p)\cdot \cos^2\theta + k_2(p)\cdot \sin^2\theta.$$

The last identity is called *Euler's formula.*

7.10. Meusnier's theorem. *Let γ be a regular smooth curve that runs in a smooth oriented surface Σ. Suppose $p = \gamma(t_0)$ and $\mathrm{v} = \gamma'(t_0)$ and $\alpha = \angle(\nu(p), \mathrm{N}(t_0))$; 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 for the curvature $\kappa(t_0)$ and the normal curvature $k_n(t_0)$ of γ at t_0:*

$$\kappa(t_0)\cdot \cos\alpha = k_n(t_0).$$

Proof. Since $\mathrm{T}' = \kappa\cdot\mathrm{N}$, 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 are proved by Jean Baptiste Meusnier [19].

7.11. Exercise. *Let Σ be a smooth surface, $p \in \Sigma$ and $\mathrm{v} \in \mathrm{T}_p\Sigma$ is a unit vector. Assume that $k_{\mathrm{v}}(p) \neq 0$; 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 with center $p + \frac{1}{k_{\mathrm{v}}}\cdot\nu$ and radius $\frac{1}{|k_{\mathrm{v}}|}$.

7.12. Exercise. *Let $\gamma(s) = (x(s), y(s))$ be a smooth unit-speed simple plane curve in the upper half-plane. Suppose that Σ is the surface of revolution of γ with respect to the x-axis.*

(a) *Show that parallels and meridians form lines of curvature on Σ.*

(b) *Show that*

$$\frac{|x'(s)|}{y(s)} \quad \textit{and} \quad \frac{-y''(s)}{|x'(s)|}$$

are principle curvatures of Σ at $(x(s), y(s), 0)$ in the direction of parallel and meridian correspondingly.

Lagunov's example

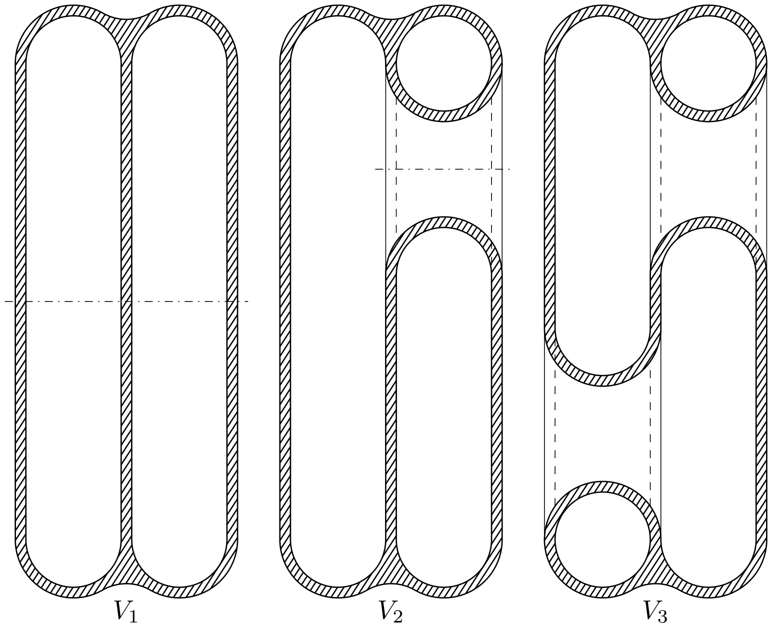
7.13. Exercise. *Assume V is a body in $\mathbb{R}^3$ bounded by a smooth surface of revolution with principle curvatures at most 1 in absolute value. Show that V contains a unit ball.*

The following question is a 3-dimensional analog of the moon in a puddle problem (5.14).

7.14. Question. *Assume a set $V \subset \mathbb{R}^3$ is bounded by a closed connected surface Σ with principle curvatures bounded in absolute value by 1. Is it true that V contains a ball of radius 1?*

According to 7.13, the answer is "yes" for surfaces of revolution. We also have "yes" for convex surfaces; see 8.6. It turns out that in general the answer is "no". The following example was constructed by Vladimir Lagunov [20].

Construction. Let us start with a body of revolution V_1 with cross section shown on the diagram. The boundary curve of the cross section consists of 6 long vertical line segments smoothly jointed into 3 closed simple curves. The boundary of V_1 has 3 components, each of which is a smooth sphere.



We assume that the curvature of the curves have curvature at most 1. Moreover with the exception of vertical segments, the curve have absolute curvature near 1 almost all time. The only thick part in V is the place where all three boundary components come together; the remaining part of V is assumed to be very thin. It could be arranged that the radius of the maximal ball in V is just a little bit above the value

$$r_2 = \tfrac{2}{\sqrt{3}} - 1 < \tfrac{1}{6}.$$

The value r_2 is the radius of the smallest circle tangent to three unit circles that are tangent to each other.

Exercise 7.12 gives formulas for the principle curvatures of the boundary of V; note that both of them are at most 1 by absolute value.

It remains to modify V_1 to make its boundary connected without alloing larger balls inside.

Note that each sphere in the boundary contains two flat discs; they come into pairs close 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; its 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; the cross section of the obtained body is shown on the diagram.

Note that the boundary of V_3 is connected. Assuming that the holes are large, its boundary can be made so that its principle curvatures are still at most 1; the latter can be proved the same way as for V_1. □

Note that the surface of V_3 in the Lagunov's example has genus 2; that is it can be parameterized 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 shpere. We reconnect two spheres by a tube and obtain one sphere. Connecting the two holes of the other sphere by a tube we get a torus. 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 torus by a tube. This way we get a sphere with two handles; that is, it has genus 2.

7.15. Exercise. *Modify Lagunov's construction to make the boundary surface a sphere with 4 handles.*

Variations*

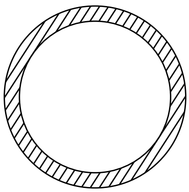
In this section we will discuss few results related to 7.14. Recall that r_2

is the radius of the smallest circle tangent to three unit circles that are tangent to each other. 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

$$r_2 = \tfrac{2}{\sqrt{3}} - 1 < \tfrac{1}{6} \quad \text{and} \quad r_3 = \sqrt{\tfrac{3}{2}} - 1 > \tfrac{1}{5}.$$

The following example shows that the bound obtained in the construction of the Lagunov's example is optimal.

7.16. Advanced exercise. *Suppose a connected body $V \subset \mathbb{R}^3$ is bounded by a finite number of closed smooth surfaces with principle 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 components of the same topological type; that is, both can be written in parametric form with the same parameter domain.*



For example consider the region between two large concentric spheres with almost equal radiuses. This region can be made arbitrary thin and the curvature of the boundary can be made arbitrary close to zero.

The same example works in the plane — a pair of circles with arbitrary small curvature can bound an arbitrary thin region.

In fact one can show that if a body $V \subset \mathbb{R}^3$ is bounded by a sphere Σ with principle curvatures at most 1, then V contains a ball of radius r_3, which is the radius of the smallest sphere tangent to three unit spheres that are tangent to each other. Moreover, this bound is optimal.

An example of such a body can be obtained by thickening a smoothed version of the so called Bing's house. It is a certain surface whose singularities are formed by three curves meeting at two points; four ends at each point. At the singular curves the three pieces of surface have to cross at angles $\frac{2}{3}\cdot\pi$ and at the sigular points 6 pieces of surface should come together forming 6 triangles with vertex in the center of a regular tetrahedron and the bases at its 6 edges. Thickening of a sufficiently large smooth Bing's house of that type produces the optimal bound r_3 on the maximal ball that it contains.

Chapter 8

Supporting surfaces

Definitions

Assume two orientable 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 a tangent-normal coordinates at p. If Σ_2 *locally supports* Σ_1 at p, then we may assume that all points of Σ_1 near p lie above the graph $z = f_2(x, y)$. In particular the tangent plane of Σ_1 at p is horizontal; that is the tangent planes of Σ_1 and Σ_2 at p coincide.

It follows that, we can assume that Σ_1 and Σ_2 are cooriented at p; that is, they have common unit normal vector at p. If not, we can revert the orientation of one of the surfaces.

If Σ_2 locally supports Σ_1 and 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 surface Σ_2 in U.

More precisely, we can use for Σ_1 and Σ_2 one tangent-normal coordinate system at p. This way we write Σ_1 and Σ_2 locally as graphs: $z = f_1(x, y)$ and $z = f_2(x, y)$ correspondingly. Then Σ_1 locally supports Σ_2 from inside (from outside) if $f_1(x, y) \geqslant f_2(x, y)$ (correspondingly $f_1(x, y) \leqslant f_2(x, y)$) for (x, y) in a sufficiently small neighborhood of the origin.

Note that Σ_1 locally supports Σ_2 from inside at the point p is equivalent to Σ_2 locally supports Σ_1 from outside. Further if we revert the orientation of both surfaces, then supporting from inside becomes supporting from outside and the other way around.

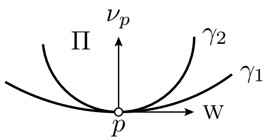
8.1. Proposition. *Let Σ_1 and Σ_2 be oriented surfaces. Assume Σ_1 locally supports Σ_2 from inside at the point p (equivalently Σ_2 locally* *supports Σ_1 from outside). Then*

$$k_1(p)_{\Sigma_1} \geqslant k_1(p)_{\Sigma_2} \quad \textit{and} \quad k_2(p)_{\Sigma_1} \geqslant k_2(p)_{\Sigma_2}.$$

8.2. Exercise. *Give an example of two surfaces Σ_1 and Σ_2 that have common point p with 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 a 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 therefore cooriented. Note that in this case the curve γ_1 supports the curve γ_2 from the right.

By 5.3, we have the following inequality for the normal curvatures of Σ_1 and Σ_2 at p in the direction of w:

❶ $$k_{\mathrm{w}}(p)_{\Sigma_1} \geqslant k_{\mathrm{w}}(p)_{\Sigma_2}.$$

According to 7.9,

$$k_1(p)_{\Sigma_i} = \min\{\, k_{\mathrm{w}}(p)_{\Sigma_i} \;:\; \mathrm{w} \in \mathrm{T}_p, |\mathrm{w}| = 1 \,\}$$

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

$$\begin{aligned} k_1(p)_{\Sigma_1} &= k_{\mathrm{w}}(p)_{\Sigma_1} \geqslant \\ &\geqslant k_{\mathrm{w}}(p)_{\Sigma_2} \geqslant \\ &\geqslant \min\{\, k_{\mathrm{w}}(p)_{\Sigma_2} \,\} = \\ &= k_1(p)_{\Sigma_2}; \end{aligned}$$

that is, $k_1(p)_{\Sigma_1} \geqslant k_1(p)_{\Sigma_2}$.

Similarly, by 7.9, we have that

$$k_2(p)_{\Sigma_i} = \max\{\, k_{\mathrm{w}}(p)_{\Sigma_i} \,\}.$$

Let us fix W so that $k_2(p)_{\Sigma_2} = k_{\mathrm{W}}(p)_{\Sigma_2}$. Then

$$\begin{aligned} k_2(p)_{\Sigma_2} &= k_{\mathrm{W}}(p)_{\Sigma_2} \leqslant \\ &\leqslant k_{\mathrm{W}}(p)_{\Sigma_1} \leqslant \\ &\leqslant \max\{\, k_{\mathrm{W}}(p)_{\Sigma_1} \,\} = \\ &= k_2(p)_{\Sigma_1}; \end{aligned}$$

that is, $k_2(p)_{\Sigma_1} \geqslant k_2(p)_{\Sigma_2}$. □

8.3. 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. By 8.1, we get that $k_1(p)_{\Sigma_1} \geqslant k_1(p)_{\Sigma_2}$ and $k_2(p)_{\Sigma_2} \geqslant k_2(p)_{\Sigma_2}$. Therefore part (a) follows since

$$H(p)_{\Sigma_i} = k_1(p)_{\Sigma_i} + k_2(p)_{\Sigma_i}.$$

(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 principle 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 8.1, it implies that

$$\begin{aligned} K(p)_{\Sigma_1} &= k_1(p)_{\Sigma_1} \cdot k_2(p)_{\Sigma_1} \geqslant \\ &\geqslant k_1(p)_{\Sigma_2} \cdot k_2(p)_{\Sigma_2} = \\ &= K(p)_{\Sigma_2}. \end{aligned}$$

□

8.4. Exercise. *Show that any closed surface in a unit ball has a point with Gauss curvature at least 1.*

Convex surfaces

A proper surface without boundary that bounds a convex region is called *convex.*

8.5. Exercise. *Show that Gauss curvature of any convex smooth surface is nonnegative at each point.*

8.6. Exercise. *Assume R is a convex body in $\mathbb{R}^3$ bounded by a surface with principle 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 convex set which is not strictly convex. It is easy to see that a convex region is strictly convex if and only if its boundary does not contain a line segment.

8.7. Lemma. *Let $z = f(x, y)$ be the local description of a smooth surface Σ in a tangent-normal coordinates at some point $p \in \Sigma$. Assume both principle 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 principle curvatures are positive, we have that

$$D^2_{\mathrm{w}} f(0, 0) = \langle S_p(\mathrm{w}), \mathrm{w}\rangle > 0$$

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 $D^2_{\mathrm{w}} f(x, y) > 0$ for (x, y) in a neighborhood of the origin. That is, f is a strictly convex function in a neighborhood of the origin in the (x, y)-plane.

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. $\square$

8.8. 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 is a curve or a single point.*

The following theorem gives a global description of surfaces with positive Gauss curvature.

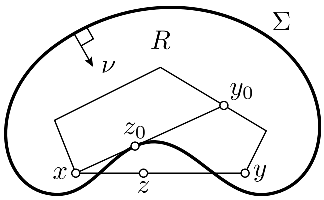
8.9. Theorem. *Assume Σ is an open or closed smooth surface with positive Gauss curvature. Then Σ bounds a strictly convex region.*

Note that in the proof we have to use that 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 principle curvatures are positive at any point. Let R be the region bounded by Σ that lies on the side of ν; that is, ν points inside of R at any point of Σ.

Fix $p \in \Sigma$; let $z = f(x, y)$ be a local description of Σ in the tangent-normal coordinates at p. By 8.7, 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. In other words, 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.

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, $R \cap B(z_0, \varepsilon)$ 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. $\square$

Note that the proof above imply the following general statement: *any connected locally convex region is convex.*

8.10. Exercise. *Assume that a closed surface Σ surrounds a unit circle. Show that Gauss curvature of Σ is at most 1 at some point.*

8.11. Exercise. *Let Σ be a closed 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 arbitrarily small Gauss curvature at most 1.*

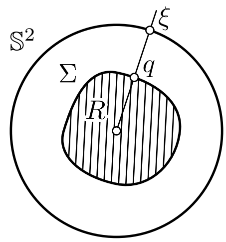
8.12. Theorem. *Suppose Σ is a smooth convex surface.*

(a) If Σ is compact then it is a smooth sphere; that is, Σ admits a smooth regular parametrization by $\mathbb{S}^2$.

(b) If Σ is open then there is a coordinate system such that Σ is a graph $z = f(x, y)$ of a convex function f defined on a convex open region of the (x, y)-plane.

The following exercises will guide you thru the proof of both parts of the theorem.

8.13. Exercise. *Assume a convex compact region R contains the origin in its interior and bounded by a smooth surface Σ.*

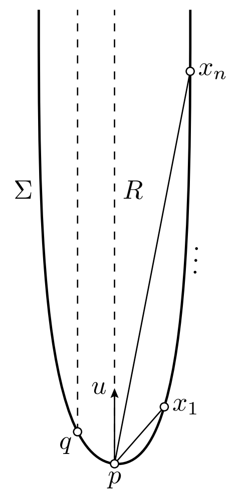


(i) Show that any half-line that starts at the origin intersects Σ at a single point; that is, there is a positive function $\rho\colon \mathbb{S}^2 \to \mathbb{R}$ such that Σ is formed by points $q = \rho(\xi)\cdot\xi$ for $\xi \in \mathbb{S}^2$.

(ii) Show that $\rho\colon \mathbb{S}^2 \to \mathbb{R}$ is a smooth function.

(iii) Conclude that $\xi \mapsto \rho(\xi)\cdot\xi$ is a smooth regular parametrization $\mathbb{S}^2 \to \Sigma$.

8.14. Exercise. *Assume a strictly convex closed noncompact region R contains the origin in its interior and bounded by a smooth surface Σ.*



(i) Show that R contains a half-line ℓ.

(ii) Show that any line parallel to ℓ intersects Σ at most at one point.

(iii) Consider (x, y, z)-coordinate system such that the z-axis points in the direction of ℓ. Show that projection of Σ to the (x, y) plane is an open convex set; denote it by Ω.

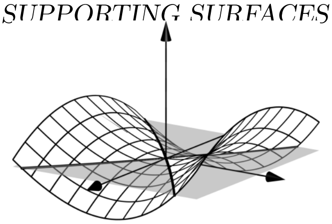
(iv) Conclude that Σ is a graph $z = f(x, y)$ of a convex function f defined on Ω.

8.15. 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.

8.16. Exercise. *Show that any open smooth surface Σ with positive Gauss curvature lies inside of an infinite circular cone.*

Saddle surfaces



A surface is called *saddle* if its Gauss curvature at each point is nonpositive; in other words principle curvatures at each point have opposite signs or one of them is zero.

If the Gauss curvature is negative at each point, then the surface is called *strictly saddle*; equivalently it means that the principle curvatures have opposite signs at each point. Note that in this case the tangent plane does not support the surface even locally — moving along the surface in the principle directions at a given point, one goes above and below the tangent plane at this point.

8.17. 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 $\ell_p\subset\Sigma_p$ thru p that is infinite or has its endpoint(s) on the boundary line of Σ.

8.18. Exercise. *Show that any ruled surface Σ is saddle.*

8.19. Exercise. *Suppose Σ is an open saddle surface. Show that for any point $p\in\Sigma$ there is a curve $\gamma\colon[0,\infty)\to\Sigma$ that starts at p and monotonically escapes to infinity; that is, the function $t\mapsto|\gamma(t)|$ is increasing and $|\gamma(t)|\to\infty$ as $t\to\infty$.*

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*.

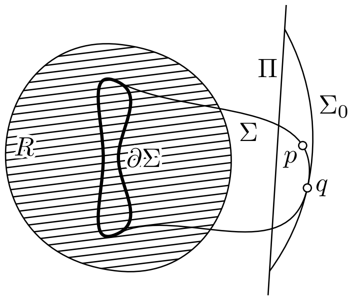
8.20. Advanced exercise. *Let $\Sigma\subset\mathbb{R}^3$ be the graph $z=f(x,y)$ of a smooth function f and γ be a closed smooth asymptotic line in Σ. Assume Σ is strictly saddle in a neighborhood of γ. Show that the projection of γ to the (x,y)-plane cannot be star-shaped.*

Hats

Note that a closed surface cannot be saddle. Indeed consider a smallest sphere that contains a closed surface Σ inside; it supports Σ at some point p and at this point the principle curvature must have the same sign. The following more general statement is proved using the same idea.

8.21. Lemma. *Assume Σ is a compact saddle surface and its boundary line lies in a convex closed region R. Then whole surface Σ lies in R.*

Proof. Assume contrary; that is, there is point $p \in \Sigma$ that does not lie in R. Let Π be a plane that separates p from R; it exists by A.22. Denote by Σ' the part of Σ that lies with p on the same side from 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 principle curvatures. In this case Σ_0 supports Δ from outside. By 8.3, $K(q)_\Sigma \geqslant K(q)_{\Sigma_0} > 0$ — a contradiction. $\square$

Note that if we assume that Σ is strictly saddle, then we could arrive to a contradiction by taking a point $q \in \Sigma$ on the maximal distance from R.

8.22. Exercise. *Let Δ be a smooth regular saddle disc and $p \in \Delta$. Assume that the boundary line $\partial\Delta$ lies in the unit sphere centered at p. Show that* $\operatorname{length}(\partial\Delta) \geqslant 2\cdot\pi$*.*

If Δ is as in the exercise, then in fact $\operatorname{area}\Delta \geqslant \pi$. The proof of this statement can be obtained by applying the so called *coarea fromula* together to with the inequality in the exercise.

8.23. 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 the remaining points of Δ lie on one side of Π.

8.24. 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 centered at the origin. However, a plane curve cannot bound a disc (as well any compact set) in a saddle surface.

Proof. Since plane is a convex set, the "only if" part follows from 8.21; 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 principle curvatures $k_1(p)$ and $k_2(p)$ have the same sign.

Let $z = f(x, y)$ be a graph representation of Σ in the tangent-normal coordinates at p. By 8.7, f is convex in a small neighborhood of $(0, 0)$. In particular the set F_ε defined by the inequality $f(x, y) \leqslant \varepsilon$ is a convex set for sufficiently small $\varepsilon > 0$; in particular it 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 whole disc lies below it; that is, Δ_ε is a hat of Σ. $\square$

The following exercise shows that Δ_ε is in fact a smooth disc. It can be used to prove slightly stronger version of 8.24; namely one can the disc in the definition of hat is a smooth disc.

8.25. 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 defined by the inequality $f(x, y) \leqslant \varepsilon$ is a smooth disc for any $\varepsilon > 0$; that is, there is a diffeomorphism $\mathbb{D} \to F_\varepsilon$, where $\mathbb{D} = \left\{ (x, y) \in \mathbb{R}^2 \ : \ x^2 + y^2 \leqslant 1 \right\}$ is the unit disc.*

8.26. Exercise. *Let $T\colon \mathbb{R}^3 \to \mathbb{R}^3$ be a linear transformation, that is T is an invertable map $\mathbb{R}^3 \to \mathbb{R}^3$ that sends any plane to a plane. Show that for any saddle surface Σ the image $T(\Sigma)$ is also a saddle surface.*

Saddle graphs

The following theorem was proved by Sergei Bernstein [21].

8.27. 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 $(x, y) \in \mathbb{R}^2$.*

Before going into the proof let us discuss some examples.

Note that the theorem states that a saddle graph cannot lie between parallel horizontal planes; applying 8.26 we get that saddle graphs cannot lie between parallel planes, not necessarily horizontal.

The following exercise shows that the theorem does not hold for saddle surfaces which are not graphs.

8.28. Exercise. *Construct an open strictly saddle surface that lies between parallel planes.*

The following exercise shows that there are saddle graphs with functions bounded on one side; that is, both (upper and lower) bounds are needed in the proof of Bernshtein's theorem.

8.29. Exercise. *Show that there are positive functions with strictly saddle graphs. In fact the graph* $z = \exp(x - y^2)$ *is strictly saddle.*

Note that according to 8.21, there are no proper saddle surfaces in a parallelepiped that boundary line lies on one of its faces. The following lemma gives an analogous statement for a parallelepiped with an infinite side.

8.30. Lemma. *There is no proper strictly saddle smooth surface that lies on bounded distance from a line and has its boundary line in a plane.*

Proof. Note that in a suitable coordinate system, the statement can be reformulated the following way: *There is no proper strictly saddle smooth surface with the boundary line in the* (x, y)*-plane that lies in a region of the following form:*

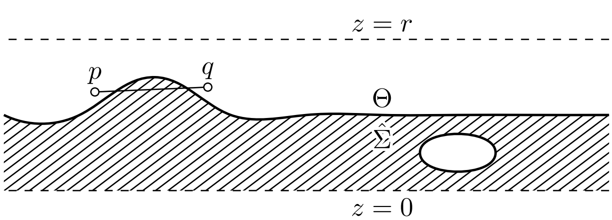
$$R = \left\{ (x, y, z) \in \mathbb{R}^3 \, : \, 0 \leqslant z \leqslant r, 0 \leqslant y \leqslant r \right\}.$$

Let us prove this statement.

Assume 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 $z = r$.

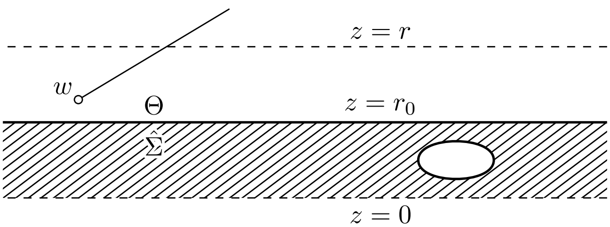
Consider the open upper half-plane $H = \left\{ (x, z) \in \mathbb{R}^2 \, : \, z > 0 \right\}$. Let Θ be the connected component of the complement $H \backslash \hat{\Sigma}$ that contains all the points above the line $z = r$.

Note that Θ is convex. If not, then there is a line segment $[pq]$ for some $p, q \in \Theta$ that cuts from $\hat{\Sigma}$ a compact piece. Consider the plane



Π thru $[pq]$ that is perpendicular to the (x, z)-plane. Note that Π cuts from Σ a compact region Δ. By general position argument (see 6.18) 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 8.21.

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 point up from it. Whence it contain all points with z-coordinate larger than the z-coordinate of w. Since Θ is open it can be described by inequality $z > r_0$. It follows that

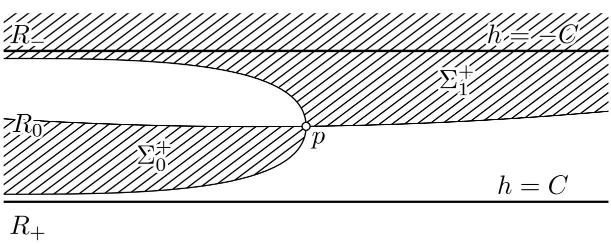


the plane $z = r_0$ supports Σ at some point (in fact at many points). By 8.1, the latter is impossible — a contradiction. $\square$

Proof of 8.27. Denote by Σ the graph $z = f(x, y)$. Assume 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 principle direction with positive principle curvature. Indeed if there would be 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 contradicts 8.21.



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.

Moving the plane T_p slightly upward, we can cut from Σ_0^+ a proper surface with boundary line lying in this plane (see 6.18). The obtained surface is still on a bounded distance to a line which is impossible by 8.30. □

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.

8.31. Advanced exercise. *Let Σ be a smooth saddle disk in $\mathbb{R}^3$. Assume that the orthogonal projection to the (x,y)-plane maps the boundary line of Σ injectively to a convex closed curve. Show that the orthogonal projection to the (x,y)-plane is injective on Σ.*

In particular, Σ is the graph $z = f(x,y)$ of a function f defined on a convex figure in the (x,y)-plane.

Remarks

Note that Bernstein's theorem and the lemma in its proof do not hold for 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 more technical.

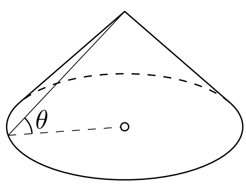
By 8.24, saddle surfaces can be defined as smooth surfaces without hats. This definition can be used for arbitrary surfaces, not necessarily smooth. Some results, for example Bernshtein's characterization of saddle graphs can be extended to generalized saddle surfaces, but this class of surfaces is far from being understood. Some nontrivial properties were proved by Samuil Shefel [22]; see also [23, Capter 4].

Chapter 9

Geodesics

We start to study the intrinsic geometry of surfaces. The following exercise should help you to be in the right mood for this; it might look like a tedious problem in calculus, but actually it is an easy problem in geometry.

9.1. Exercise. *There is a mountain of frictionless ice with the shape of a perfect cone with a circular base. A cowboy is at the bottom and he wants to climb the mountain. So, he throws up his lasso which slips neatly over the top of the cone, he pulls it tight and starts to climb. If the angle of inclination θ is large, there is no problem; the lasso grips tight and up he goes. On the other hand if θ is small, the lasso slips off as soon as the cowboy pulls on it.*



What is the critical angle θ_0 at which the cowboy can no longer climb the ice-mountain?

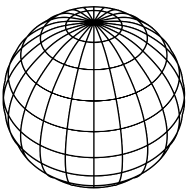
Shortest paths

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 exact lower bound on 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 of points $p, q \in \Sigma$.

A path γ from p to q in Σ that minimizes the length is called a *shortest path* from p to q.

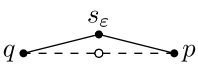
The image of a shortest path between p and q in Σ is usually denoted by $[p,q]_\Sigma$. In general there might be no shortest path between two given points on the surface and it might be many of them; this is shown in the following two examples. However if we write $[p,q]_\Sigma$, then we assume that a shortest path exists and we made a choice of one of them.



Nonuniqueness. There are plenty of shortest paths between the poles on the sphere — each meridian is a shortest path.

Nonexistence. Let Σ be the (x,y)-plane with removed origin. Consider two points $p=(1,0,0)$ and $q=(-1,0,0)$ in Σ.

Note that $|p-q|_\Sigma = 2$. Indeed, given $\varepsilon > 0$, consider the point $s_\varepsilon = (0,\varepsilon,0)$. Note that the polygonal path $ps_\varepsilon q$ lies in Σ and its length $2\cdot\sqrt{1+\varepsilon^2}$ approaches 2 as $\varepsilon \to 0$. It follows that $|p-q|_\Sigma \leqslant 2$. On the other hand $|p-q|_\Sigma \geqslant |p-q|_{\mathbb{R}^3} = 2$; that is, $|p-q|_\Sigma = 2$.



Therefore a shortest path from p to q (if it exists) must have length 2. By 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

9.2. Proposition. *Any two points in a proper smooth surface can be joined by a shortest path.*

Proof. Fix a proper smooth surface Σ with two points p and q. Set $\ell = |p-q|_\Sigma$.

By the definition of induced length metric, 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 parameterized 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 $t_0, t_1 \in [0,1]$. Further 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 A.17, we can pass to a converging subsequence of γ_n; denote by $\gamma_\infty\colon [0,1] \to \mathbb{R}^3$ its limit. As a limit of uniformly continuous sequence, γ_∞ is continuous; that is, γ_∞ is a path. Evidently γ_∞ runs from p to q. Since Σ is a closed set, γ_∞ lies in Σ. Finally, by 2.13,

$$\gamma_\infty \leqslant \ell;$$

that is, γ_∞ is a shortest path from p to q. $\square$

Closest point projection

9.3. 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 $|p - x|$ among all points $x \in R$.*

Moreover the map $p \mapsto \bar{p}$ is short; that is,

❶ $$|p - q| \geqslant |\bar{p} - \bar{q}|$$

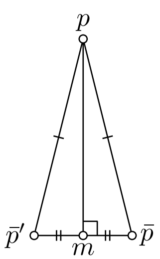
for any pair of points $p, q \in \mathbb{R}^3$.

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 $n \to \infty$.



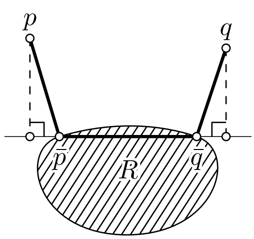
Without loss of generality, we can assume that all the points x_n lie in a ball of radius $\ell + 1$ centered at p. Therefore we can pass to a partial limit $\bar{p}$ of x_n; that is, $\bar{p}$ is a limit of a subsequence of x_n. Since R is closed $\bar{p} \in R$. By construction

$$\begin{aligned} |p - \bar{p}| &= \lim_{n \to \infty} |p - x_n| = \\ &= \ell. \end{aligned}$$

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 $m = \tfrac{1}{2} \cdot (\bar{p} + \bar{p}')$ lies in R. Note that $|p - \bar{p}| = |p - \bar{p}'| = \ell$; that is, $\triangle p\bar{p}\bar{p}'$ is isosceles and therefore $\triangle p\bar{p}m$ 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 inequality ❶.



We can assume that $\bar{p} \neq \bar{q}$, otherwise there is nothing to prove. Note that if $p \neq \bar{p}$ (that is, if $p \notin R$), then $\angle p\bar{p}\bar{q}$ is right or obtuse. Otherwise there would be a point x on the line segment $[\bar{q}, \bar{p}]$ that is closer to p than $\bar{p}$. Since R is convex, the line segment $[\bar{q}, \bar{p}]$ and therefore x lie in R. Hence $\bar{p}$ is not closest to p — a contradiction.

The same way we can show that if $q \neq \bar{q}$, then $\angle q\bar{q}\bar{p}$ is right or obtuse.

We have to consider the following 4 cases: (1) $p \neq \bar{p}$ and $q \neq \bar{q}$, (2) $p = \bar{p}$ and $q \neq \bar{q}$, (3) $p \neq \bar{p}$ and $q = \bar{q}$, (4) $p = \bar{p}$ and $q = \bar{q}$. In all these cases the obtained angle estimates imply 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

$$|p - q| \geqslant |\bar{p} - \bar{q}|.$$

□

9.4. 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 for any curve γ in W that runs from p to q we have*

$$\operatorname{length} \gamma \geqslant |p - q|_\Sigma.$$

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

$$\operatorname{length} \bar{\gamma} \geqslant |p - q|_\Sigma.$$

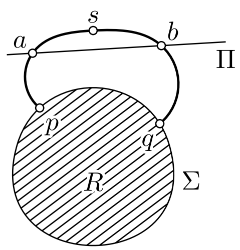
Consider an inscribed polygonal line $p_0 \dots p_n$ in γ. Denote by $\bar{p}_i$ the closest point projection of p_i to R. Note that the polygonal line $\bar{p}_0 \dots \bar{p}_n$ is inscribed in $\bar{\gamma}$; moreover any inscribed polygonal line in $\bar{\gamma}$ can appear this way. By 9.3 $|p_i - p_{i-1}| \geqslant |\bar{p}_i - \bar{p}_{i-1}|$ for any i. Therefore

$$\operatorname{length} p_0 \dots p_n \geqslant \operatorname{length} \bar{p}_0 \dots \bar{p}_n.$$

Taking least upper bound of each side of the inequality for all inscribed polygonal lines $p_0 \dots p_n$ in γ, we get

$$\operatorname{length} \gamma \geqslant \operatorname{length} \bar{\gamma}.$$

Whence the first statement follows.



To prove the second statement, note that if $s = \gamma(t_1) \notin \Sigma$, then $s \notin R$. Hence there is a plane Π that cuts s from Σ. The curve γ must intersect at least 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 on the way γ visits s that is not on the plane Π and therefore not on the lie segment $[a, b]$.

Remove from γ the arc from a to b and glue in the line segment $[a, b]$; denote the obtained curve by γ_1. From above,

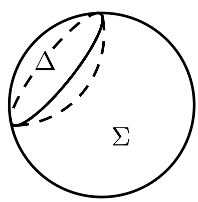
$$\operatorname{length}\gamma > \operatorname{length}\gamma_1$$

Note that γ_1 runs in W. Therefore by the first part of corollary, we have

$$\operatorname{length}\gamma_1 \geqslant |p - q|_\Sigma.$$

Whence the second statement follows. □

9.5. 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 with respect to 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 geodesic lies in Δ.*



Let us define the *intrinsic diameter* of a closed surface Σ as the exact upper bound on the lengths of shortest paths in the surface.

9.6. Exercise. *Assume that a closed smooth surface Σ with positive Gauss curvature lies in a unit ball. Show that the intrinsic diameter of Σ cannot exceed π.*

Geodesics

A smooth curve γ on a smooth surface Σ is called *geodesic* if its acceleration $\gamma''(t)$ is perpendicular to the tangent plane $\mathrm{T}_{\gamma(t)}$ for each t.

Geodesics can be understood as the trajectories of a particle that slides on Σ without friction. In this case the force that keeps the particle on Σ must be perpendicular to Σ. By the second Newton's laws of motion, we get that the acceleration γ'' is perpendicular to $\mathrm{T}_{\gamma(t)}$.

9.7. Exercise. *Assume that a smooth surface Σ is mirror symmetric with respect to a plane Π. Suppose that Σ and Π intersect along a curve γ. Show that γ is a geodesic of Σ.*

9.8. Exercise. *Show that the helix*

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

is a geodesic on the cylindrical surface described by the equation $x^2 + y^2 = 1$.

Recall that asymptotic line is defined on page 86.

9.9. Exercise. *Show that if a curve that is a geodesic and an asymptotic line on a smooth surface, then the curve is a line segment.*

9.10. Lemma. *Any geodesic γ has constant speed; that is, $|\gamma'(t)|$ is constant.*

Proof. Since $\gamma'(t)$ is a tangent vector at $\gamma(t)$, we have that $\gamma''(t) \perp \perp \gamma'(t)$, or equivalently $\langle\gamma'', \gamma'\rangle = 0$ for any t. Whence

$$\begin{aligned}\langle\gamma', \gamma'\rangle' &= 2{\cdot}\langle\gamma'', \gamma'\rangle = \\ &= 0.\end{aligned}$$

That is, $|\gamma'(t)|^2 = \langle\gamma'(t), \gamma'(t)\rangle$ is constant. ☐

9.11. 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 $\gamma'(0) = v$.*

Moreover

(a) the map $(p, v, t) \mapsto \gamma(t)$ is smooth in its domain of definition.

(b) if Σ is proper, then $\mathbb{I} = \mathbb{R}$; that is, the maximal interval is whole real line.

Sketch of proof. The first part of the proposition and part (a) follows from existence and uniqueness of a solution of initial value problem (A.14). One only needs to rewrite the condition $\gamma''(t) \perp \mathrm{T}_{\gamma(t)}$ as a differential equation $\gamma''(t) = \mathrm{II}_{\gamma(t)}(\gamma'(t), \gamma'(t))$.

The part (b) follows from 9.10. Indeed by A.14, if the maximal interval is not whole real line, then the curve γ must escape to infinity. But the latter is impossible since γ runs with constant speed. ☐

Exponential map

Let Σ be smooth regular surface and $p \in \Sigma$. Given a tangent vector $v \in \mathrm{T}_p$ consider a geodesic γ_v in Σ that runs from p with the initial velocity v; that is, $\gamma(0) = p$ and $\gamma'(0) = v$.

The point $q = \gamma_v(1)$ is called *exponential map* of v, or briefly $q = \exp_p v$. (There is a reason to call this map *exponential*, but it will take us too far from the subject.) By 9.11, the map $\exp_p \colon \mathrm{T}_p \to \Sigma$ is smooth and defined in a neighborhood of zero in T_p; moreover, if Σ is proper, then $\exp_p$ is defined on the whole space T_p.

Note that the Jacobian of $\exp_p$ at zero is the identity matrix. Indeed, let $z = f(x,y)$ be a local graph representation of Σ in the tangent-normal coordinates. The tangent plane at p is the (x,y)-plane. Let γ_x and γ_y be the geodesics starting from p in the directions $(1,0,0)$ and $(0,1,0)$ correspondingly. A general tangent vector can be written as $v = (x,y,0)$. Note that $\frac{\partial \exp_p}{\partial x}(0,0) = \gamma_x'(0) = (1,0,0)$ and $\frac{\partial \exp_p}{\partial x}(0,0) = \gamma_y'(0) = (0,1,0)$. That is, the Jacobian matrix of $\exp_p$ at $(0,0)$ is

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

It follows that the Jacobian matrix of the projection of $\exp_p$ to the (x,y)-plane is the identity matrix. Therefore by the inverse function theorem (A.9), we get the following statement:

9.12. Proposition. *Let Σ be smooth surface and $p \in \Sigma$. Then the exponential map $\exp_p \colon \mathrm{T}_p \to \Sigma$ is a smooth regular parametrization of a neighborhood of p in Σ by a neighborhood of 0 in the tangent plane T_p.*

Moreover for any $p \in \Sigma$ there is $\varepsilon > 0$ such that for any $x \in \Sigma$ such that $|x - p|_\Sigma < \varepsilon$ the map $\exp_x \colon \mathrm{T}_x \to \Sigma$ is a smooth regular parametrization of the ε-neighborhood of x in Σ by the ε-neighborhood of zero in the tangent plane T_x.

Shortest paths are geodesics

9.13. Claim. *Let Σ be a smooth regular surface. Then any shortest path γ in Σ parameterized proportional to its 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, given a smooth surface Σ there is a positive function ρ on Σ such that if a *geodesic γ starts at $p \in \Sigma$ and has length at most $\rho(p)$, then it is a shortest path.*

A geodesic might not form a shortest path, but if this is the case, then it is called *minimizing geodesic*. Note that according to the claim, any shortest path is a reparametrization of a minimizing geodesic.

This claim provides connection between intrinsic geometry of the surface and its extrinsic geometry. This connection will be important later; in particular it will play the key role in the proof of the so-called *remarkable theorem* (11.19).

Intrinsic means that it can be expressed in terms of measuring things inside the surface, for example length of curves or angles between the curves that lie in the surface. Extrinsic means that we have to use ambient space in order to measure it.

For instance, a shortest path is an object of intrinsic geometry of a surface, while definition of geodesic is not intrinsic — it requires acceleration which needs the ambient space. Note that there is a smooth bijection between the cylinder $z = x^2$ and the plane $z = 0$ that preserves the lengths of all curves; in other words the cylinder can be *unfolded* on the plane. Such a bijection sends geodesics in the cylinder to geodesics on the plane and the other way around; however a geodesic on the cylinder might have nonvanishing second derivative while geodesics on the plane are straight lines with vanishing second derivative.

Informal sketch. The smoothness should be intuitively obvious; at least the curve should be twice differentiable otherwise it can be shortened.

Let us give an informal physical explanation why $\gamma''(t) \perp \mathrm{T}_{\gamma(t)}\Sigma$. One may think about the geodesic γ as of stable position of a stretched elastic thread that is forced to lie on a frictionless surface. Since it is frictionless, the force density $N(t)$ that keeps the geodesic 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 by T the tension. We can assume that γ has unit speed, In this case the net force from tension to the arc $\gamma_{[t_0,t_1]}$ is $T{\cdot}(\gamma'(t_1) - \gamma'(t_0))$. Hence the density of net force from tension at t is $F(t) = T{\cdot}\gamma''(t)$. According to the second Newton's law of motion, we have

$$F(t) + N(t) = 0;$$

which implies that $\gamma''(t)$ is perpendicular to $\mathrm{T}_{\gamma(t)}\Sigma$.

Fix a point $p \in \Sigma$. Let $\varepsilon > 0$ be as in 12.3. 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 if parameterized 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 $v, w \in \mathrm{T}_p$ such that $|v| < \varepsilon$, $|w| < \varepsilon$ and $q = \exp_p v = \exp_p w$. But according to 12.3, the exponential map is injective in ε-neighborhood of zero — a contradiction. $\square$

9.14. Exercise. *Show that two 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 contain an arc from p to q.*

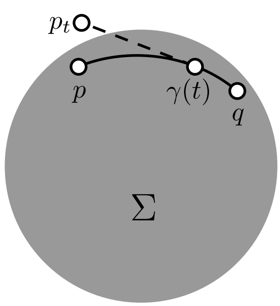
Show by example that nonoverlapping geodesics can cross each other an arbitrary number of times.

9.15. Exercise. *Assume that a smooth regular surface Σ is mirror symmetric with respect to a plane Π. Show that no shortest path in Σ can cross Π more than once.*

9.16. 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.

Liberman's lemma

The following lemma is a smooth analog of lemma proved by Joseph Liberman [26].

9.17. Liberman's lemma. *Assume γ is a geodesic on the graph $z = f(x,y)$ of a smooth convex function f defined on an open subset* *of the plane. Suppose that* $\gamma(t) = (x(t), y(t), z(t))$ *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, it has positive z-coordinate.

Since γ is a geodesic, we have $\gamma''(t) \perp \mathrm{T}_{\gamma(t)}$, or equivalently $\gamma''(t)$ is proportional to $\nu_{\gamma(t)}$ for any t. By **??**, we have

$$\langle \gamma''(t), \nu_{\gamma(t)} \rangle = \mathrm{II}_{\gamma(t)}(\gamma'(t), \gamma'(t));$$

hence

$$\gamma''(t) = \nu_{\gamma(t)} \cdot \mathrm{II}_{\gamma(t)}(\gamma'(t), \gamma'(t))$$

for any t.

Therefore

❷ $$z''(t) = \cos(\theta_\gamma(t)) \cdot \nu_{\gamma(t)} \cdot \mathrm{II}_{\gamma(t)}(\gamma'(t), \gamma'(t)),$$

where $\theta_\gamma(t)$ denotes the angle between $\nu_{\gamma(t)}$ and the z-axis.

Since ν points up, we have $\theta_\gamma(t) < \frac{\pi}{2}$, or equivalently

$$\cos(\theta_\gamma(t)) > 0$$

for any t.

Since f is convex, we have that tangent plane supports the graph from below at any point; in particular $\mathrm{II}_{\gamma(t)}(\gamma'(t), \gamma'(t)) \geqslant 0$. It follows that the right hand side in ❷ is nonnegative; whence the statement follows. □

9.18. Exercise. *Assume* γ *is a unit-speed geodesic on a smooth convex surface* Σ *and* p *in the interior of a convex set bounded by* Σ*. Set* $\rho(t) = |p - \gamma(t)|^2$*. Show that* $\rho''(t) \leqslant 2$ *for any* t*.*

Bound on total curvature

9.19. Theorem. *Assume* Σ *is a 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* $2 \cdot \ell$*.*

The above theorem was proved by Vladimir Usov [27], later David Berg [28] pointed out that the same proof works for geodesics in closed epigraphs of ℓ-Lipschitz functions which are not necessary convex; that is, sets of the type

$$W = \{\, (x, y, z) \in \mathbb{R}^3 \,:\, z \geqslant f(x, y) \,\}$$

Proof. Let $\gamma(t) = (x(t), y(t), z(t))$ be a unit-speed geodesic on Σ. According to Liberman's lemma $z(t)$ is convex.

Since the slope of f is at most ℓ, we have

$$|z'(t)| \leqslant \tfrac{\ell}{\sqrt{1+\ell^2}}.$$

If γ is defined on the interval $[a, b]$, then

$$\begin{aligned}\int\limits_a^b z''(t) = z'(b) - z'(a) \leqslant \\ \leqslant 2\cdot \tfrac{\ell}{\sqrt{1+\ell^2}}.\end{aligned}$$

Further, 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 slope greater than ℓ for any t. Because γ'' is perpendicular to that plane,

$$|\gamma''(t)| \leqslant z''(t)\cdot\sqrt{1+\ell^2}.$$

Recall that $\Phi(\gamma)$ denotes the total curvature of curve γ. It follows that

$$\begin{aligned}\Phi(\gamma) &= \int\limits_a^b |\gamma''(t)|\cdot dt \leqslant \\ &\leqslant \sqrt{1+\ell^2}\cdot\int\limits_a^b z''(t)\cdot dt \leqslant \\ &\leqslant 2\cdot\ell.\end{aligned}$$

□

9.20. Exercise. *Note that the graph $z = \ell\cdot\sqrt{x^2+y^2}$ with removed origin is a smooth surface; denote it by Σ. Show that it has an both side infinite geodesic γ with total curvature exactly $2\cdot\ell$.*

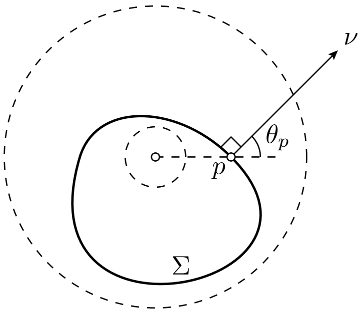
Note that the function $f(x, y) = \ell\cdot\sqrt{x^2+y^2}$ is ℓ-Lipschitz. The graph $z = f(x, y)$ in the exercise can be smoothed in a neighborhood of the origin while keeping it convex. It follows that the estimate in the Usov's theorem is optimal.

9.21. Exercise. *Assume f is a convex $\frac{3}{2}$-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 on its graph $z = f(x, y)$.

9.22. 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.



9.23. 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 a point $p \in \Sigma$. Set $\rho(t) = |\gamma(t)|^2$; let $k(t)$ be 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) Use the the statements above to conclude that

$$\Phi(\gamma) \leqslant \frac{100}{\varepsilon^2}.$$

Note that our bound on total curvature given above goes to infinity as $\varepsilon \to 0$, but in fact there is a bound independent of ε; it is good of any closed convex surface [29].

Chapter 10

Spherical map

Differential

Let $f\colon \Sigma \to \mathbb{R}^3$ be a smooth map defined on a smooth surface Σ; that is, for any chart s of Σ the composition $f \circ s$ is smooth.

Given a smooth curve γ in Σ consider the smooth curve $\hat{\gamma} = f \circ \gamma$. Assume γ starts at a point $p \in \Sigma$ with velocity vector $v \in \mathrm{T}_p\Sigma$; that is, $p = \gamma(0)$ and $v = \gamma'(0)$. The differential of v at p is defined as

❶ $$d_p f(v) = \hat{\gamma}'(0).$$

The domain of definition of $d_p f$ is the tangent plane $\mathrm{T}_p\Sigma$. The differential is an operator that produces a map $d_p f\colon \mathrm{T}_p\Sigma \to \mathbb{R}^3$ for given smooth map $f\colon \Sigma \to \mathbb{R}^3$ and $p \in \Sigma$.

Note that the value $d_p f(v)$ does not depend on the choice of γ; that is, if γ_1 is another curve in Σ such that $\gamma_1(0) = p$ and $\gamma_1'(0) = v$, then

❷ $$(f \circ \gamma)'(0) = (f \circ \gamma_1)'(0).$$

Indeed, $v = \gamma_1'(0) = \gamma'(0)$; therefore we have that

$$|\gamma_1(\varepsilon) - \gamma(\varepsilon)| = o(\varepsilon).$$

Since f is smooth,

$$|f \circ \gamma_1(\varepsilon) - f \circ \gamma(\varepsilon)| = o(\varepsilon);$$

whence ❷ follows.

Note that if Σ is a plane, then

$$d_p f(v) = D_v f,$$

where D_v denotes the direction derivative. For general surface $D_v f$ is not well defined since the point $p+t\cdot v$ may not lie in Σ for small values t.

10.1. Exercise. *Assume f is a smooth map from one surface Σ_0 to another Σ_1 and $p\in\Sigma_0$. Show that the range of $d_p f$ lies in the tangent plane $\mathrm{T}_{f(p)}\Sigma_1$.*

10.2. Proposition. *The differential is a linear map. That is, for any smooth map $f\colon \Sigma\to\mathbb{R}^3$ defined on a smooth surface Σ and $p\in\Sigma$, the map $d_p f\colon \mathrm{T}_p\to\mathbb{R}^3$ is linear.*

Proof. Fix a chart $(u,v)\mapsto s(u,v)$ on Σ that covers a neighborhood of p. Without loss of generality we may assume that $p=s(0,0)$.

Any smooth curve γ that starts at p can be written locally in the chart as $\gamma(t)=s(u(t),v(t))$; since γ stats at p, we have $u(0)=v(0)=$ $=0$. Applying the chain rule, we get

$$\begin{aligned}\gamma'(0)&=\tfrac{\partial s}{\partial u}(0,0)\cdot u'(0)+\tfrac{\partial s}{\partial v}(0,0)\cdot v'(0),\\ (f\circ\gamma)'(0)&=\tfrac{\partial f\circ s}{\partial u}(0,0)\cdot u'(0)+\tfrac{\partial f\circ s}{\partial v}(0,0)\cdot v'(0).\end{aligned}$$

The statement follows since $d_p(\gamma'(0))=(f\circ\gamma)'(0)$. □

Shape operator

Suppose Σ is an oriented surface with the unit normal field ν; in other words $\nu\colon\Sigma\to\mathbb{S}^2$ is its spherical map.

Fix a point $p\in\Sigma$. The *shape operator* at p is defined as

❸ $$S_p:=-d_p\nu.$$

The shape operator S_p is defined on the tangent plane $\mathrm{T}_p\Sigma$ and it returns a vector in the same plane (otherwise we could not call it an *operator*). The latter is shown in the following proposition, which also gives a reason for the change of sign in ❸.

Recall that

$$\mathrm{II}_p(w,v)=\langle M_p\cdot v,w\rangle,$$

where M_p is the Hessian matrix at p. It follows that

$$S_p(v)=M_p\cdot v$$

if v is written in the standard basis of the (x,y)-plane. Whence we get the following theorem.

10.3. Theorem. *Let Σ be a smooth oriented surface and $p \in \Sigma$. A nonzero tangent vector $v \in \mathrm{T}_p$ points in a principle direction at p if and only if $S_p(v) \parallel v$ and if so, then the unique coefficient k such that $S_p(v) = k \cdot v$ is the principle curvature in this direction.*

In particular $K(p) = \det S_p$ and $H(p) = \operatorname{trace} S_p$.

10.4. Exercise. *Suppose that a geodesic γ on the surface Σ is also a curvature line. Show that γ lies in a plane.*

Area

Let Σ be a smooth surface and $h\colon \Sigma \to \mathbb{R}$ be a smooth function. Let us define the integral $\int_R h$ of the function h along a region $R \subset \Sigma$.

First 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

$$❹ \qquad \int\limits_R h := \iint\limits_{s^{-1}(R)} h \circ s(u, v) \cdot \left|\tfrac{\partial s}{\partial v}(u, v) \times \tfrac{\partial s}{\partial u}(u, v)\right| \cdot du \cdot dv.$$

By substitution rule for multiple variables (A.12), the right hand side in ❹ does not depend on choice of s; that is, if $s_1\colon U_1 \to \Sigma$ is another chart such that $s_1(U_1) \supset R$, then

$$\iint\limits_{s^{-1}(R)} h \circ s \cdot \left|\tfrac{\partial s}{\partial v} \times \tfrac{\partial s}{\partial u}\right| \cdot du \cdot dv = \iint\limits_{s_1^{-1}(R)} h \circ s_1 \cdot \left|\tfrac{\partial s_1}{\partial v} \times \tfrac{\partial s_1}{\partial u}\right| \cdot du \cdot dv.$$

(In fact the factor $\left|\frac{\partial s}{\partial v} \times \frac{\partial s}{\partial u}\right|$ is chosen so to meet this property.)

For a general region R one could subdivide it 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

$$\int\limits_R h = \int\limits_{R_1} h + \int\limits_{R_2} h + \dots$$

The area of R is defined as the integral

$$\operatorname{area} R = \int\limits_R 1.$$

Spherical image

10.5. Theorem. *Let Σ be an oriented proper surface without boundary and with positive Gauss curvature. Then the spherical map $\nu\colon\Sigma\to\mathbb{S}^2$ is injective and*

$$\int\limits_R K = \operatorname{area}[\nu(R)]$$

for any region R in Σ.

Proof. Lets show that the spherical map $\nu\colon\Sigma\to\mathbb{S}^2$ is injective. Fix two distinct points $p,q\in\Sigma$. Recall that Σ bounds a strictly convex region. Therefore the ν_p makes an obtuse angle with the line segment $[p,q]$. The same way we can show that ν_q makes an obtuse angle with the line segment $[q,p]$. In other words the projections of ν_p and ν_q on the line pq point in the opposite directions. In particular $\nu_p\ne\nu_q$; that is, the spherical map is injective.

Note that it is sufficient to prove the identity assuming that the region R is covered by one chart $(u,v)\mapsto s(u,v)$ of Σ; if not cut R into smaller regions and sum up the results. Applying the definition of integral, we have the following expression for the left hand side

$$\int\limits_R K := \iint\limits_{s^{-1}(R)} K[s(u,v)]{\cdot}|\tfrac{\partial s}{\partial v}(u,v)\times\tfrac{\partial s}{\partial u}(u,v)|{\cdot}du{\cdot}dv.$$

Applying the definition of area, we have the following expression for the right hand side

$$\operatorname{area}[\nu(R)] := \iint\limits_{s^{-1}(R)} |\tfrac{\partial\nu\circ s}{\partial v}(u,v)\times\tfrac{\partial\nu\circ s}{\partial u}(u,v)|{\cdot}du{\cdot}dv.$$

Therefore it is sufficient to show that

❺ $$\tfrac{\partial\nu\circ s}{\partial v}(u,v)\times\tfrac{\partial\nu\circ s}{\partial u}(u,v) = K[s(u,v)]{\cdot}\tfrac{\partial s}{\partial v}(u,v)\times\tfrac{\partial s}{\partial u}(u,v)$$

for any (u,v) in the domain of definition.

Fix a point $p=s(u,v)$. Recall that

$$\tfrac{\partial\nu\circ s}{\partial u} = S_p(\tfrac{\partial s}{\partial u}) \quad\text{and}\quad \tfrac{\partial\nu\circ s}{\partial v} = S_p(\tfrac{\partial s}{\partial v}).$$

Therefore

$$\tfrac{\partial\nu\circ s}{\partial v}\times\tfrac{\partial\nu\circ s}{\partial u} = \det S_p{\cdot}\tfrac{\partial s}{\partial v}\times\tfrac{\partial s}{\partial u}.$$

Since $K(p)=\det S_p$, ❺ follows. □

10.6. Exercise. *Let Σ be a closed surface with positive Gauss curvature. Show that*

$$\int\limits_{\Sigma} K = 4\cdot\pi.$$

10.7. Exercise. *Let Σ be an open surface with positive Gauss curvature. Show that*

$$\int\limits_{\Sigma} K \leqslant 2\cdot\pi.$$

Chapter 11

Parallel transport

Parallel 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 $t \mapsto v(t)$ is called a *tangent field* on γ if the vector $v(t)$ lies in the tangent plane $\mathrm{T}_{\gamma(t)}\Sigma$ for each t.

A tangent field $v(t)$ on γ is called *parallel* if $v'(t) \perp \mathrm{T}_{\gamma(t)}$ 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 $v(t)$ with $v' \equiv 0$. The condition $v'(t) \perp \mathrm{T}_{\gamma(t)}$ means that this 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 $v(t) = \gamma'(t)$ of any geodesic γ is parallel along γ.

11.1. Exercise. *Let Σ be a smooth regular surface in the Euclidean space, $\gamma\colon [a,b] \to \Sigma$ a smooth curve and $v(t)$, $w(t)$ parallel vector fields along γ.*

(a) Show that $|v(t)|$ is constant.

(b) Show that the angle $\theta(t)$ between $v(t)$ and $w(t)$ is constant.

Parallel transport

Assume $p = \gamma(a)$ and $q = \gamma(b)$. Given a tangent vector $v \in \mathrm{T}_p$ there is unique parallel field $v(t)$ along γ such that $v(a) = v$. The latter follows from A.14; the uniqueness also follows from Exercise 11.1.

The vector $v(b) \in \mathrm{T}_q$ is called the *parallel transport* of v along γ and denoted as $\iota_\gamma(v)$.

From the Exercise 11.1, it follows that parallel transport $\iota_\gamma\colon \mathrm{T}_p \to \to \mathrm{T}_q$ is an an isometry; it depends on the choice of γ — for another curve γ_1 connecting p to q in Σ, the parallel transport $\iota_{\gamma_1}\colon \mathrm{T}_p \to \mathrm{T}_q$ might be different.

To interpret the parallel transport physically, think of walking along γ and carrying a perfectly balanced bike wheel in such a way that you touch only its axis and keep it normal to Σ. It should be physically evident that if the wheel is non-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.

This physical interpretation was suggested by Mark Levi [30]; it will be used further.

On a more formal level, one can choose a partition $a = t_0 < \dots < < t_n = b$ 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

$$\varphi_n \circ \dots \circ \varphi_1\colon T_p \to \mathrm{T}_q$$

gives an approximation of ι_γ. Each φ_i does not increase the magnitude of a vector and neither the composition. It is straightforward to see that if if the partition is sufficiently fine, then it is almost isometry; in particular it almost preserves the magnitudes of tangent vectors.

11.2. Exercise. *Let γ be a smooth closed loop with base point p on a smooth oriented surface Σ with the unit normal field ν. Suppose that the spherical image of γ lies in a great circle. Show that the parallel translation $\iota_\gamma\colon \mathrm{T}_p \to \mathrm{T}_p$ along γ is the identity map.*

Geodesic curvature

Plane is the simplest example of smooth surface. Earlier we introduced signed curvature of a plane curve. For a smooth curve γ in general oriented smooth surface Σ the analogous notion is called *geodesic curvature* which we are about to introduce.

Let $\nu\colon \Sigma \to \mathbb{S}^2$ be the spherical map that defines the orientation on Σ. Without loss of generality we can assume that γ has unit speed. Then for any t the vectors $\nu(t) = \nu(\gamma(t))_\Sigma$ and the velocity vector $\mathrm{T}(t) = \gamma'(t)$ are unit vectors that are normal to each other. Denote by $\mu(t)$ the unit vector that is normal to both $\nu(t)$ and $\mathrm{T}(t)$ that points to the left from γ; that is, $\mu = \nu \times \mathrm{T}$. Note that the triple $\mathrm{T}(t), \mu(t), \nu(t)$ is an oriented orthonormal basis for any t.

Since γ is unit-speed, the acceleration $\gamma''(t)$ is perpendicular to $\mathrm{T}(t)$; therefore at any parameter value t, we have

$$\gamma''(t) = k_g(t)\cdot\mu(t) - k_n(t)\cdot\nu(t),$$

for some real numbers $k_n(t)$ and $k_g(t)$. The numbers $k_n(t)$ and $k_g(t)$ are called *normal* and *geodesic curvature* of γ at t correspondingly.

Note that the geodesic curvature vanishes if γ is a geodesic. It measures how much a given curve diverges from being a geodesic; it is positive if γ turns left and negative if γ turns right.

Total geodesic curvature

The total geodesic curvature is defined as integral

$$\Psi(\gamma) := \int\limits_{\mathbb{I}} k_g(t)\cdot dt,$$

assuming that γ is defined on the real interval $\mathbb{I}$. Note that if Σ is a plane and γ lies in Σ, then geodesic curvature of γ equals to signed curvature and therefore total geodesic curvature equals to the 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 a sum of all total geodesic curvature of its arcs and the sum signed exterior angles of γ at the joints. More precisely, if γ is a concatenation of smooth regular curves $\gamma_1, \dots, \gamma_n$, 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 γ_i and γ_{i+1}; it is positive if we turn left and negative if we turn right, it is undefined if we turn to the opposite direction. If γ is closed, then

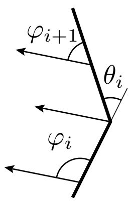
$$\Psi(\gamma) = \Psi(\gamma_1) + \dots + \Psi(\gamma_n) + \theta_1 + \dots + \theta_n,$$

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

Note that if each arc γ_i in the concatenation is a geodesic, then γ is called *broken geodesic*. Note that 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.

11.3. 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* $\iota_\gamma\colon T_p \to \mathrm{T}_p$ *is a rotation of the the plane* T_p *clockwise by angle* $\Psi(\gamma)$.

Moreover, the same statement holds 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 $w_i(t) = \gamma_i'(t)$ is parallel along γ_i, the angle φ_i between v and w_i stays constant on each γ_i.

If θ_i denotes the external angle at this vertex of 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 around we get that

$$\varphi_{n+1} - \varphi_1 = -\theta_1 - \dots - \theta_n = -\Psi(\gamma).$$

Hence the the first statement follows.

For the smooth unit-speed curve $\gamma\colon [a, b] \to \Sigma$, the proof is analogous. If $\varphi(t)$ denotes the angle between $v(t)$ and $w(t) = \gamma'(t)$, then

$$\varphi'(t) + k_g(t) \equiv 0$$

Whence the angle of rotation

$$\begin{aligned}\varphi(b) - \varphi(a) &= \int_a^b \varphi'(t){\cdot}dt = \\ &= -\int_a^b k_g{\cdot}dt = \\ &= -\Psi(\gamma)\end{aligned}$$

The case of piecewise regular smooth curve is a straightforward combination of the above two cases. □

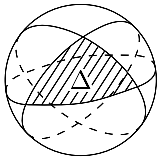
Spherical area

11.4. 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.



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. □

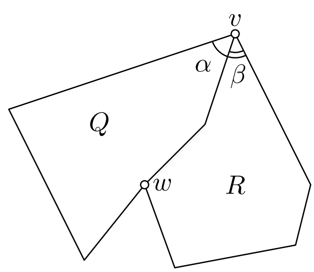
If the contour $\partial\Delta$ of a spherical triangle with angles α, β and γ is oriented such that the triangle lies on the left, then its external angles are $\pi-\alpha$, $\pi-\beta$ and $\pi-\gamma$. Therefore the total geodesic curvature of $\partial\Delta$ is $\Psi(\partial\Delta) = 3\cdot\pi-\alpha-\beta-\gamma$. The identity ❶ can be rewritten as

❹ $$\Psi(\partial\Delta) + \operatorname{area}\Delta = 2\cdot\pi.$$

The formula ❹ holds for an arbitrary spherical polygon bounded by a simple broken geodesic. The latter can be proved by triangulating the poygon, applying the formula for each triangle in the triangulation and summing up the results.

If a spherical polygon is P divided in two polygons Q and R by polygonal line between vertexes v and w then

$$\Psi(\partial P) + 2\cdot\pi = \Psi(\partial Q) + \Psi(\partial R).$$



Indeed, for the internal angles Q and R at v are α and β, then their external angles are $\pi-\alpha$ and $\pi-\beta$ respectfully. The internal angle of P in this case is $\alpha+\beta$ and its external angle is $\pi-\alpha-\beta$ Clearly we have that

$$(\pi-\alpha)+(\pi-\beta) = (\pi-\alpha-\beta)+\pi;$$

that is, the sum of external angles of Q and R at v is π plus the external angle of P at v. The same holds for the external angles at

w and the rest of the external angles of P appear once on Q or R. Therefore if the formula ➍ holds for Q and R, then it holds for P.

The following proposition gives a spherical analog of 4.4.

11.5. 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 from ∂P. Then*

$$\Psi(\partial P) + \operatorname{area} P = 2{\cdot}\pi.$$

Moreover the same formula holds for any spherical region P bounded by piecewise smooth simple closed curve ∂P.

Sketch of proof. The proof of the first statement is given above.

The second statement can be proved by approximation. One has to show that the total geodesic curvature of an inscribed broken geodesic approximates the total geodesic curvature of the original curve. We omit the proof of the latter statement, but it can be done along the same lines as 3.23. □

11.6. Exercise. *Assume γ is a simple piecewise smooth loop on $\mathbb{S}^2$ that divides its area into two equal parts. Denote by p the base point of γ. Show that $\iota_\gamma\colon \mathrm{T}_p\mathbb{S}^2 \to \mathrm{T}_p\mathbb{S}^2$ is the identity map.*

11.1 Gauss–Bonnet formula

11.7. Theorem. *Let Δ be a topological disc in a smooth oriented surface Σ bounded by a simple piecewise smooth and regular curve $\partial\Delta$ that is oriented in such a way that Δ lies on its left. Then*

➊ $$\Psi(\partial\Delta) + \int\limits_\Delta K = 2{\cdot}\pi,$$

where K denotes the Gauss curvature of Σ.

For geodesic triangles this theorem was proved by Carl Friedrich Gauss [31]; Pierre Bonnet and Jacques Binet independently generalized the statement for arbitrary curves.

Note that if Σ is a plane, then the Gauss curvatue vanished; therefore the statement of theorem follows from 4.4.

If Σ is the unit sphere, then $K \equiv 1$. Therefore formula ➍ can be rewritten as

$$\Psi(\partial\Delta) + \operatorname{area}\Delta = 2{\cdot}\pi,$$

which follows from 11.5.

We will give an informal proof of 11.15 in a partial case based on the bike wheel interpretation described above. We suppose that it is intuitively clear that moving the axis of the wheel without changing its direction does not change the direction of the wheel's spikes.

More precisely, assume we keep the axis of a non-spinning bike wheel and perform the following two experiments:

(i) We move it around and bring the axis back to the original position. As a result the wheel might turn by some angle; let us measure this angle.

(ii) We move the direction of the axis the same way as before without moving the center of the wheel. After that we measure the angle of rotation.

Then the resulting angles in these two experiments is the same.

Consider a surface Σ with a Gauss map $\nu\colon \Sigma \to \mathbb{S}^2$. Note that for any point p on Σ, the tangent plane $\mathrm{T}_p\Sigma$ is parallel to the tangent plane $\mathrm{T}_{\nu(p)}\mathbb{S}^2$; so we can identify these tangent spaces. From the experiments above, we get the following:

11.8. Lemma. *Suppose α is a piecewise smooth regular curve in a smooth regular surface Σ which has a Gauss map $\nu\colon \Sigma \to \mathbb{S}^2$. Then the parallel transport along α in Σ coincides with the parallel transport along the curve $\beta = \nu \circ \alpha$ in $\mathbb{S}^2$.*

Proof of partial case of 11.15. We will prove the formula for proper surface Σ with positive Gauss curvature. In this case, by 10.5 the formula can be rewritten as

$$\Psi(\partial\Delta) + \mathrm{area}[\nu(\Delta)] = 2\cdot\pi. \tag{❷}$$

The general case can be proved similarly, but one has to use the area formula (A.13) and oriented area surrounded by a spherical curve.

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 $\iota\colon \mathrm{T}_p \to \mathrm{T}_p$ along α. According to 11.3, ι is a clockwise rotation by angle $\Psi(\alpha)_\Sigma$.

Set $\beta = \nu \circ \alpha$. According to 11.16, ι is also parallel translation along β in $\mathbb{S}^2$. In particular ι is a clockwise rotation by angle $\Psi(\beta)_{\mathbb{S}^2}$. By 11.5

$$\Psi(\beta)_{\mathbb{S}^2} + \mathrm{area}[\nu(\Delta)] = 2\cdot\pi.$$

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

$$\Psi(\partial\Delta)_\Sigma + \mathrm{area}[\nu(\Delta)] = 2\cdot\pi\cdot n \tag{❸}$$

for an integer n.

It remains to show that $n=1$. By ❸, this is so for a topological disc in a plane. One can think of a general disc Δ as about a 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 ❸ is continuous along the deformation.

Let us redo the last argument more formally.

First assume that Δ lies in a local graph realization $z=f(x,y)$ of Σ. Consider 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 $\Delta_1=\Delta$ amd Δ_0 is its projection to the (x,y)-plane. Since Σ_0 is a plane domain, we have $\mathrm{area}[\nu_0(\Delta_0)]=0$. Therefore by 4.4 we gave

$$\Psi(\partial\Delta_0)_{\Sigma_0}+\mathrm{area}[\nu_0(\Delta_0)]=2{\cdot}\pi.$$

Note that

$$\Psi(\partial\Delta_t)_{\Sigma_t}+\mathrm{area}[\nu_t(\Delta_t)]$$

depends continuously on t. According to ❸, its value is a multiple of $2{\cdot}\pi$; therefore it has to be constant. Whence the Gauss–Bonnet formula follows.

If Δ does not lie in one graph, then one could divide it into smaller discs, apply the formula for each and sum up the result. The proof is done along the same lines as 11.5. □

11.9. 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*

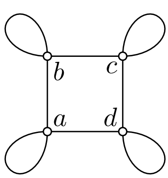
$$\mathrm{length}\,\gamma\leqslant 2{\cdot}\pi;$$

that is, the length of γ cannot exceed the length of the unit circle in the plane.

11.10. Exercise. *Let γ be a closed simple 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 regions of equal area.*

Conclude that

$$\mathrm{length}\,\alpha\geqslant 2{\cdot}\pi.$$

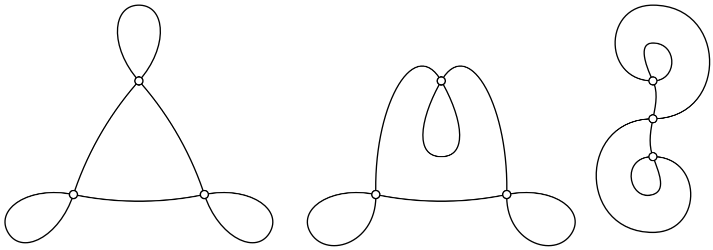


11.11. Exercise. *Let Σ be a smooth closed surface with a closed geodesic γ. Assume γ has exactly 4 self-intersection at the points a, b, c and d that appear on γ in the order a,a,b,b,c,c,d,d. Show that Σ cannot have positive Gauss curvature.*

The following exercise gives the optimal bound on Lipschitz constant of a convex function that guarantees that its geodesics have no self-intersections; compare to 9.21.

11.12. 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 surface defined by the graph $z = f(x, y)$ has no self-intersections.*

11.13. Advanced exercise. *Let Σ be a smooth regular sphere with positive Gauss curvature and $p \in \Sigma$. Suppose γ is a closed geodesic that does not pass thru p. Assume $\Sigma\backslash\{p\}$ parametrized by the plane. Can it happen that in this parametrization, γ looks like one of the curves on the diagram? Say as much as possible about possible/impossible*



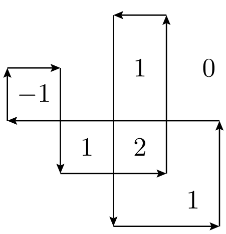
diagrams of that type.

Signed area

The formula ❹ holds modulo $2{\cdot}\pi$ for any closed broken geodesic, if one use *signed area* surrounded by curve instead of usual area; that is, we count area of the regions taking into account how many times the curve goes around the region.

Namely, we have to choose a *south pole* and state that its region has zero multiplicity. When you cross the curve the mulitplicity changes by ±1; we add 1 if the curve crosses your path from left to right and we subtract 1 otherwise. The signed area surrounded by a closed curve is the sum of area of all the regions counted with multiplicities.

Here is an example of a broken line with multiplicities assuming that the big region has the south pole inside.



This signed-area formula can be proved in a similar way: Apply the formula for each triangle with vertex at the north pole and base at each edge of the broken geodesic. Sum the resulting identities taking each with a sign: plus

if the triangle lies on the left from the edge and
minus if the triangle lies on the right from edge.

Choosing a different pole will change all the coefficients by the same number. So the resulting formula holds only modulo the area of $\mathbb{S}^2$, which is $4\cdot\pi$ — this will not destroy identity modulo $2\cdot\pi$.

Furthermore, by approximation, the signed-area formula holds for any reasonable curve, say piecewise smooth regular curves on the sphere. Summarizing, we hope the discussion above convinced the reader that the following statement holds.

A domain Δ in a surface is called a *disc* (or more precisely *topological disc*) if it is bounded by a closed simple curve and can be parameterized by a unit plane disc

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

That is, there is a continuous bijection $\mathbb{D} \to \Delta$.

11.14. Proposition. *For any closed piecewise smooth regular curve α on the sphere, we have that*

$$\Psi(\alpha) + \operatorname{area}\alpha = 0 \pmod{2\cdot\pi},$$

where $\operatorname{area}\alpha$ *denotes the signed area surrounded by α and $\Psi(\alpha)$ the total geodesic curvature of α.*

Moreover, if α is a simple curve that bounds a disc Δ on the left from it, then we have

$$\Psi(\alpha) + \operatorname{area}\Delta = 2\cdot\pi.$$

Gauss–Bonnet formula

11.15. Theorem. *Let Δ be a disc in a smooth oriented surface Σ bounded by a simple piecewise smooth and regular curve $\partial\Delta$ that is oriented in such a way that Δ lies on its left. Then*

❹
$$\Psi(\partial\Delta) + \iint\limits_{\Delta} G = 2\cdot\pi,$$

where G denotes the Gauss curvature of Σ.

For geodesic triangles this theorem was proved by Carl Friedrich Gauss [31]; Pierre Bonnet and Jacques Binet independently generalized the statement for arbitrary curves. The modern formulation described below was given by Wilhelm Blaschke.

Remarks; (1). For a general compact domain Δ (not necessary a disc) we have that

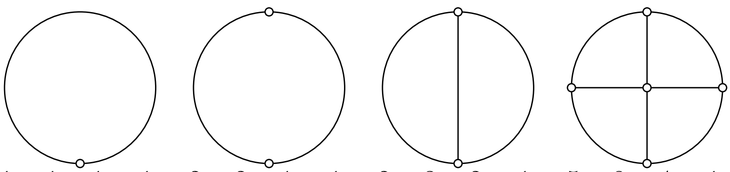
❺ $$\Psi(\partial\Delta) + \iint\limits_{\Delta} G = 2{\cdot}\pi{\cdot}\chi(\Delta),$$

where $\chi(\Delta)$ is the so called *Euler's characteristic* of Δ. The Euler's characteristic is *topological invariant*, in particular preserved in a continuous deformation.

If a surface Σ (possibly with boundary) can be divided into f discs by drawing e edges connecting v vertexes, then

$$\chi(\Sigma) = v - e + f.$$

For example the disc $\mathbb{D}$ has Euler's characteristic 1; it can be divided



$1 - 1 + 1 = 1$; $2 - 2 + 1 = 1$; $2 - 3 + 2 = 1$; $5 - 8 + 4 = 1$.

into discs many ways, but each time we have $v - e = f = 1$. The latter agrees with ❹ and ❺. It is useful to know that $\chi(\mathbb{S}^2) = 2$; $\chi(\mathbb{T}^2) = 0$ where $\mathbb{T}^2$ denotes torus; $\chi(S_g) = 2 - 2{\cdot}g$, where S_g is a surface of genus g; that is, sphere with g handles.

(2). Note that if Σ is a plane, then a geodesic in Σ are formed by line segments. In this case the statement of theorem follows from Exercise **??**.

(3). If Σ is the unit sphere, then $G \equiv 1$ and therefore formula ❹ can be rewritten as

$$\Psi(\partial\Delta) + \operatorname{area}\Delta = 2{\cdot}\pi,$$

which follows from Proposition 11.14.

We will give an informal proof of 11.15 based on the bike wheel interpretation described above. We suppose that it is intuitively clear that moving the axis of the wheel without changing its direction does not change the direction of the wheel's spikes.

More precisely, assume we keep the axis of a non-spinning bike wheel and perform the following two experiments:

(i) We move it around and bring the axis back to the original position. As a result the wheel might rotate by some angle; let us measure this angle.

(ii) We move the direction of the axis the same way as before without moving the center of the wheel. After that we measure the angle of rotation.

Then the resulting angle in these two experiments is the same.

Consider a oriented smooth surface Σ with the spherical; map $\nu\colon\Sigma\to\mathbb{S}^2$. Note that for any point p on Σ, the tangent plane $\mathrm{T}_p\Sigma$ is parallel to the tangent plane $\mathrm{T}_{\nu(p)}\mathbb{S}^2$; so we can identify these tangent spaces. From the experiments above, we get the following:

11.16. Lemma. *Suppose α is a piecewise smooth regular curve in a smooth regular surface Σ which has a Gauss map $\nu\colon\Sigma\to\mathbb{S}^2$. Then the parallel transport along α in Σ coincides with the parallel transport along the curve $\beta=\nu\circ\alpha$ in $\mathbb{S}^2$.*

11.17. Exercise. *Let Σ be a smooth closed surface with positive Gauss curvature. Given a line ℓ denote by ω_ℓ the closed curve formed by points with tangent planes parallel to ℓ.*[1] *Show that parallel transport around ω_ℓ is the identity map.*

Now we are ready to prove the theorem.

Proof of 11.15. Let α be the boundary $\partial\Delta$ parameterized in such a way that Δ lies on the left from it. Assume p is the point where α starts and ends.

Set $\beta=\nu\circ\gamma$ and $q=\nu(p)$, so the spherical curve β starts and ends at q.

By Lemma 11.16 the parallel transport along α in Σ coincides with the parallel transport along the curve β in $\mathbb{S}^2$. By Proposition 11.3, it follows that

$$\Psi(\alpha,\Sigma)=\Psi(\beta,\mathbb{S}^2)\pmod{2\cdot\pi}.$$

By Proposition 11.14,

$$\Psi(\beta,\mathbb{S}^2)+\operatorname{area}\beta=0\pmod{2\cdot\pi}.$$

Therefore

$$\Psi(\alpha,\Sigma)+\operatorname{area}\beta=0\pmod{2\cdot\pi}.$$

Recall that the shape operator $s_p\colon T_p\Sigma\to\mathrm{T}_{\nu(p)}\mathbb{S}^2=T_p\Sigma$ is the Jacobian of the Gauss map $\nu\colon\Sigma\to\mathbb{S}^2$ at the point p. In appropriately chosen coordinates in T_p, the shape operator can be presented by a diagonal matrix $\left(\begin{smallmatrix}k_1&0\\0&k_2\end{smallmatrix}\right)$, where k_1 and k_2 are the principle curvatures at p. Therefore, the determinant of s_p is the Gauss curvature at p.

[1] Equivalently the normal vector at any point of ω_ℓ is perpendicular to ℓ. If the light falls on Σ from one side parallel to ℓ, then ω_ℓ divides the bright and dark sides of Σ.

If Σ is a closed surface with positive Gauss curvature, then the Gauss map $\nu\colon \Sigma \to \mathbb{S}^2$ is a smooth bijection. Therefore

$$\iint\limits_{\Delta} G = \operatorname{area}[\nu(\Delta)].$$

In the general case we have to count the area $\nu(\Delta)$ taking orientation and multiplicity of the Gauss map into account. In this case

$$\iint\limits_{\Delta} G = \operatorname{area}\beta,$$

where $\operatorname{area}\beta$ is the signed area surrounded by β; it is defined above. Therefore

❻ $$\Psi(\alpha,\Sigma) + \iint\limits_{\Delta} G = 0 \pmod{2\cdot\pi}.$$

If Δ is a disc in the plane, then Gauss curvature vanishes and by Exercise **??**, we have

$$\Psi(\partial\Delta) + \iint\limits_{\Delta} G = 2\cdot\pi.$$

Assunme that Σ_t is a smooth one parameter family of surfaces with a one parameter family of discs $\Delta_t \subset \Sigma_t$ and α_t is the boundary $\partial\Delta_t$ parameterized in such a way that Δ_t lies on the left from it. The value

$$f(t) = \Psi(\alpha_t) + \iint\limits_{\Delta} G$$

is continuous in t and by ❻ it has to be constant.

If Σ_0 is a plane, then

$$\Psi(\partial\Delta_0) + \iint\limits_{\Delta_0} G = 2\cdot\pi.$$

Intuitively it is clear that any disc can be obtained as a resultof continuous deformation of plane disc. Therefore

$$\Psi(\partial\Delta_1) + \iint\limits_{\Delta_1} G = 2\cdot\pi$$

for arbitrary disc Δ_1; whence ❹ follows. □

The remarkable theorem

Let Σ_1 and Σ_2 be two smooth regular surfaces in the Euclidean space. A map $f\colon \Sigma_1 \to \Sigma_2$ is called length-preserving if for any curve γ_1 in Σ_1 the curve $\gamma_2 = f \circ \gamma_1$ in Σ_2 has the same length. If in addition f is smooth and bijective, then it is called *intrinsic isometry.*

A simple example of intrinsic isometry can obtained by warping a plane into a cylinder. The following exercise produce slightly more interesting example.

11.18. Exercise. *Suppose $\gamma(t) = (x(t), y(t))$ is a smooth unit-curve in the plane such that $y(t) = a\cdot \cos t$. Let Σ_γ be the surface of revolution of γ around the x-axis. Show that a small open domain in Σ_γ admits a smooth length-pereserving map to the unit sphere.*

Conclude that any round disc Δ in $\mathbb{S}^2$ of intrinsic radius smaller than $\frac{\pi}{2}$ 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 any $t \neq 0$.[2]

11.19. Theorem. *Suppose $f\colon \Sigma_1 \to \Sigma_2$ is an intrinsic isometry between two smooth regular surfaces in the Euclidean space; $p_1 \in \Sigma_1$ and $p_2 = f(p_1) \in \Sigma_1$. Then*

$$G(p_1)_{\Sigma_1} = G(p_2)_{\Sigma_2};$$

that is, the Gauss curvature of Σ_1 at p_1 is the same as the Gauss curvature of Σ_2 at p_2.

This theorem was proved by Carl Friedrich Gauss [31] who called it *Remarkable theorem* (Theorema Egregium). The theorem is indeed remarkable because the Gauss curvature is defined as a product of principle curvatures which might be different at these points; however, according to the theorem, their product can not change.

In fact Gauss curvature of the surface at the given point can be found *intrinsically*, by measuring the lengths of curves in the surface. For example, Gauss curvature $G(p)$ in the following formula for the circumference $c(r)$ of a geodesic circle centered at p in a surface:

$$c(r) = 2\cdot\pi\cdot r - \tfrac{\pi}{3}\cdot G(p)\cdot r^3 + o(r^3).$$

Note that the theorem implies there is no smooth length-preserving map that sends an open region in the unit sphere to the plane.[3] It

[2] In fact any disc in $\mathbb{S}^2$ of intrinsic radius smaller than π admits a smooth length preserving deformation.

[3] There are plenty of non-smooth length-preserving maps from the sphere to the plane; see [32] and the references there in.

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 a region on Earth without distortion.

Proof. Set $g_1 = G(p_1)_{\Sigma_1}$ and $g_2 = G(p_2)_{\Sigma_2}$; we need to show that

❼ $$g_1 = g_2.$$

Suppose Δ_1 is a small geodesic triangle in Σ_1 that contains p_1. Set $\Delta_2 = f(\Delta_1)$. We may assume that the Gauss curvature is almost constant in Δ_1 and Δ_2; that is, given $\varepsilon > 0$, we can assume that

❽ $$\begin{aligned} |G(x_1)_{\Sigma_1} - g_1| &< \varepsilon, \\ |G(x_2)_{\Sigma_2} - g_2| &< \varepsilon \end{aligned}$$

for any $x_1 \in \Delta_1$ and $x_2 \in \Delta_2$.

Since f is length-preserving the triangles Δ_2 is geodesic and

❾ $$\operatorname{area}\Delta_1 = \operatorname{area}\Delta_2.$$

Moreover, triangles Δ_1 and Δ_2 have the same corresponding angles; denote them by α, β and γ.

By Gauss–Bonnet formula, we get that

❿ $$\iint\limits_{\Delta_1} G_{\Sigma_1} = \alpha + \beta + \gamma - \pi = \iint\limits_{\Delta_2} G_{\Sigma_2}.$$

By ❽,

$$\left| g_1 - \frac{1}{\operatorname{area}\Delta_1} \cdot \iint\limits_{\Delta_1} G_{\Sigma_1} \right| < \varepsilon,$$

$$\left| g_2 - \frac{1}{\operatorname{area}\Delta_2} \cdot \iint\limits_{\Delta_2} G_{\Sigma_2} \right| < \varepsilon.$$

By ❾ and ❿,

$$\frac{1}{\operatorname{area}\Delta_1} \cdot \iint\limits_{\Delta_1} G_{\Sigma_1} = \frac{1}{\operatorname{area}\Delta_2} \cdot \iint\limits_{\Delta_2} G_{\Sigma_2},$$

therefore

$$|g_1 - g_2| < 2{\cdot}\varepsilon.$$

Since $\varepsilon > 0$ is arbitrary, ❼ follows. □

Simple geodesic

The following theorem provides an interesting application of Gauss–Bonnet formula; it is proved by Stephan Cohn-Vossen [Satz 9 in 33].

11.20. Theorem. *Any open smooth regular surface with positive Gauss curvature has a simple two-sided infinite geodesic.*

11.21. Lemma. *Suppose Σ is an open surface in with positive Gauss curvature in the Euclidean space. Then there is a convex function f defined on a convex open region of (x,y)-plane such that Σ can be presented as a graph $z=f(x,y)$ in some (x,y,z)-coordinate system of the Euclidean space.*

Moreover

➔ $$\iint\limits_{\Sigma} G \leqslant 2\cdot\pi.$$

Proof. The surface Σ is a boundary of an unbounded closed convex set K.

Fix $p\in\Sigma$ and consider a sequence of points x_n such that $|x_n-p|\to\to\infty$ as $n\to\infty$. Set $u_n=\frac{x_n-p}{|x_n-p|}$; the unit vector in the direction from p to 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.

Note that for any $q\in\Sigma$, the directions $v_n=\frac{x_n-q}{|x_n-q|}$ converge to u as well. The half-line from q in the direction of u lies in K. Indeed any point on the half-line is a limit of points on the line segments $[q,x_n]$; since K is closed, all of these poins lie in K.

Let us choose the z-axis in the direction of u. Note that line segments can not lie in Σ, otherwise its Gauss curvature would vanish. It follows that any vertical line can intersect Σ at most at one point. That is, Σ is a graph of a function $z=f(x,y)$. Since K is convex, the function f is convex and it is defined in a region Ω which is convex. The domain Ω is the projection of Σ to the (x,y)-plane. This projection is injective and by the inverse function theorem, it maps open sets in Σ to open sets in the plane; hence Ω is open.

It follows that the outer normal vectors to Σ at any point, points to the south hemisphere $\mathbb{S}^2_-=\left\{\,(x,y,z)\in\mathbb{S}^2\;:\;z<0\,\right\}$. Therefore the area of the spherical image of Σ is at most $\operatorname{area}\mathbb{S}^2_-=2\cdot\pi$. The area of this image is the integral of the Gauss curvature along Σ. That is,

$$\begin{aligned}\iint\limits_{\Sigma} G &= \operatorname{area}[\nu(\Sigma)] \leqslant \\ &\leqslant \operatorname{area} \mathbb{S}^2_- = \\ &= 2\cdot\pi,\end{aligned}$$

where $\nu(p)$ denotes the outer unit normal vector at p. Hence ➔ follows. □

Proof of 11.20. Let Σ be an open surface in with positive Gauss curvature and γ a two-sided infinite geodesic in Σ. The following is the key statement in the proof.

11.22. Claim. *The geodesic γ contains at most one simple loop.*

Assume γ has a simple loop ℓ. By Lemma 11.21, Σ is parameterized by a open convex region Ω in the plane; therefore ℓ bounds a disc in Σ; denote it by Δ. If φ is the angle at the base of the loop, then by Gauss–Bonnet,

$$\iint\limits_{\Delta} G = \pi + \varphi.$$

By Lemma 11.21, $\varphi < \pi$; 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 disks Δ_1 and Δ_2 have to overlap, otherwise the curvature of Σ would exceed $2\cdot\pi$.

We may assume that $\Delta_1 \not\subset \Delta_2$; the loop ℓ_2 appears after ℓ_1 on γ and there are no other simple loops between them. In this case, after going around ℓ_1 and before closing ℓ_2, the curve γ must enter Δ_1 creating a concave loop. The latter contradicts the above observation.

If a geodesic γ has a self-intersection, then it contains a simple loop. From above, there is only one such loop; it cuts a disk from Σ and goes around it either clockwise or counterclockwise. This way we divide all the self-intersecting geodesics into two sets which we will 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. By shooting unit-speed geodesics in all directions at a given point $p = \gamma(0)$, we get a one parameter family of geodesics γ_s for $s \in [0, \pi]$ connecting the geodesic $t \mapsto \gamma(t)$ with the $t \mapsto \gamma(-t)$; that is, $\gamma_0(t) = \gamma(t)$ and $\gamma_\pi(t) = \gamma(-t)$. It follows that there are geodesics which aren't clockwise nor counterclockwise. Those geodesics have no self-intersections. □

Chapter 12

Local comparison

12.1 First variation formula

12.1. Proposition. *Assume $(s,t) \mapsto w(s,t)$ be a local parametrization of an oriented smooth regular surface Σ such that $\frac{\partial}{\partial s}w \perp \frac{\partial}{\partial t}w$, $|\frac{\partial}{\partial s}w| = 1$ and the vector $\frac{\partial}{\partial s}w$ points to the right from $\frac{\partial}{\partial t}w$ at any parameter value (s,t).*

Fix a closed real interval $[a,b]$ and consider a one parameter family of curves $\sigma_s\colon [a,b] \to \Sigma$ defined as the coordinate lines $\sigma_s(t) = w(s,t)$. Set $\ell(s) = \operatorname{length}\sigma_s$. Then

$$\ell'(s) = \Theta_{\sigma_s}$$

for any s.

The proof is done by direct calculations.

Proof. Since $\frac{\partial}{\partial s}w \perp \frac{\partial}{\partial t}w$, we have that

$$\langle \tfrac{\partial}{\partial s}w, \tfrac{\partial}{\partial t}w\rangle = 0$$

and therefore

$$\langle \tfrac{\partial^2}{\partial s \partial t}w, \tfrac{\partial}{\partial t}w\rangle + \langle \tfrac{\partial}{\partial s}w, \tfrac{\partial^2}{\partial t^2}w\rangle = \tfrac{\partial}{\partial t}\langle \tfrac{\partial}{\partial s}w, \tfrac{\partial}{\partial t}w\rangle = 0.$$

Note that $|\gamma_s'(t)| = |\tfrac{\partial}{\partial t} w(s,t)|$ and therefore

$$\begin{aligned}
\tfrac{\partial}{\partial s}|\gamma_s'(t)| &= \tfrac{\partial}{\partial s}\sqrt{\langle \tfrac{\partial}{\partial t}w(s,t), \tfrac{\partial}{\partial t}w(s,t)\rangle} = \\
&= \frac{\langle \tfrac{\partial^2}{\partial s \partial t}w(s,t), \tfrac{\partial}{\partial t}w(s,t)\rangle}{\sqrt{\langle \tfrac{\partial}{\partial t}w(s,t), \tfrac{\partial}{\partial t}w(s,t)\rangle}} = \\
&= -\frac{\langle \tfrac{\partial}{\partial s}w, \tfrac{\partial^2}{\partial t^2}w\rangle}{|\gamma_s'(t)|} = \\
&= -\frac{\langle \tfrac{\partial}{\partial s}w, \gamma_s''(t)\rangle}{|\gamma_s'(t)|}.
\end{aligned}$$

The values $\ell(s)$ do not change if we reparametrize γ_s, so we can assume that for a fixed value s the curve σ_s is unit-speed. Since $|\tfrac{\partial}{\partial s}w| = 1$ and $\tfrac{\partial}{\partial s}w$ points to the right from $\tfrac{\partial}{\partial t}w = \gamma_s'(t)$, the last expression equals to $k_g(s,t)$, where $k_g(s,t)$ denotes the geodesic curvature of σ_s at t. Therefore, for this particular s we have

$$\begin{aligned}
\ell'(s) &= \int\limits_a^b \tfrac{\partial}{\partial s}|\gamma_s'(t)|\cdot dt = \\
&= \int\limits_a^b k_g(s,t)\cdot dt = \\
&= \Theta_{\sigma_s}.
\end{aligned}$$

Since the left hand side and the right hands side of this formula do not depend on the parametrization of σ_s, this formula holds for all s.[1] $\square$

The parametrization of a surface satisfying the conditions in the proposition are called *semigeodesic coordinates.* The following exercise explains the reason for this name.

12.2. Exercise. *Assume $(s,t) \mapsto w(s,t)$ be a local parametrization of an oriented smooth regular surface Σ as in the proposition above. Show that for any fixed t the curve $\gamma_t(s) = w(s,t)$ is a geodesic.*

---

[1] One may avoid passing the a unit-speed parametrization by using the following formula for geodesic curvature which holds for any regular parametrization:

$$k_g(t,s) = \langle \nu(\sigma_s(t)), [\sigma_s'(t), \sigma_s''(t)]\rangle / |\sigma_s'(t)|^3;$$

it saves thinking but makes the calculations longer.

12.2 Exponential map

Let Σ be smooth regular surface and $p \in \Sigma$. Given a tangent vector $v \in \mathrm{T}_p$ consider a geodesic γ_v in Σ that runs from p with the initial velocity v; that is, $\gamma(0) = p$ and $\gamma'(0) = v$.

The point $q = \gamma_v(1)$ is called *exponential map* of v, or briefly $q = \exp_p v$. The map $\exp_p \colon \mathrm{T}_p \to \Sigma$ is defined in a neighborhood of zero. We assume that it is intuitively obvious that the map $\exp_p$ is smooth; formally it follows since the solution of the initial value problem for the equation $\gamma_v''(t) \perp \mathrm{T}_{\gamma_v(t)}$ which describes the geodesic γ_v smoothly depend on the initial data v. Note that the Jacobian of $\exp_p$ at zero is the identity matrix. Therefore frm the inverse function theorem we get the following statement:

12.3. Proposition. *Let Σ be smooth regular surface and $p \in \Sigma$. Then the exponential map $\exp_p \colon \mathrm{T}_p \to \Sigma$ is a smooth regular parametrization of a neighborhood of p in Σ by a neighborhood of 0 in the tangent plane T_p.*

Moreover for any $p \in \Sigma$ there is $\varepsilon > 0$ such that for any $x \in \Sigma$ such that $|x - p|_\Sigma < \varepsilon$ the map $\exp_x \colon \mathrm{T}_x \to \Sigma$ is a smooth regular parametrization of the ε-neighborhood of x in Σ by the ε-neighborhood of zero in the tangent plane T_x.

Note that if there are two minimizing geodesics between two points x and y in a surface, then there are two distinct vectors $v, v' \in \mathrm{T}_x$ such that $y = \exp_x v = \exp_x v'$. Therefore by the above proposition we get the following:

12.4. Corollary. *Let Σ be a smooth regular surface. Then for any point $p \in \Sigma$ there is $\varepsilon > 0$ such that any two points x and y in the ε-neightbohood of p in Σ can be connected by a unique minimizing geodesic $[xy]_\Sigma$.*

12.3 Polar coordinates

Proposition 12.3 implies existence of polar coordinates in a neighborhood of any point in p in Σ. That is, any point x in Σ sufficiently close to p can be uniquely described by the distance $|x - p|_\Sigma$ and the direction from p to x.

Assume (θ, r) are the described polar coordinates at p. Namely, assume $\tilde{w}(\theta, r)$ denotes the tangent vector at p with polar coordinates (θ, r) and $w(\theta, r) = \exp_p[\tilde{w}(\theta, r)]$. By the definition of exponential map, for a fixed θ, the curve $\gamma_\theta(t) = w(\theta, t)$ is a unit-speed geodesic that starts at p; in particular $|\frac{\partial}{\partial r} w| = |\gamma_\theta'(r)| = 1$ and $\gamma_\theta''(r) \perp \mathrm{T}_{\gamma_\theta(r)}$.

The curve $\sigma_r(t)=w(t,r)$ is a parametrization of the circle of radius r and center at p in Σ; that is, if $q=\sigma_r(t)$, then $|q-p|_\Sigma=r$. If the latter is not the case, then a minimizing geodesic $[pq]_\Sigma$ would be shorter than r and therefore q would not be described uniquely in the polar coordinates.

Note that $\frac{\partial}{\partial r}w\perp\frac{\partial}{\partial\theta}w$ if $r>0$; otherwise for small $\varepsilon>0$ the intrinsic distance from p to $w(\theta\pm\varepsilon,r)$ would be shorter than r, which contradicts the previous statement.

12.5. Proposition. *Let $w(\theta,r)$ and $\tilde w(\theta,r)$ be the polar coordinates of a surface Σ at p and its tangent plane T_p at zero, so $w(\theta,r)=$ $=\exp_p[\tilde w(\theta,r)]$. Given a real interval $[a,b]$ consider the one parameter families of circular arcs $\sigma_r\colon[a,b]\to\Sigma$ and $\tilde\sigma_r\colon[a,b]\to\mathrm{T}_p$ $\sigma_r(t)=$ $=w(t,r)$ and $\tilde\sigma_r(t)=\tilde w(t,r)$. Set $\ell(r)=\operatorname{length}\sigma_r$ and $\tilde\ell(r)=$ $=\operatorname{length}\tilde\sigma_r$.*[2]

(i) If the Gauss curvature of Σ is nonnegative, then

$$\ell(r)\leqslant\tilde\ell(r)$$

for all small $r>0$.

(ii) If the Gauss curvature of Σ is nonpositive, then

$$\ell(r)\geqslant\tilde\ell(r)$$

for all small $r>0$.

Taking a limit as $b\to a$, we obtain the following corollary.

12.6. Corollary. *Let $w(\theta,r)$ and $\tilde w(\theta,r)$ be the polar coordinates of a surface Σ at p and its tangent plane T_p at zero, so $w(\theta,r)=$ $=\exp_p[\tilde w(\theta,r)]$.*

(i) If the Gauss curvature of Σ is nonnegative, then

$$|\tfrac{\partial}{\partial\theta}w|\leqslant|\tfrac{\partial}{\partial\theta}\tilde w|$$

for all small $r>0$.

(ii) If the Gauss curvature of Σ is nonpositive, then

$$|\tfrac{\partial}{\partial\theta}w|\geqslant|\tfrac{\partial}{\partial\theta}\tilde w|$$

for all small $r>0$.

Proof. From the above discussion, the polar coordinates $w(\theta,r)$ are semigeodesic; that is, $w(\theta,r)$ satisfies the conditions in the first variation formula (12.1). In particular if $\ell(r)=\operatorname{length}\sigma_r$, then

$$\ell'(r)=\Theta_{\sigma_r}$$

[2]Note that angular measure of $\tilde\sigma_r$ is $b-a$; therefore $\tilde\ell(r)=r\cdot(b-a)$.

for any $r > 0$.

By Gauss–Bonnet formula, the last identity can be rewritten as

❶ $$\ell'(r) = 2\cdot(b-a) - \iint\limits_{\Delta_r} G,$$

where Δ_r is the sector in Σ in the polar coordinates at p

$$\{\, w(t,s) \;:\; a \leqslant t \leqslant b,\ 0 \leqslant s \leqslant r \,\};$$

which is bounded by two geodesics from p with angle $b-a$ and a circular arc that meets these geodesics at right angle.

Since the plane has vanishing Gauss curvature, we have

❷ $$\tilde\ell'(r) = 2\cdot(b-a),$$

which agrees with the formula for the length of the arc $\tilde\ell(r) = 2\cdot\pi\cdot r$.

If the Gauss curvature of Σ is nonnegative, the equations ❶ and ❷ imply that

$$\ell'(r) \leqslant \tilde\ell'(r)$$

for any small r.

If the Gauss curvature of Σ is nonnegative, the same equations imply that

$$\ell'(r) \geqslant \tilde\ell'(r)$$

for any small r.

Since $\ell(0) = \tilde\ell(0)$, integrating the inequalities proves both statements. □

The following exercise provides a stronger statement. It almost follow from the proof above, but one has to make an extra observation.

12.7. Exercise. *Assume Σ is a smooth regular surface and $p \in \Sigma$, denote by $\ell(r)$ the circumference of the circle with the center at p and radius r in Σ and let $\tilde\ell(r) = 2\cdot\pi\cdot r$ the circumference of the plane circle of radius r.*

(i) *Show that if Gauss curvature of Σ is nonnegative, then the function $r \mapsto \ell(r)$ is concave for small $r > 0$. Conclude that the function $r \mapsto \frac{\ell(r)}{\tilde\ell(r)}$ is nonincreasing for small $r > 0$.*

(ii) *Show that if Gauss curvature of Σ is nonpositive, then the function $r \mapsto \ell(r)$ is convex for small $r > 0$. Conclude that the function $r \mapsto \frac{\ell(r)}{\tilde\ell(r)}$ is nondecresing for small $r > 0$.*

12.4 Local comparison

The following proposition is a special case of the so a comparison theorem, proved by Harry Rauch [34].

12.8. Theorem. *Let Σ be a smooth regular surface and $p \in \Sigma$. Assume $\tilde\gamma\colon [a,b]$ is a curve the tangent plane $\mathrm{T}_p\Sigma$ that runs in a sufficiently small neighborhood of the origin; consider the curve*

$$\gamma = \exp_p \circ \gamma$$

in Σ.

(i) If Gauss curvature of Σ is nonnegative, then

$$\operatorname{length}\gamma \leqslant \operatorname{length}\tilde\gamma$$

(ii) If Gauss curvature of Σ is nonpositive, then

$$\operatorname{length}\gamma \geqslant \operatorname{length}\tilde\gamma.$$

The proof is a direct application of Corollary 12.6.

Proof. Let us denote $\tilde w(\theta, r)$ and $w(\theta, r)$ the polar coordinates of T_p and Σ at p Recall that

$$\begin{aligned} &\tfrac{\partial\tilde w}{\partial\theta} \perp \tfrac{\partial\tilde w}{\partial r}; && |\tfrac{\partial\tilde w}{\partial r}| = 1;\\ &\tfrac{\partial w}{\partial\theta} \perp \tfrac{\partial w}{\partial r}; && |\tfrac{\partial w}{\partial r}| = 1; \end{aligned}$$

By Corollary 12.6, we also have

$$|\tfrac{\partial\tilde w}{\partial\theta}| \geqslant |\tfrac{\partial\tilde w}{\partial\theta}|; \qquad |\tfrac{\partial\tilde w}{\partial\theta}| \leqslant |\tfrac{\partial\tilde w}{\partial\theta}|;$$

if Gauss curvature is nonnegative or nonpositive correspondingly.

It is sufficient to show that

$$❶ \qquad |\gamma'(t)| \leqslant |\tilde\gamma'(t)| \quad \text{or, correspondingly} \quad |\gamma'(t)| \geqslant |\tilde\gamma'(t)|$$

for any t.

Note that both curves $\gamma(t)$ and $\tilde\gamma(t)$ described the same way in the polar coordinates; denote these coordinates by $(\theta(t), r(t))$. Then

$$\begin{aligned} \gamma'(t)|^2 &= |\tfrac{\partial w}{\partial\theta}\cdot\theta'(t) + \tfrac{\partial w}{\partial r}\cdot r'(t)|^2 = \\ &= |\tfrac{\partial w}{\partial\theta}|^2\cdot|\theta'(t)|^2 + |r'(t)|^2 \end{aligned}$$

The same way

$$\tilde\gamma'(t)|^2 = |\tfrac{\partial\tilde w}{\partial\theta}|^2\cdot|\theta'(t)|^2 + |r'(t)|^2;$$

hence ❶ follows. $\square$

Chapter 13

Global comparison

13.1 Formulation

A minimizing geodesic between points x and y in a surface Σ will be denoted as $[xy]$ or $[xy]_\Sigma$; the latter notation is used if we need to emphasise that the geodesic lies in Σ. If we write $[xy]$, then we assume that a minimizing geodesic exists and we made a choice of one of them.

In general minimizing geodesic might be not unique for example any meridian in the sphere is a minimizing geodesic between its poles. If Σ is proper, then a minimizing geodesic always exists.

A *geodesic triangle* in a surface Σ is a triple of points $x, y, z \in \Sigma$ with choice of minimizing geodesics $[xy]$, $[yz]$ and $[zx]$. The points x, y, z are called *vertexes* of the geodesic triangle, the minimizing geodesics $[xy]$, $[yz]$ and $[zx]$ are called its sides; the triangle itself is denoted by $[xyz]$.

The length of one (and therefore any) minimizing geodesic $[xy]_\Sigma$ will be denoted by $|x-y|_\Sigma$; it is called *intrinsic distance* from x to y in Σ. If defined, then $|x-y|_\Sigma$ is the exact lower bound on the lengths of curves from x to y in Σ.

A triangle $[\tilde{x}\tilde{y}\tilde{z}]$ in the plane $\mathbb{R}^2$ is called *model triangle* of the triangle $[xyz]$, briefly $[\tilde{x}\tilde{y}\tilde{z}] = \tilde{\triangle}xyz$, if its corresponding sides are equal; that is,

$$|\tilde{x}-\tilde{y}|_{\mathbb{R}^2} = |x-y|_\Sigma, \quad |\tilde{y}-\tilde{z}|_{\mathbb{R}^2} = |y-z|_\Sigma, \quad |\tilde{z}-\tilde{x}|_{\mathbb{R}^2} = |z-x|_\Sigma.$$

A pair of minimizing geodesics $[xy]$ and $[xz]$ starting from one point x is called *hinge* and denoted as $[x\,{}^y_z]$. 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 the model triangle $[\tilde{x}\tilde{y}\tilde{z}] = \tilde{\triangle}xyz$ is denoted by $\tilde{\measuredangle}(x\,{}^y_z)$.

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 (trivial means that it stays at one point). A plane or sphere are examples of simply connected surfaces, while torus or cylinder are not simply connected.

13.1. Comparison theorem. *Let Σ be a proper smooth regular surface with a geodesic triangle $[xyz]$.*

(i) If Σ has nonnegative Gauss curvature, then

$$\measuredangle[x\,{}^y_z] \geqslant \tilde\measuredangle(x\,{}^y_z).$$

(ii) If Σ is simply connected and has nonpositive Gauss curvature, then

$$\measuredangle[x\,{}^y_z] \leqslant \tilde\measuredangle(x\,{}^y_z).$$

Let us make few remarks on the formulation.

The angle $\measuredangle[x\,{}^y_z]$ is a number in the interval $[0,\pi]$. If the triangle $[xyz]$ bounds a disc Δ and θ is the external angle at x which used in Gauss–Bonnet formula, then $\measuredangle[x\,{}^y_z] = |\pi - \theta|$. The corresponding internal angle might be $\measuredangle[x\,{}^y_z]$ or $2\cdot\pi - \measuredangle[x\,{}^y_z]$ depending on which side lies the disc Δ.

- ⋄ Since the angles of any plane triangle sum up to π, the part (i) of the theorem implies that angles of any triangle in a surface with nonnegative Gauss curvature have sum at least π.
- ⋄ The triangle may not bound a disc[1], but if it does, then by Gauss–Bonnet formula the sum of its *internal* angles is at least π.

These two statements are closely related, but they are not the same. Note that if α is the angle in the comparison theorem, then the internal angle might be α or $2\cdot\pi-\alpha$; while Gauss–Bonnet formula gives a lower bound on the sum of internal angles it does not forbid that each of these angles is close to $2\cdot\pi$. However the latter is impossible by the comparison theorem.

First note that without condition that Σ is simply connected, the statement *(ii)* does not hold. For example the equator $z = 0$ of the hyperboloid (which is not simply connected)

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

forms a triangle with all angles π, which contradict the comparison in *(ii)*.

13.2. Exercise. *Let Σ be an open smooth regular simply connected surface with nonpositive Gauss curvature. Show that any two points in Σ are connected by a unique geodesic.*

[1]For example equator on the cyclinder is formed by a geodesic triangle that does not bound a disc.

13.2 Names and history

Part *(i)* of this theorem is called *Toponogov comparison theorem*; it is was proved by Paolo Pizzetti [35] and latter independently by Alexandr Alexandrov [36]; generalizations were obtained by Victor Toponogov [37], Mikhael Gromov, Yuri Burago and Grigory Perelman [38].

Part *(ii)* is called *Cartan–Hadamard theorem*; it was proved by Hans von Mangoldt [39] and generalized by Elie Cartan [40], Jacques Hadamard [25], Herbert Busemann [41], Willi Rinow in [42], Mikhael Gromov [43, p. 119], Stephanie Alexander and Richard Bishop in [44].

13.3 Local part

First we prove the following local version of comparison theorem and then use it to prove the global version.

13.3. Theorem. *The comparison theorem (13.1) holds in a small neighborhood of any point.*

That is, if Σ be a smooth regular surface without boundary, then any point $p \in \Sigma$ admits a neighborhood $U \ni p$ such that

(i) If Σ has nonnegative Gauss curvature, then for any geodesic triangle $[xyz]$ in U we have

$$\measuredangle[x\,{}^y_z] \geqslant \tilde\measuredangle(x\,{}^y_z).$$

(ii) If Σ has nonpositive Gauss curvature, then for any geodesic triangle $[xyz]$ in U we have

$$\measuredangle[x\,{}^y_z] \leqslant \tilde\measuredangle(x\,{}^y_z).$$

Note that we can assume that U is simply connected therefore this condition is not necessary to include in part *(ii)*.

Proof. Assume $y = \exp_x v$ and $z = \exp_x w$ for two small vectors $v, w \in \mathrm{T}_x$. Note that

$$\begin{aligned}\measuredangle[x\,{}^v_w]_{\mathrm{T}_x} &= \measuredangle[x\,{}^y_z]_\Sigma,\\ |x-v|_{\mathrm{T}_x} &= |x-y|_\Sigma,\\ |x-w|_{\mathrm{T}_x} &= |x-z|_\Sigma.\end{aligned}$$

If the Gauss curvature is nonnegative, 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 theorem (12.8), we have

$$\operatorname{length}\gamma \leqslant \operatorname{length}\tilde\gamma.$$

Since $|v-w|_{\mathrm{T}_x}=\operatorname{length}\tilde\gamma$ and $|y-z|_\Sigma\leqslant\operatorname{length}\gamma$, we get

$$|v-w|_{\mathrm{T}_x}\geqslant|y-z|_\Sigma.$$

Therefore

$$\tilde\measuredangle(x\,{}^z_y)\geqslant\measuredangle[x\,{}^z_y].$$

If the Gauss curvature is nonpositive, consider a minimizing geodesic γ joining y to z in Σ and let $\tilde\gamma$ be the corresponding curve joining v to w in T_x; that is, $\gamma=\exp_x\circ\tilde\gamma$. By Rauch comparison theorem (12.8), we have

$$\operatorname{length}\gamma\geqslant\operatorname{length}\tilde\gamma.$$

Since $|v-w|_{\mathrm{T}_x}\leqslant\operatorname{length}\tilde\gamma$ and $|y-z|_\Sigma=\operatorname{length}\gamma$, we get

$$|v-w|_{\mathrm{T}_x}\geqslant|y-z|_\Sigma.$$

Therefore

$$\tilde\measuredangle(x\,{}^z_y)\geqslant\measuredangle[x\,{}^z_y].$$ □

13.4 Alexandrov's lemma

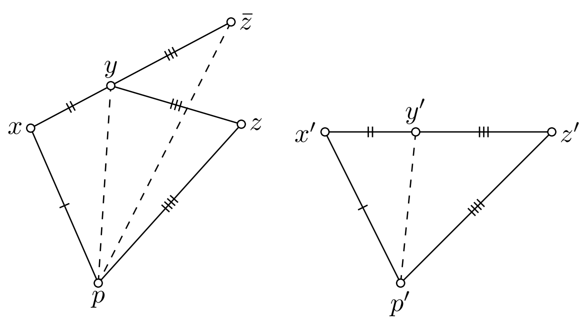
It this section we prove the following lemma in the plane geometry.

13.4. Lemma. *Assume $[pxyz]$ and $[p'x'y'z']$ be two quadraliterals 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. In the proof we use the following *monotonicity property*: if two sides adjacent to an angle in a plane triangle are fixed, then the angle is increases if the opposite side increase.

Take a point $\bar{z}$ on the extension of $[xy]$ beyond y so that $|y - \bar{z}| = |y - z|$ (and therefore $|x - \bar{z}| = |x' - z'|$).

From monotonicity, 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) $|p - \bar{z}| - |p' - z'|$;

(iv) $\measuredangle[y\,{}^{\bar{z}}_{p}] - \measuredangle[y'\,{}^{z'}_{p'}]$;

The statement follows since

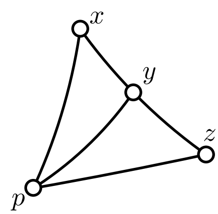
$$\measuredangle[y'\,{}^{z'}_{p'}] + \measuredangle[y'\,{}^{x'}_{p'}] = \pi$$

and

$$\measuredangle[y\,{}^{\bar{z}}_{p}] + \measuredangle[y\,{}^{x}_{p}] = \pi. \qquad \square$$

Further we will use the following reformulation of this lemma that is using language of comparison triangles and angles.

13.5. Reformulation. *Assume $[pxz]$ be a triangle in a surface Σ and the point y lies on the side $[xz]$. 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) $\pi - \tilde{\measuredangle}(y\,{}^{p}_{x}) - \tilde{\measuredangle}(y\,{}^{p}_{z})$;

13.5 Reformulations of comparison

In this section we formulate conditions equivalent to the conclusion of the comparison theorem (13.1).

A triangle $[xyz]$ in a surface is called *fat* (or correspondingly *thin*) if for any two points p and q on the sides of the triangle and the corresponding points $\tilde{p}$ and $\tilde{q}$ on the sides of its model triangle $[\tilde{x}\tilde{y}\tilde{z}] = \tilde{\triangle}xyz$ we have $|p-q| \geqslant |\tilde{p}-\tilde{q}|$ (or correspondingly $|p-q| \leqslant |\tilde{p}-\tilde{q}|$).

13.6. Proposition. *Let Σ be a proper smooth regular surface. Then the following three conditions are equivalent:*

(i⁺) *For any geodesic triangle $[xyz]$ in Σ we have*

$$\measuredangle[x\,{}^{y}_{z}] \geqslant \tilde{\measuredangle}(x\,{}^{y}_{z}).$$

(ii⁺) *For any geodesic triangle $[pxz]$ in Σ and y on the side $[xz]$ we have*

$$\tilde{\measuredangle}(x\,{}^{p}_{y}) \geqslant \tilde{\measuredangle}(x\,{}^{p}_{z}).$$

(iii$^+$) *Any geodesic triangle in Σ is fat.*

Similarly, following three conditions are equivalent:

(i$^-$) *For any geodesic triangle $[xyz]$ in Σ we have*

$$\measuredangle[x\,{}^y_z] \leqslant \tilde\measuredangle(x\,{}^y_z).$$

(ii$^-$) *For any geodesic triangle $[pxz]$ in Σ and y on the side $[xz]$ we have*

$$\tilde\measuredangle(x\,{}^p_y) \leqslant \tilde\measuredangle(x\,{}^p_z).$$

(iii$^-$) *Any geodesic triangle in Σ is thin.*

Proof. We will prove the implications *(i$^+$)⇒(ii$^+$)⇒(iii$^+$)⇒(i$^+$).* The implications *(i$^-$)⇒(ii$^-$)⇒(iii$^-$)⇒(i$^-$)* can be done the same way.

(i$^+$)⇒(ii$^+$). Note that $\measuredangle[y\,{}^p_x] + \measuredangle[y\,{}^p_z] = \pi$. By *(i$^+$)*,

$$\tilde\measuredangle(y\,{}^p_x) + \tilde\measuredangle(y\,{}^p_z) \leqslant \pi.$$

It reamains to apply Alexandrov's lemma (13.5).

(ii$^+$)⇒(iii$^+$). Applying *(i$^+$)* twice, first for $y \in [xz]$ and then for $w \in [px]$, 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. Hence the implication follows.

(iii$^+$)⇒(i$^+$). 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. $\square$

In the following exercises you can apply the globalization theorem.

13.7. Exercise. *Let Σ be a closed (or open) regular surface and with nonnegative Gauss curvature. Show that for any four distinct points the following inequality holds:*

$$\tilde\measuredangle(p\,{}^x_y) + \tilde\measuredangle(p\,{}^y_z) + \tilde\measuredangle(p\,{}^z_x) \leqslant 2{\cdot}\pi.$$

13.8. Exercise. *Let Σ be a open smooth regular surface and γ be 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 the Gauss curvature of* Σ *is nonnegative, then* h *is a concave function.*
(b) *Show that if* Σ *is simply connected and the Gauss curvature of* Σ *is nonpositive, then* h *is a convex function.*

13.9. Exercise. *Let* $\tilde{x}_1 \dots \tilde{x}_n$ *be a convex plane polygon and* $x_1 \dots x_n$ *be a broken geodesic in an open simply connected surface* Σ *with nonpositive curvature. Assume that* $|x_i - x_{i-1}|_\Sigma = |\tilde{x}_i - \tilde{x}_{i-1}|_{\mathbb{R}^2}$ *and* $\measuredangle[x_{i-1}\,{}^{x_i}_{x_{i+1}}] \geqslant \measuredangle[x_{i-1}\,{}^{x_i}_{x_{i+1}}]$ *for each* i*. Show that*

$$|x_1 - x_n|_\Sigma \geqslant |\tilde{x}_1 - \tilde{x}_n|_{\mathbb{R}^2}.$$

For $\Sigma = \mathbb{R}^2$, the exercise above is the so called *arm lemma*; you can use it without proof.

13.10. Exercise. *Let* x' *and* y' *be the midpoints of minimizing geodesics* $[px]$ *and* $[py]$ *in an open smooth regular surface* Σ*.*
(a) *Show that if the Gauss curvature of* Σ *is nonnegative, then*

$$2{\cdot}|x' - y'|_\Sigma \geqslant |x - y|_\Sigma.$$

(b) *Show that if* Σ *is simply connected and has nonpositive Gauss curvature, then*

$$2{\cdot}|x' - y'|_\Sigma \leqslant |x - y|_\Sigma.$$

13.6 Nonnegative curvature

In this section we will prove part *(i)* of the comparison theorem (13.1) assuming that Σ is compact; the general case requre only minor modifications.

Since Σ is compact, from the local theorem (13.3), we get that there is $\varepsilon > 0$ such that the inequality

$$\measuredangle[x\,{}^y_z] \geqslant \tilde\measuredangle(x\,{}^y_z).$$

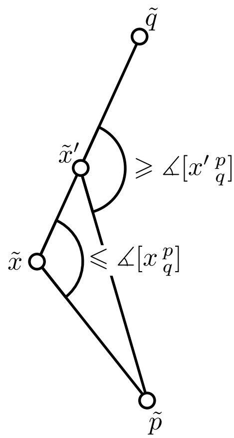
holds for any hinge $[x\,{}^y_z]$ such that $|x - y| + |x - z| < \varepsilon$. The following lemma states that in this case the same holds for any hinge $[x\,{}^y_z]$ such that $|x - y| + |x - z| < \frac{3}{2}{\cdot}\varepsilon$. Applying the lemma few times we will get that the comparison holds for arbitrary hinge, which will prove part *(i)*.

13.11. Key lemma. *Let* Σ *be an open smooth regular surface. Assume that the comparison*

➊ $$\measuredangle[x\,{}^y_z] \geqslant \tilde\measuredangle(x\,{}^y_z)$$

holds for any hinge $[x\,{}^{y}_{z}]$ *with* $|x-y|+|x-z|<\tfrac{2}{3}\cdot\ell$. *Then the comparison* ❶ *holds for any hinge* $[x\,{}^{y}_{z}]$ *with* $|x-y|+|x-z|<\ell$.

Proof. Given a hinge $[x\,{}^{p}_{q}]$ consider a triangle in the plane with angle $\measuredangle[x\,{}^{p}_{q}]$ and two adjacent sides $|x-p|$ and $|x-q|$. Let us denote by $\tilde{\Upsilon}[x\,{}^{p}_{q}]$ the third side of this triangle; let us call it *model side* of the hinge.



Note that the inequalities

$$\measuredangle[x\,{}^{p}_{q}] \geqslant \tilde{\measuredangle}(x\,{}^{p}_{q}) \quad\text{and}\quad \tilde{\Upsilon}[x\,{}^{p}_{q}] \geqslant |p-q|$$

are equivalent. So it is sufficient to prove that

❷ $$\tilde{\Upsilon}[x\,{}^{p}_{q}] \geqslant |p-q|.$$

for any hinge $[x\,{}^{p}_{q}]$ with $|x-p|+|x-q|<\ell$.

Given a hinge $[x\,{}^{p}_{q}]$ such that

$$\tfrac{2}{3}\cdot\ell \leqslant |p-x|+|x-q|<\ell,$$

let us construct a new smaller hinge $[x'\,{}^{p}_{q}]$; that is,

❸ $$|p-x|+|x-q| \geqslant |p-x'|+|x'-q|$$

and such that

❹ $$\tilde{\Upsilon}[x\,{}^{p}_{q}] \geqslant \tilde{\Upsilon}[x'\,{}^{p}_{q}].$$

Assume $|x-q|\geqslant|x-p|$, otherwise switch the roles of p and q in the following construction. Take $x'\in[xq]$ such that

❺ $$|p-x|+3\cdot|x-x'|=\tfrac{2}{3}\cdot\ell$$

Choose a geodesic $[x'p]$ and consider the hinge $[x'\,{}^{p}_{q}]$ fromed by $[x'p]$ and $[x'q]\subset[xq]$. Then ❸ follows since the length of $[x'p]$ can not exceed the total length of $[x'x]$ and $[x'p]$.

Further, note that $|p-x|+|x-x'|, |p-x'|+|x'-x| < \tfrac{2}{3}\cdot\ell$. In particular,

❻ $$\measuredangle[x\,{}^{p}_{x'}] \geqslant \tilde{\measuredangle}(x\,{}^{p}_{x'}) \ \text{ and }\ \measuredangle[x'\,{}^{p}_{x}] \geqslant \tilde{\measuredangle}(x'\,{}^{p}_{x}).$$

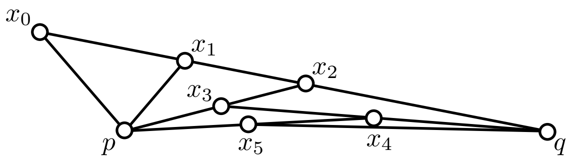
Consider the model triangle $[\tilde{x}\tilde{x}'\tilde{p}] = \tilde{\triangle}xx'p$. Take $\tilde{q}$ on the extension of $[\tilde{x}\tilde{x}']$ beyond x' such that $|\tilde{x}-\tilde{q}|=|x-q|$ (and therefore $|\tilde{x}'-\tilde{q}|=|x'-q|$). From ❻,

$$\measuredangle[x\,{}^{p}_{q}] = \measuredangle[x\,{}^{p}_{x'}] \geqslant \tilde{\measuredangle}(x\,{}^{p}_{x'}) \quad\Rightarrow\quad \tilde{\Upsilon}[x\,{}^{q}_{p}] \geqslant |\tilde{p}-\tilde{q}|.$$

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 $|\tilde p - \tilde q| \geqslant \tilde\Upsilon[x'\,{}^{q}_{p}]$ and ❹ follows.



Set $x_0 = x$. Let us apply inductively the above construction to get a sequence of hinges $[x_n\,{}^{p}_{q}]$ with $x_{n+1} = x'_n$. By ❹ and triangle inequality, both sequences

$$s_n = \tilde\Upsilon[x_n\,{}^{p}_{q}] \quad\text{and}\quad r_n = |p - x_n| + |x_n - q|$$

are nonincreasing.

The sequence might terminate at some n only if $r_n < \tfrac{2}{3}\cdot\ell$. In this case, by the assumptions of the lemma,

$$s_n = \tilde\Upsilon[x_n\,{}^{p}_{q}] \geqslant |p - q|.$$

Since sequence s_n is nonincreasing;

$$s_0 = \tilde\Upsilon[x\,{}^{p}_{q}] \geqslant |p - q|,$$

whence inequality ❷ follows.

If the sequence does not terminate, then $r_n \geqslant \tfrac{2}{3}\cdot\ell$ for all n. Since (r_n) is nonincreasing, $r_n \to r \geqslant |p - q|_\Sigma$ as $n \to \infty$.

Let us show that $\measuredangle[x_n\,{}^{p}_{q}] \to \pi$ as $n \to \infty$.

Indeed assume $\measuredangle[x_n\,{}^{p}_{q}] \leqslant \pi - \varepsilon$ for some $\varepsilon > 0$. Without loss of generality we can assume that $x_{n+1} \in [x_n q]$; otherwise switch p and q further. Note that $|x_n - x_{n+1}|, |p - x_n| > \tfrac{\ell}{100}$. Therefore by comparison

$$|p - x_{n+1}| < \tilde\Upsilon[x_n\,{}^{p}_{x_{n+1}}] < |p - x_n| + |x_n - x_{n+1}| - \delta$$

for some fixed $\delta = \delta(\varepsilon) > 0$. Therefore $r_n - r_{n+1} > \delta$. The latter can not hold for large n, otherwise the sequence r_n would not converge.

It follows that for any $\varepsilon > 0$ we have that $\measuredangle[x_n\,{}^{p}_{q}] > \pi - \varepsilon$ for all large n; that is, $\measuredangle[x_n\,{}^{p}_{q}] \to \pi$ as $n \to \infty$.

Since $\measuredangle[x_n\,{}^{p}_{q}] \to \pi$, we have $s_n - r_n \to 0$ as $n \to \infty$; that is, $s_n \to r$.

Since the sequence (s_n) is nonincreasing and $r \geqslant |p - q|$, we get

$$s_n \geqslant |p - q|$$

for any n. In particular

$$\tilde{\Upsilon}[s\,{}^{p}_{q}] = s_0 \geqslant |p-q|,$$

so we obtain ❷. □

13.12. Exercise. *Assume a disc Δ lies in open smooth regular surface Σ with nonnegative Gauss curvature and bounded by a closed broken geodesic $x_1 \dots x_n$ with positive exterior angles; that is, when you travel along the boundary, you always turn to the side where Δ is.*

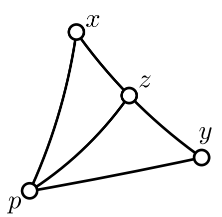
Show that there is a convex plane polygon $\tilde{x}_1 \dots \tilde{x}_n$ which sides are equal to the corresponding sides of $x_1 \dots x_n$ and with internal angles at not bigger than in the corresponding angles of $x_1 \dots x_n$.

13.7 Inheritance lemma

The following lemma will play key role in the proof of part *(ii)* of the comparison theorem (13.1).

13.13. Inheritance Lemma. *Assume that a triangle $[pxy]$ in a surface Σ* decomposes *into two triangles $[pxz]$ and $[pyz]$; that is, $[pxz]$ and $[pyz]$ have common side $[pz]$, and the sides $[xz]$ and $[zy]$ together form the side $[xy]$ of $[pxy]$.*

If both triangles $[pxz]$ and $[pyz]$ are thin, then so is $[pxy]$.



We shall need the following lemma in plane geometry.

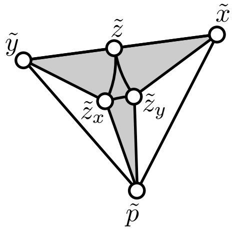
13.14. Lemma. *Let $\blacktriangle\tilde{p}\tilde{x}\tilde{y}$ be a solid plane triangle; that is, $\blacktriangle\tilde{p}\tilde{x}\tilde{y} = \mathrm{Conv}\{\tilde{p},\tilde{x},\tilde{y}\}$. Given $\tilde{z} \in [\tilde{x}\tilde{y}]$, consider points $\dot{p},\dot{x},\dot{z},\dot{y}$ in the plane such that*

$$\begin{aligned} &|\dot{p}-\dot{x}| = |\tilde{p}-\tilde{x}|, && |\dot{p}-\dot{y}| = |\tilde{p}-\tilde{y}|, && |\dot{p}-\dot{z}| \leqslant |\tilde{p}-\tilde{z}|, \\ &|\dot{x}-\dot{z}| = |\tilde{x}-\tilde{z}|, && |\dot{y}-\dot{z}| = |\tilde{y}-\tilde{z}|, \end{aligned}$$

where points $\dot{x}$ and $\dot{y}$ lie on either side of $[\dot{p}\dot{z}]$. Then there is a short map

$$F\colon \blacktriangle\tilde{p}\tilde{x}\tilde{y} \to \blacktriangle\dot{p}\dot{x}\dot{z} \cup \blacktriangle\dot{p}\dot{y}\dot{z}$$

that maps $\tilde{p}$, $\tilde{x}$, $\tilde{y}$ and $\tilde{z}$ to $\dot{p}$, $\dot{x}$, $\dot{y}$ and $\dot{z}$ respectively.



Proof. Note that

$$\begin{aligned}|\dot x-\dot y| &\leqslant |\dot x-\dot z|+|\dot z-\dot y| = \\ &= |\tilde x-\tilde z|+|\tilde z-\tilde y| \\ &= |\tilde x-\tilde y|.\end{aligned}$$

Applying monotonicity propery, we get that

$$\measuredangle[\dot p\,{}^{\dot x}_{\dot y}] \leqslant \measuredangle[\tilde p\,{}^{\tilde x}_{\tilde y}].$$

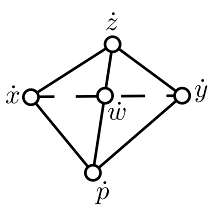
It follows that there are nonoverlapping triangles $[\tilde p\tilde x\tilde z_y]\cong[\dot p\dot x\dot z]$ and $[\tilde p\tilde y\tilde z_x]\cong[\dot p\dot y\dot z]$ inside triangle $[\tilde p\tilde x\tilde y]$.

Connect points in each pair $(\tilde z,\tilde z_x)$, $(\tilde z_x,\tilde z_y)$ and $(\tilde z_y,\tilde z)$ with arcs of circles centered at $\tilde y$, $\tilde p$, and $\tilde x$ respectively. Define F as follows.

- ◇ Map $\blacktriangle\tilde p\tilde x\tilde z_y$ isometrically onto $\blacktriangle\dot p\dot x\dot y$; similarly map $\blacktriangle\tilde p\tilde y\tilde z_x$ onto $\blacktriangle\dot p\dot y\dot z$.
- ◇ If a point w lies in one of the three circular sectors, say at distance r from center of the circle, let $F(w)$ be the point on the corresponding segment $[\dot p\dot z]$, $[\dot x\dot z]$ or $[\dot y\dot z]$ whose distance from the lefthand endpoint of the segment is r.
- ◇ Finally, if w lies in the remaining curvilinear triangle $\tilde z\tilde z_x\tilde z_y$, set $F(w)=\dot z$.

By construction, F satisfies the remaining conditions of the lemma. □

Proof of Inheritance lemma 13.13. Construct model triangles $[\dot p\dot x\dot z]= =\tilde\triangle(pxz)$ and $[\dot p\dot y\dot z]=\tilde\triangle(pyz)$ so that $\dot x$ and $\dot y$ lie on opposite sides of $[\dot p\dot z]$.



Suppose

$$\tilde\measuredangle(z\,{}^{p}_{x})+\tilde\measuredangle(z\,{}^{p}_{y})<\pi.$$

Then for some point $\dot w\in[\dot p\dot z]$, we have

$$|\dot x-\dot w|+|\dot w-\dot y|<|\dot x-\dot z|+|\dot z-\dot y|=|x-y|.$$

Let $w\in[pz]$ correspond to $\dot w$; that is, $|z-w|=|\dot z-\dot w|$. Since $[pxz]$ and $[pyz]$ are thin, we have

$$|x-w|+|w-y|<|x-y|,$$

contradicting the triangle inequality.

Thus

$$\tilde\measuredangle(z\,{}^{p}_{x})+\tilde\measuredangle(z\,{}^{p}_{y})\geqslant\pi.$$

By Alexandrov's lemma (13.5), this is equivalent to

$$❶\qquad \tilde\measuredangle(x\,{}^{p}_{z}) \leqslant \tilde\measuredangle(x\,{}^{p}_{y}).$$

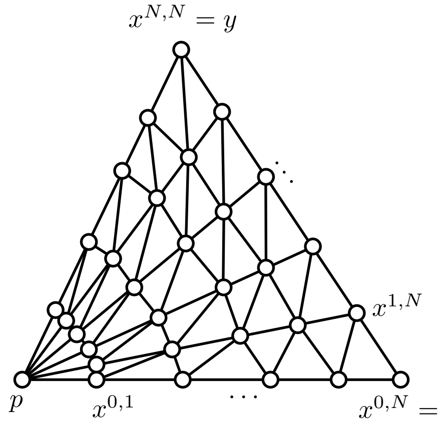
Let $[\tilde p\tilde x\tilde y] = \tilde\triangle(pxy)$ and $\tilde z \in [\tilde x\tilde y]$ correspond to z; that is, $|x - z| = |\tilde x - \tilde z|$. Inequality ❶ is equivalent to $|p - z| \leqslant |\tilde p - \tilde z|$. Hence Lemma 13.14 applies; let $F\colon \blacktriangle\tilde p\tilde x\tilde y \to \blacktriangle\dot p\dot x\dot z \cup \blacktriangle\dot p\dot y\dot z$ be the provided short map.

Fix v, w on the sides of $[pxy]$; let $\tilde v, \tilde w$ be the corresponding points on the sides of the model triangle $[\tilde p\tilde x\tilde y] = \tilde\triangle pxy$ and $\dot v, \dot w$ be the corresponding points on the sides of the model triangles $[\dot p\dot x\dot z] = \tilde\triangle pxz$ and $[\dot p\dot y\dot z] = \tilde\triangle pyz$. Denote by ℓ the length of shortest curve from $\dot v$ to $\dot w$ in $\blacktriangle\dot p\dot x\dot z \cup \blacktriangle\dot p\dot y\dot z$. Since F is short, $|\tilde v - \tilde w|_{\mathbb{R}^2} \geqslant \ell$. Since both triangles $[pxz]$ and $[pyz]$ are thin, $\ell \geqslant |v - w|_\Sigma$.

It follows that $|\tilde v - \tilde w|_{\mathbb{R}^2} \geqslant |v - w|_\Sigma$ for any v and w; that is, the triangle $[pxy]$ is thin. □

13.8 Nonpositive curvature

Assume Σ is an open smooth regular surface with nonpositive curvature. As it follow from Exercise 13.2 any two points x and y in Σ are joined by unique geodesic $[xy]$.



Note that the geodesic $[xy]$ depends continuously on its endpoints x and y. That is, if $\gamma_{[xy]}\colon [0,1] \to \Sigma$ is the constant speed parametrization of $[xy]$ from x to y, then the map $(x, y, t) \mapsto \gamma_{[xy]}(t)$ is continuous in three arguments. Indeed, assume contrary, that is, $x_n \to x$, $y_n \to y$ and $t_n \to t$ as $n \to \infty$ and $\gamma_{[x_ny_n]}(t_n)$ does not converge to $\gamma_{[xy]}(t)$. Then we can pass to a subsequence such that $\gamma_{[x_ny_n]}(t_n)$ converges to a point distinct from $w \ne \gamma_{[xy]}(t)$. Note that $w \notin [xy]$. Therefore there will be two distinct geodesics from x to y; one is $[xy]$ and the other is the limit of $[x_ny_n]$ which passes thru w.

Proof of part (ii) of the comparison theorem (13.1). Fix a triangle $[pxy]$; by Proposition 13.6, it is sufficient to show that the triangle $[pxy]$ is thin.

Fix large integer N and divide $[xy]$ by points $x = x^{0,N}, \dots, x^{N,N} = y$ into N equal parts. Further divide each geodesic $[p\,x^{i,N}]$ into N equal parts by points $p = x^{i,0}, \dots, x^{i,N}$. Since the geodesic depends continuously on its end points, we can assume that each triangle $[x^{i,j}\,x^{i,j+1}\,x^{i+1,j+1}]$ and $[x^{i,j}\,x^{i+1,j}\,x^{i+1,j+1}]$ is small; in particular, by local comparison (13.3), each of these triangles is thin.

Now we show that the thin property propagates to $[pxy]$ by repeated application of the inheritance lemma (13.13):

- ⋄ First, for fixed i, sequentially applying the lemma shows that the triangles $[x\,x^{i,1}\,x^{i+1,2}]$, $[x\,x^{i,2}\,x^{i+1,2}]$, $[x\,x^{i,2}\,x^{i+1,3}]$, and so on are thin.

In particular, for each i, the long triangle $[x\,x^{i,N}\,x^{i+1,N}]$ is thin.

- ⋄ Applying the lemma again shows that the triangles $[x\,x^{0,N}\,x^{2,N}]$, $[x\,x^{0,N}\,x^{3,N}]$, and so on are thin.

In particular, $[pxy] = [p\,x^{0,N}\,x^{N,N}]$ is thin. □

13.15. Exercise. *Assume γ_1 and γ_2 be 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.

Appendix A

Review

Here we state and discuss results from different branches of mathematics which were used further in the book. The reader is not expected to know proofs of these statements, but it is better to check that his intuition agrees with each.

A.1 Metric spaces

Metric is a function that returns a real value $\operatorname{dist}(x,y)$ 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:
$$\operatorname{dist}(x,y) \geqslant 0.$$

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

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

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

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 a shortcut notations

$|x-y|$ or $|x-y|_\mathcal{X}$ for the *distance* $\operatorname{dist}(x,y)$ *from* x *to* y *in* $\mathcal{X}$. For example, the triangle inequality can be written as

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

Examples. Euclidean space and plane as well as 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 a norm of the vector $x-y$. However, in general metric space the expression $x-y$ has no sense, but anyway we use expression $|x-y|$ for the distance.

If we say *plane* or *space* we mean *Eucledean* 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.

A.1. Exercise. *Consider the function*

$$\operatorname{dist}(p,q) = |x_p - x_q| + |y_p - y_q|,$$

where $p = (x_p, y_p)$ *and* $q = (x_q, y_q)$ *are points in the coordinate plane* $\mathbb{R}^2$. *Show that* dist *is a metric on* $\mathbb{R}^2$.

Let us mention another example: the *discrete space* — 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 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 (or at most R) from p is called open (or correspondingly closed) ball of radius R with center at 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

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

A.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* $B(p,\tfrac{3}{2}) \subset B(q,1)$ *and the inclusions is strict.*

A.2 Continuity

In this section we will extend standard notions from calculus to the metric spaces.

A.3. Definition. *Let $\mathcal{X}$ be a metric space. A sequence of points $x_1, x_2, \dots$ 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*

$$|x_\infty - x_n| < \varepsilon.$$

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 $x_\infty = \lim_{n\to\infty} x_n$.*

A.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

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

A.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,*

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

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

A.6. Definition. *Let $\mathcal{X}$ and $\mathcal{Y}$ be metric spaces. A continous 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.

A.7. Definition. *A subset A of a metric space $\mathcal{X}$ is called* closed *if whenever a sequence (x_n) of points from A converges in $\mathcal{X}$, we have that $\lim_{n\to\infty} x_n \in A$.*

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

An open set Ω that contains a given point p is called *neighborhood of* p.

A.8. Exercise. *Let Q be a subset of a metric space $\mathcal{X}$. Show that A is closed if and only if its complement $\Omega = \mathcal{X}\backslash Q$ is open.*

A.3 Regular values

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

$$f_1, \dots, f_k \colon \mathbb{R}^n \to \mathbb{R}.$$

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 $\boldsymbol{f}$.

Inverse function theorem gives a sufficient condition for a smooth function to be invertible in a neighborhood of a given point p in its domain. The condition is formulated in terms of partial derivative of f_i at p.

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

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

These two theorems are discussed in any course of multivariable calculus, the classical book of Walter Rudin [45] is one of my favorites.

A.9. Inverse function theorem. *Let $\boldsymbol{f} = (f_1, \dots, f_n)\colon \mathbb{R}^n \to \mathbb{R}^n$ be a smooth map. Assume that the Jacobian matrix*

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

is invertible at some point p in the domain of definition of $\boldsymbol{f}$. Then there is a smooth function $\boldsymbol{h}\colon \mathbb{R}^m \to \mathbb{R}^n$ defined is a neighborhood Ω_q of $q = \boldsymbol{f}(p)$ that is local inverse of $\boldsymbol{f}$ at *p; that is, there are neighborhoods $\Omega_p \ni p$ such that $\boldsymbol{f}$ defines a bijection $\Omega_p \to \Omega_q$ and $\boldsymbol{f}(x) = y$ if and only if $x = \boldsymbol{h}(y)$ for any $x \in \Omega_p$ and any $y \in \Omega_q$.*

A.10. Implicit function theorem. *Let $\boldsymbol{f} = (f_1, \dots, f_n)\colon \mathbb{R}^{n+m} \to \mathbb{R}^n$ be a smooth map, $m, n \geqslant 1$. Let us consider $\mathbb{R}^{n+m}$ as a product* *space $\mathbb{R}^n \times \mathbb{R}^m$ with coodinates $x_1, \dots, x_n, y_1, \dots, y_m$. Consider the following matrix*

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

formed by first n columns of the Jacobian matrix. Assume M is invertible at some point p in the domain of definition of $\boldsymbol{f}$ and $\boldsymbol{f}(p) = 0$. Then there is a neighborhood $\Omega_p \ni p$ and smooth function $\boldsymbol{h}\colon \mathbb{R}^m \to \mathbb{R}^n$ defined is a neighborhood $\Omega_0 \ni 0$ that for any $(x_1, \dots, x_n, y_1, \dots y_m) \in \Omega_p$ the equality

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

holds if and only if

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

If the assumption in the theorem holds for any point p such that $\boldsymbol{f}(p) = 0$, then we say that 0 is a regular value of $\boldsymbol{f}$. The following lemma states that most of the values of smooth map are regular; in particular generic smooth function satisfies the assumption of the theorem.

A.11. Sard's lemma. *Almost all values of a smooth map $\boldsymbol{f}\colon U \to \to \mathbb{R}^m$ defined on an open set $U \subset \mathbb{R}^n$ are regular.*

The words *almost all* means that with exception o set of zero Lebesgue measure. In particular if one chooses a random value equidistributed in arbitrarily small ball $B \subset \mathbb{R}^m$, then it is a regular value of f with probability 1.

A.4 Multiple integral

The following theorem is a substitution rule for multiple variables.

A.12. Theorem. *Let $K \subset \mathbb{R}^n$ be a compact set and $h\colon K \to \mathbb{R}$ be a bounded measurable function. Assume $\boldsymbol{f}\colon K \to \mathbb{R}^n$ is an injective smooth map. Then*

$$\int\limits_K h(\boldsymbol{x}) \cdot |J_{\boldsymbol{f}}(\boldsymbol{x})| \cdot d\boldsymbol{x} = \int\limits_{\boldsymbol{f}(K)} h \circ \boldsymbol{f}^{-1}(\boldsymbol{y}) \cdot d\boldsymbol{y},$$

where $J_{\boldsymbol{f}}(\boldsymbol{x})$ denotes the Jacobian of $\boldsymbol{f}$ at $\boldsymbol{x}$; that is, the determinant of the Jacobian matrix of $\boldsymbol{f}$ at $\boldsymbol{x}$.

A.13. Area formula. *Let $\boldsymbol{f}\colon K \to \mathbb{R}^n$ be a smooth map defined on a compact set $K \subset \mathbb{R}^n$; denote by $J_{\boldsymbol{f}}$ the Jacobian of $\boldsymbol{f}$. Then for any function $h\colon K \to \mathbb{R}$*

$$\int\limits_K h(\boldsymbol{x})\cdot|J_{\boldsymbol{f}}(\boldsymbol{x})|\cdot d\boldsymbol{x} = \int\limits_{\boldsymbol{f}(K)} H_K(\boldsymbol{y})\cdot d\boldsymbol{y},$$

where

$$H_K(\boldsymbol{y}) = \sum_{\substack{\boldsymbol{x}\in K\\ \boldsymbol{f}(\boldsymbol{x})=\boldsymbol{y}}} h(\boldsymbol{x}).$$

(The integrals are understood in the sense of Lebesgue.)

Let us sketch the proof of area formula using Sard's lemma (A.11) and the substitution rule (A.12).

Sketch of proof. Denote by $S \subset K$ the set of critical points of $\boldsymbol{f}$; that is, $\boldsymbol{x} \in S$ if $J_{\boldsymbol{f}}(\boldsymbol{x}) = 0$. By Sard's lemma, $\boldsymbol{f}(S)$ *has vanishing measure.* Note that

$$\int\limits_S h(\boldsymbol{x})\cdot|J_{\boldsymbol{f}}(\boldsymbol{x})|\cdot d\boldsymbol{x} = 0$$

since $J_{\boldsymbol{f}}(\boldsymbol{x}) = 0$ and

$$\int\limits_{\boldsymbol{f}(S)} H_S(\boldsymbol{y})\cdot d\boldsymbol{y}$$

since $\boldsymbol{f}(S)$ has vanishing measure. In particular,

$$\int\limits_S h(\boldsymbol{x})\cdot|J_{\boldsymbol{f}}(\boldsymbol{x})|\cdot d\boldsymbol{x} = \int\limits_{\boldsymbol{f}(S)} H_S(\boldsymbol{y})\cdot d\boldsymbol{y};$$

that is the area formula holds for S.

It remains to prove that

$$\int\limits_{K\backslash S} h(\boldsymbol{x})\cdot|J_{\boldsymbol{f}}(\boldsymbol{x})|\cdot d\boldsymbol{x} = \int\limits_{\boldsymbol{f}(K\backslash S)} H_{K\backslash S}(\boldsymbol{y})\cdot d\boldsymbol{y}.$$

Since $J_{\boldsymbol{f}}(\boldsymbol{x}) \neq 0$ for any $\boldsymbol{x} \in K\backslash S$, by inverse function theorem, the restriction of $\boldsymbol{f}$ to a neighborhood $U \ni \boldsymbol{x}$ has a smooth inverse. Therefore for any compact set $K' \subset U$ we have that

$$\int\limits_{K'} h(\boldsymbol{x})\cdot|J_{\boldsymbol{f}}(\boldsymbol{x})|\cdot d\boldsymbol{x} = \int\limits_{\boldsymbol{f}(K_1)} h(\boldsymbol{f}^{-}1(\boldsymbol{y}))\cdot d\boldsymbol{y}.$$

It remains to subdivide K_1 into a countable collection of subsets of that type and sum up the corresponding formulas. $\square$

A.5 Initial value problem

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

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

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

The array functions $(f_1, \dots, f_n)$ can be considered as one vector-valued function $\boldsymbol{f}\colon \mathbb{R}^n \times \mathbb{I} \to \mathbb{R}^n$ and the array $(x_1, \dots, x_n)$ can be considered as a vector $\boldsymbol{x} \in \mathbb{R}^n$. Therefore the system can be rewritten as one vector equation

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

A.14. Theorem. *Suppose $\mathbb{I}$ is a real interval and $\boldsymbol{f}\colon \mathbb{R}^n \times \mathbb{I} \to \mathbb{R}^n$ is a smooth function. Then for any initial data $\boldsymbol{x}(t_0) = \boldsymbol{u}$ the differential equation*

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

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

(a) if $\mathbb{J} \ne \mathbb{I}$ (that is, if an end a of $\mathbb{J}$ lies in the interior of $\mathbb{I}$) then $\boldsymbol{x}(t)$ diverges for $t \to a$;

(b) the function $(\boldsymbol{u}, t_0, t) \mapsto \boldsymbol{x}(t)$ is smooth.

A.6 Lipschitz condition

Recall that a function f is called Lipschitz if there is a constant L such that

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

for values x and y in the domain of definition of f. This definition works for maps between metric spaces, but we will use it for real-to-real functions only.

A.15. Rademacher's theorem. *Let $f\colon [a, b] \to \mathbb{R}$ be a Lipschitz function. then derivative $f'(x)$ is defined for alomst all $x \in [a, b]$. Moreover the derivative 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*

$$f(b) - f(a) = \int_a^b f'(x)\cdot dx,$$

holds if the integral understood in the sense of Lebesgue.

It is often helps to work with measurable functions; it makes possible to extend many statements about continuous function to measurable functions.

A.16. 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, φ is bounded above and/or below by some constants, then we can assume that so is ψ_ε.*

A.7 Uniform continuity

Let f be a real function defined on a real interval. If for any $\varepsilon > 0$ there is $\delta > 0$ such that

$$|x_1 - x_2|_X < \delta \quad \Longrightarrow \quad |f(x_1) - f(x_2)|_Y < \varepsilon,$$

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. Indeed, in this case

❶ $$|f(x_1) - f(x_2)| = |x_1 + x_2|\cdot|x_1 - x_2|$$

If $|x_1 - x_2|$ is arbitrarily small positive value, then one is free to chose $|x_1 + x_2|$ sufficiently large, so that the product in ❶ is bigger that any given number. However if f is continuous and defined on a closed interval $[a,b]$, then f is uniformly continuous

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

$$|x_1 - x_2|_X < \delta \quad \Longrightarrow \quad |f_n(x_1) - f_n(x_2)|_Y < \varepsilon$$

for any n. The following lemma is a partial case of [45, Theorem 7.25].

A.17. Lemma. *Any uniformly equicontinuous sequence of function $f_n\colon [a,b] \to [c,d]$ has a subsequence that converges pointwise to a continuous function.*

A.8 Jordan's theorem

We sometimes characterization of homeomorphism.

A.18. Theorem. *A continuous bijection f between compact metric spaces has continuous inverse; that is, f is a homeomorphism.*

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

A.19. Theorem. *The complement of any closed simple plane curve γ has exactly two connected components.*

Moreover the 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 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 somewhat developed technique is given by Patrick Doyle [46], among elementary proofs, one of my favorites is the proof given by Aleksei Filippov [47].

We use the following smooth analog of this theorem.

A.20. Theorem. *The complement of any closed simple smooth regular plane curve γ has exactly two connected components.*

Moreover the there is a diffeomorphism $h\colon \mathbb{R}^2 \to \mathbb{R}^2$ that maps the unit circle to γ.

The proof of this statement is much simpler. An amusing proof of can be found in [48].

A.9 Connectedness

Recall that a continuous map from the unit interval $[0,1]$ to a space is called a *path.*

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; that is, there is a path $\alpha\colon [0,1] \to X$ such that $\alpha(0) = x$ and $\alpha(1) = y$.

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

A.21. 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 containing x is called *connected component* of x in X.

A.10 Convexity

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 intersection of arbitrary family of convex sets is convex. The intersection of all convex sets containing X is called *convex hull* of X; it is the minimal convex set containing the given set X.

A.22. Lemma. *Let $K \subset \mathbb{R}^3$ be a closed convex subset. Then for any point $p \notin K$ there is a plane Π that separates K from p; that is, K and p lie on the opposite open half-spaces of Π.*

A function of two variables $(x, y) \mapsto f(x, y)$ is called convex if its epigraph $z \geqslant f(x, y)$ is a convex set. This is equivalent to the so-called *Jensen's inequality*

$$f\left(t{\cdot}x_1 + (1-t){\cdot}x_2\right) \le t{\cdot}f(x_1) + (1-t){\cdot}f(x_2)$$

for $t \in [0, 1]$. If f is smooth, then the condition is equivalent to the following inequality for second directional derivative:

$$D_w^2 f \geqslant 0$$

for any vector $w \neq 0$ in the (x, y)-plane.

A.11 Elementary geometry

A.23. Theorem. *The sum of sum of all the internal angles of a simple n-gon is $(n-2){\cdot}\pi$.*

Proof. The proof is by induction on n. For $n = 3$ it says that sum of internal angles of a triangle is π, which is assumed to be known.

First let us show that for any $n \geqslant 4$, any n-gon has a diagonal that lies inside of it. Assume this is holds true for all polygons with at most $n - 1$ vertex.

Fix an n-gon P, $n \geqslant 4$. Applying rotation if necessary, we can assume that all its vertexes have different x-coordinates. Let v be a vertex of P that minimizes the x-coordinate; denote by u and w its adjacent vertexes. Let us choose the diagonal uw if it lies in P. Otherwise the triangle $\triangle uvw$ contains another vertex of P. Choose a vertex s in the interior of $\triangle uvw$ that maximizes the distance to line uw. Note that the diagonal vs lyes in P; if it is not the case, then vs crosses another side pq of P, one of the vertexes p or q has larger distance to the line and it lies in the interior of $\triangle uvw$ — a contradiction.

Note that the diagonal divides P into two polygons, say Q and R, with smaller number of sides in each, say k and m correspondingly. Note that

❶ $$k + m = n + 2;$$

indeed each side of P appears once as a side of P or Q plus the diagonal appears twice — once as a side in Q and once as a side of R. Note that the sum of angles of P is the sum of angles of Q and R, which by the induction hypothesis are $(k-2)\cdot\pi$ and $(m-2)\cdot\pi$ correspondingly. It remains to note that ❶ implies

$$(k-2)\cdot\pi + (m-2)\cdot\pi = (n-2)\cdot\pi.$$ □

A.12 Triangle inequality for angles

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

A.24. Theorem. *The inequality*

$$\measuredangle aob + \measuredangle boc \geqslant \measuredangle aoc$$

holds for any three half-lines oa, ob and oc in the Euclidean space.

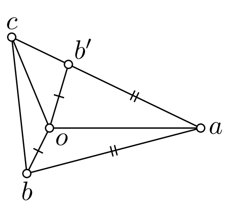
The following lemma says that angle of a triangle monotonically depends on the opposit side, assuming the we keep the remaining two sides fixed. It is a simple statement in elementary geometry; in particular it follows directly from the cosine rule.

A.25. 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*

$$\measuredangle xyz \geqslant \measuredangle x'y'z' \quad \textit{if and only if} \quad |x-z| \geqslant |x'-y'|.$$

Proof of A.24. We can assume that $\measuredangle aob < \measuredangle aoc$; otherwise the statement is evident. In this case there is a half-line ob' in the angle aoc such that

$$\measuredangle aob = \measuredangle aob',$$



so in particular we have that

$$\measuredangle aob' + \measuredangle b'oc = \measuredangle aoc.$$

Without loss of generality we can assume that $|o-b| = |o-b'|$ and b' lies on a line segment ac, so

$$|a-b'| + |b'-c| = |a-c|.$$

Then by triangle inequality

❶
$$\begin{aligned} |a-b| + |b-c| \geqslant |a-c| &= \\ &= |a-b'| + |b'-c|. \end{aligned}$$

Note that in the triangles aob and aob' the side ao is shared, $\measuredangle aob = \measuredangle aob'$ and $|o-b| = |o-b'|$. By side-angle-side congruence condition, we have that $\triangle aob \cong \triangle aob'$; in particular $|a-b'| = |a-b|$. Therefore from ❶ we have that

$$|b-c| \geqslant |b'-c|.$$

Applying the angle monotonicity (A.25) we get that

$$\measuredangle boc \geqslant \measuredangle b'oc.$$

Whence

$$\begin{aligned} \measuredangle aob + \measuredangle boc \geqslant \measuredangle aob' + \measuredangle b'oc &= \\ &= \measuredangle aoc. \end{aligned}$$

□

Appendix B

Semisolutions

Exercise 1.2. The image of γ might have a shape of digit 8 or 9.

Exercise 2.3. 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.

Advanced exercise 2.12. Use theorems of Rademacher and Lusin (A.15 and A.16).

Exercise 3.2. Differentiate the identity $\langle\gamma(s),\gamma(s)\rangle = 1$ a couple of times.

Exercise 3.3. Prove and use the following identities:

$$\begin{gathered}\gamma''(t)-\gamma''(t)^{\perp}=\tfrac{\gamma'(t)}{|\gamma'(t)|}\cdot\langle\gamma''(t),\tfrac{\gamma'(t)}{|\gamma'(t)|}\rangle,\\ |\gamma'(t)|=\sqrt{\langle\gamma'(t),\gamma'(t)\rangle}.\end{gathered}$$

Advanced exercise 3.9. Assume that γ is unit-speed; show that $|\sigma'|\leqslant\kappa+\theta'$, where $\theta(s)=\measuredangle(\gamma(s),\gamma'(s))$.

Exercise 3.13. Use that exterior angle of a triangle equals to the sum of the two remote interior angles; for the second part apply the induction on number of vertexes.

Exercise 3.19. Choose a value $s_0\in[a,b]$ that splits the total curvature into two equal parts, θ in each. Observe that $\measuredangle(\gamma'(s_0),\gamma'(s))\leqslant\theta$ for any s. Use this inequality the same way as in the proof of the bow lemma.

Exercise 3.25. Modify the proof of semi-continuity of length (2.13).

Exercise ??. Show and use that the binormal vector is constant.

Exercise ??. Show that $\langle w, \alpha\rangle$ is constant if γ makes constant angle with a fixed vector w and α is the evolvent of γ.

Exercise ??. Use the second statement in **??**.

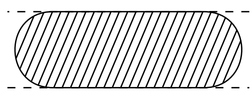
Advanced exercise ??. Note that the function

$$\rho(\ell) = |\gamma(t+\ell) - \gamma(t)|^2 = \langle \gamma(t+\ell) - \gamma(t), \gamma(t+\ell) - \gamma(t)\rangle$$

is smooth and does not depend on t. Express speed, curvature and torsion of γ in terms of derivatives $\rho^{(n)}(0)$ and apply **??**.

Exercise 5.2. Apply the spiral lemma (4.10).

Exercise 5.6. Note that γ lies in a figure F as on the diagram. More precisely, F is formed by a rectangle with pair of bases on the lines and two half discs attached to the sides of length 2.



Look at the right most or left most position of F that still contains curve loop.

To do the second part, try to modify the suggested proof.

Exercise 5.11. Note that we can assume that γ bounds a convex figure F, otherwise by 5.9 its curvature changes the 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 common chord xy that lies in F and apply supporting test (5.3).

Exercise 5.15. Note that γ contains a simple loop; apply to it **??**.

Exercise 5.15. Note that γ contains a simple loop; apply to it **??**.

Exercise 5.18. Repeat the proof of theorem for each cyclic concatenation of an arc of γ from p_i to p_{i+1} with large arc of the circle.

Exercise 4.9. Use the definition of osculating circle via order of contact and that inversion maps circles to circlines.

Exercise 6.15. Show that $\nu = \frac{\nabla h}{|\nabla h|}$ defines a unit normal field on Σ.

Exercise 7.13. Use 7.1 and 5.13.

Exercise 8.4. Consider the minimal sphere that encloses the surface.

Exercise 8.10. Look for a supporting spherical dome with the unit circle as the boundary.

Exercise 7.12. Use 7.1 and **??**.

Exercise 8.6. Assume a maximal ball in V touches its boundary at the points p and q. Consider the projection of V to a plane thru p, q and the center of the ball.

Exercise 7.15. Drill an extra hole or combine two examples together.

Exercise 8.17. Use 7.12.

Exercise 8.18. Prove and use that each point $p \in \Sigma$ has a direction with vanishing normal curvature.

Exercise 8.22. Use the 8.21 and the hemisphere lemma (2.16).

Exercise 8.25. Observe that it is sufficient to construct a smooth parametrization of Δ_ε by a closed hemisphere. To do this repeat the argument in **??** with the center at a point surrounded by the boundary line of Δ_ε in its plane.

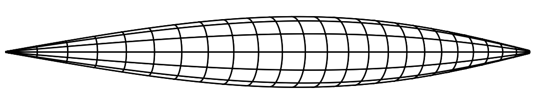
Exercise 8.28. Look for an example among the surfaces of revolution and use 7.12.

Exercise 8.29. 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 apply Meusnier's theorem (**??**).

Exercise 8.5. Show and use that any tangent plane T_p supports Σ at p.

Exercise 8.11. Note that we can assume that the surface has positive Gauss curvature, otherwise the statement is evident. Therefore the surface bounds a convex region that contains a line segment of length π.

Observe 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 (Use 3.4 and **??**). Try to support the surface Σ from inside by a surface of revolution of the described type with large R.



Exercise 9.1. Cut the lateral surface of the mountain by a line from the cowboy to the top, unfold it on the plane and try to figure out what is the image of the strained lasso.

Advanced exercise 9.16. Show that the concatenation of the line segment $[p_t, \gamma(t)]$ and the arc $\gamma|_{[t,\ell]}$ is a shortest path in the closed region W outside of Σ.

Exercise 9.21. Use 9.19 and 3.15.

The suggested argument does not give the optimal bound for the Lipschitz constant that guarantees that γ is simple, but later (see 11.12) we will show that the exact bound is $\sqrt{3} = \operatorname{tg}\frac{\pi}{3}$ — it is the same as in the exercise about mountain of with the shape of a perfect cone; see 9.1.

Exercise 7.6. Denote by $\nu_1(t)$ and $\nu_2(t)$ the unit normal vectors to Σ_1 and Σ_2 at $\gamma(t)$. Note that $\langle\nu_1(t),\nu_2(t)\rangle$ is constant; take it derivative and apply 10.3.

Exercise 10.7. Use 11.21.

Exercise 11.2. Denote by w the pole of the equator; show and use that w is a parallel vector field along γ.

Exercise 11.11. Estimate integral of Gauss curvature bounded by a simple geodesic loop and apply 10.6.

Exercise 11.12. Note that it is sufficient to show that the surface has no geodesic loops. Estimate the integral of Gauss curvature of whole surface and a disc in it surrounded by a geodesic loop.

Exercise 9.6. Use 9.3.

Exercise 12.2. Note that in order to show that $\gamma_t''(s) \perp \mathrm{T}_{\gamma_t(s)}$, it is sufficient to show that $\langle\frac{\partial^2}{\partial s^2}w, \frac{\partial}{\partial t}w\rangle = 0$.

Bibliography

[1] E. Denne. *Alternating quadrisecants of knots*. Thesis (Ph.D.)–University of Illinois at Urbana-Champaign. ProQuest LLC, Ann Arbor, MI, 2004, p. 119.

[2] G. D. Chakerian. "An inequality for closed space curves". *Pacific J. Math.* 12 (1962), pp. 53–57.

[3] István Fáry. "Sur certaines inégalites géométriques". *Acta Sci. Math. Szeged* 12.Leopoldo Fejér et Frederico Riesz LXX annos natis dedicatus, Pars A (1950), pp. 117–124.

[4] G. D. Chakerian. "An inequality for closed space curves". *Pacific J. Math.* 12 (1962), pp. 53–57.

[5] Serge Tabachnikov. "The tale of a geometric inequality". *MASS selecta.* Amer. Math. Soc., Providence, RI, 2003, pp. 257–262.

[6] John M. Sullivan. "Curves of finite total curvature". *Discrete differential geometry.* Vol. 38. Oberwolfach Semin. 2008, pp. 137–161.

[7] J. Lagarias and T. Richardson. "Convexity and the average curvature of plane curves". *Geom. Dedicata* 67.1 (1997), pp. 1–30.

[8] A. I. Nazarov and F. V. Petrov. "On a conjecture of S. L. Tabachnikov". *Algebra i Analiz* 19.1 (2007), pp. 177–193.

[9] R. Foote, M. Levi, and S. Tabachnikov. "Tractrices, bicycle tire tracks, hatchet planimeters, and a 100-year-old conjecture". *Amer. Math. Monthly* 120.3 (2013), pp. 199–216.

[10] Heinz Hopf. "Über die Drehung der Tangenten und Sehnen ebener Kurven". *Compositio Math.* 2 (1935), pp. 50–62.

[11] Heinz Hopf. *Differential geometry in the large.* Second. Vol. 1000. Lecture Notes in Mathematics. Notes taken by Peter Lax and John W. Gray, With a preface by S. S. Chern, With a preface by K. Voss. Springer-Verlag, Berlin, 1989, pp. viii+184.

[12] P. G. Tait. "Note on the circles of curvature of a plane curve." *Proc. Edinb. Math. Soc.* 14 (1896), p. 26.

[13] 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.

[14] D. Fuchs and S. Tabachnikov. *Mathematical omnibus.* Thirty lectures on classic mathematics. American Mathematical Society, Providence, RI, 2007.

[15] G. Pestov and V. Ionin. "On the largest possible circle imbedded in a given closed curve". *Dokl. Akad. Nauk SSSR* 127 (1959), pp. 1170–1172.

[16] W. Blaschke. *Kreis und Kugel.* Verlag von Veit & Comp., Leipzig, 1916.

[17] S. Mukhopadhyaya. "New methods in the geometry of a plane arc". *Bull Calcutta Math Soc* 1 (1909), pp. 31–37.

[18] R. Osserman. "The four-or-more vertex theorem". *Amer. Math. Monthly* 92.5 (1985), pp. 332–337.

[19] J. B. Meusnier. “Mémoire sur la courbure des surfaces”. *Mem des savan etrangers* 10.1776 (1785), pp. 477–510.

[20] V. N. Lagunov. “On the largest sphere contained in a closed surface”. *Sibirsk. Mat. Ž.* 1 (1960), pp. 205–232.

[21] S. N. Bernstein. “Sur un théoréme de géométrie et son application aux équations aux dérivées partielles du type elliptique”. *Comm. de la Soc. Math de Kharkov (2eme ser.), 15 (1915–17), 38–45.* 26 (1927). See also: Über ein geometrisches Theorem und seine Anwendung auf die partiellen Differential gleichungen vom elliptischen Typus, Math. Zeit, 26, pp. 551–558.

[22] ().

[23] S. Alexander, V. Kapovitch, and A. Petrunin. *An invitation to Alexandrov geometry: CAT (0) spaces.*

[24] J. J. Stoker. “Über die Gestalt der positiv gekrümmten offenen Flächen im dreidimensionalen Raume”. *Compositio Mathematica* 3 (1936), pp. 55–88.

[25] J. Hadamard. “Sur certaines propriétés des trajectoires en dynamique”. *J. math. pures appl.* 3 (1897), pp. 331–387.

[26] J. Liberman. “Geodesic lines on convex surfaces”. *C. R. (Doklady) Acad. Sci. URSS (N.S.)* 32 (1941), pp. 310–313.

[27] V. V. Usov. “The length of the spherical image of a geodesic on a convex surface”. *Sibirsk. Mat. Ž.* 17.1 (1976), pp. 233–236.

[28] I. D. Berg. “An estimate on the total curvature of a geodesic in Euclidean 3-space-with-boundary.” *Geom. Dedicata* 13 (1982), pp. 1–6.

[29] N. Lebedeva and A. Petrunin. “On the total curvature of minimizing geodesics on convex surfaces”. *Algebra i Analiz* 29.1 (2017). Reprinted in St. Petersburg Math. J. **2**9 (2018), no. 1, 139–153, pp. 189–208.

[30] M. Levi. “A “bicycle wheel” proof of the Gauss–Bonnet theorem”. *Exposition. Math.* 12.2 (1994), pp. 145–164.

[31] K. F. Gauss. “Disquisitiones generales circa superficies curvas”. *Commentationes Societatis Regiae Scientiarum Gottingensis recentiores* 6 (classis mathematicae) (1828). [Translated to English in *General investigations of curved surfaces* 1902.], pp. 99–146.

[32] A. Petrunin and A. Yashinski. “Piecewise isometric mappings”. *Algebra i Analiz* 27.1 (2015), pp. 218–247.

[33] S. Cohn-Vossen. „Totalkrümmung und geodätische Linien auf einfachzusammenhängenden offenen vollständigen Flächenstücken“. . . 1(43).2 (1936), S. 139–164.

[34] H. E. Rauch. “A contribution to differential geometry in the large”. *Ann. of Math. (2)* 54 (1951), pp. 38–55.

[35] 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), pp. 6–11.

[36] A. D. Alexandrov. *Vnutrennyaya Geometriya Vypuklyh Poverhnosteı.* OGIZ, Moscow-Leningrad,] 1948, p. 387.

[37] V. A. Toponogov. “On convexity of Riemannian spaces of positive curvature”. *Dokl. Akad. Nauk SSSR (N.S.)* 115 (1957), pp. 674–676.

[38] Yu. Burago, M. Gromov, and G. Perelman. “A. D. Aleksandrov spaces with curvatures bounded below”. *Uspekhi Mat. Nauk* 47.2(284) (1992), pp. 3–51, 222.

[39] 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), pp. 23–53.

[40] É. Cartan. *Leçons sur la Géométrie des Espaces de Riemann.* 1928.

[41] H. Busemann. “Spaces with non-positive curvature”. *Acta Math.* 80 (1948), pp. 259–310.

[42] W. Rinow. *Die innere Geometrie der metrischen Räume*. Die Grundlehren der mathematischen Wissenschaften, Bd. 105. Springer-Verlag, Berlin-Göttingen-Heidelberg, 1961, pp. xv+520.

[43] M. Gromov. "Hyperbolic groups". *Essays in group theory*. Vol. 8. Math. Sci. Res. Inst. Publ. Springer, New York, 1987, pp. 75–263.

[44] S. Alexander and R. Bishop. "The Hadamard–Cartan theorem in locally convex metric spaces". *Enseign. Math. (2)* 36.3-4 (1990), pp. 309–320.

[45] Walter Rudin. *Principles of mathematical analysis*. 1976.

[46] P. H. Doyle. "Plane separation". *Proc. Cambridge Philos. Soc.* 64 (1968), p. 291.

[47] A. F. Filippov. "An elementary proof of Jordan's theorem". *Uspehi Matem. Nauk (N.S.)* 5.5(39) (1950), pp. 173–176.

[48] Gregory R. Chambers and Yevgeny Liokumovich. "Converting homotopies to isotopies and dividing homotopies in half in an effective way". *Geom. Funct. Anal.* 24.4 (2014), pp. 1080–1100.