Introduction to Differential Geometry

Ivan Izmestiev

University of Fribourg, Spring 2016

Contents

Chapter 1

Curvature of curves

1 Curves, length, and area

1.1 What is a curve?

We will study curves in the plane or in the space. There are different ways to describe a curve.

- Graph of a function:

$$y = f(x) \quad \text{(plane curve)} \qquad \begin{cases} y = f(x) \\ z = g(x) \end{cases} \quad \text{(space curve)}$$

  Problem: not every curve is a graph of a function (think of a circle or of a spiral).

- Level set of a function:

$$F(x, y) = c \quad \text{(plane curve)}$$

  In $\mathbb{R}^3$ the level set of a function is (usually) a surface. A curve can be represented as an intersection of two surfaces:

$$\begin{cases} F_1(x, y) = c_1 \\ F_2(x, y) = c_2 \end{cases}$$

  This is good enough, but not convenient for computations (we will see this later).

- Parametrized curve: trajectory of a point. This is the most convenient representation, and we will use it most of the time.

Definition 1.1. *A* parametrized curve *is a map from an interval to* $\mathbb{R}^n$*:*

$$\gamma \colon I \to \mathbb{R}^n,$$

where $I \subset \mathbb{R}$ *is* (a, b) *or* $(-\infty, b)$ *or* $(-\infty, +\infty)$ *or* $[a, b]$ *etc.*

Example 1.2.

- A graph of a function can easily be parametrized:

$$y = f(x) \Leftrightarrow \begin{cases} x = t \\ y = f(t) \end{cases}$$

  That is, we put $\gamma(t) = (t, f(t))$.

- The circle $x^2 + y^2 = 1$ can be parametrized as

$$x = \cos t, \quad y = \sin t$$

  As the range of t we can take $[0, 2\pi)$ or $(-\infty, +\infty)$. The latter traces the circle infinitely many times.

Exercise 1.3. Find a parametrization of the curve $x^2 - y^2 = 1$ (this set consists of two components; parametrize one of them).

Note that the parametrization of a curve is not unique: different maps can have the same images. For example, the upper semicircle can be parametrized as

$$\gamma(t) = (t, \sqrt{1 - t^2}), \quad t \in [-1, 1]$$

The parabola $y = x^2$ can be parametrized as (t, t^2) or, for example, as (t^3, t^6).

Having decided for parametrized curves, let us think how good the map γ should be so that it image could be called a "curve". Of course, we want γ to be continuous (the point is moving, not jumping). But there are continuous maps that look rather strange...

- Space-filling curves, e. g. Peano curve and Hilbert curve. They map an interval onto a square (surjectively, but not injectively).
- The Koch snowflake. The boundary of the Koch snowflake is an injective image of a circle. However it has an infinite length.

Exercise 1.4. Compute the perimeter of the n-th iteration of the Koch snowflake. Compute the area of the Koch snowflake.

Remark 1.5. The boundary of the Koch snowflake has Hausdorff dimension $\frac{\log 4}{\log 3} > 1$, so this is not quite a curve. Objects of non-integer dimension are called fractals. Some beautiful pictures of fractals can be found in the book [MSW02].

Since we are doing *differential* geometry, we will require the map γ to be differentiable.

Definition 1.6. *A parametrized curve* $\gamma\colon I \to \mathbb{R}^n$ *is called* smooth *if it is a* C^∞*-map.*

By this we mean that each component of γ can be differentiated any number of times: if $\gamma(t) = (\gamma_1(t), \ldots, \gamma_n(t))$, then all of the derivatives $\frac{d\gamma_i}{dt}$, $\frac{d^2\gamma_i}{dt^2}$, $\frac{d^3\gamma_i}{dt^3}$, ... do exist.

The derivative of γ

$$\dot{\gamma} = \frac{d\gamma}{dt} = \left(\frac{d\gamma_1}{dt}, \ldots, \frac{d\gamma_n}{dt} \right)$$

is called the *velocity vector* of γ.

1.2 Length of a smooth curve

The velocity vector shows the direction and the speed of the motion. The distance traveled is the speed multiplied by the time. If the speed is variable, then the distance is the integral of the speed as a function of the time. Thus the common sense suggests us the following

Definition 1.7. *The integral*

$$L(\gamma) = \int_a^b \|\dot{\gamma}\|\, dt = \int_a^b \sqrt{\left(\frac{d\gamma_1}{dt}\right)^2 + \cdots + \left(\frac{d\gamma_n}{dt}\right)^2}\, dt$$

is called the length *of a smooth curve* $\gamma\colon [a,b] \to \mathbb{R}^n$.

There is a theorem behind this definition, because a standard way to define the length of a curve is as the supremum of lengths of inscribed polygons (think of how the length of a circle is defined):

$$\widetilde{L}(\gamma) := \sup_T \sum_{i=1}^{n} \|\gamma(t_i) - \gamma(t_{i-1})\|, \tag{1.1}$$

where T denotes a subdivision $a = t_0 < t_1 < \cdots < t_n = b$.

Theorem 1.8. *For every smooth curve* $\gamma\colon [a,b] \to \mathbb{R}^n$ *the supremum* (1.1) *exists and is equal to the integral in Definition 1.7:*

$$\widetilde{L}(\gamma) = L(\gamma).$$

We need a lemma.

Lemma 1.9. *For every smooth curve* $\gamma\colon [a,b] \to \mathbb{R}^n$ *its length is bigger or equal than the distance between the endpoints:*

$$\int_a^b \|\dot{\gamma}\|\, dt \geq \|\gamma(b) - \gamma(a)\|$$

Proof. Use the Cauchy-Schwarz inequality:

$$\|\gamma(b) - \gamma(a)\| \|\dot{\gamma}\| \geq \langle \gamma(b) - \gamma(a), \dot{\gamma} \rangle$$

Integration yields

$$\|\gamma(b)-\gamma(a)\| \int_a^b \|\dot{\gamma}\|\, dt \geq \left\langle \gamma(b) - \gamma(a), \int_a^b \dot{\gamma}\, dt \right\rangle = \langle \gamma(b)-\gamma(a), \gamma(b)-\gamma(a) = \|\gamma(b)-\gamma(a)\|^2$$

and the lemma follows. □

Proof of Theorem 1.8. For any subdivision T denote by L_T the length of the corresponding polygon and by $\gamma_{t_{i-1}}^{t_i}$ the arc of γ between t_{i-1} and t_i. By Lemma 1.9 we have

$$L(\gamma) = \sum_{i=1}^{n} L(\gamma_{t_{i-1}}^{t_i}) \geq \sum_{i=1}^{n} \|\gamma(t_i) - \gamma(t_{i-1})\| = L_T$$

As a consequence, $\widetilde{L}(\gamma) = \sup_T L_T$ exists, and $\widetilde{L}(\gamma) \leq L(\gamma)$.

Denote $S(t) = \widetilde{L}(\gamma_a^t)$ and compute the derivative of $S(t)$. It is easily seen that the generalized length is additive: $S(t+\varepsilon) - S(t) = \widetilde{L}(\gamma_t^{t+\varepsilon})$ for $\varepsilon \geq 0$. Hence we have

$$\left\| \frac{\gamma(t+\varepsilon) - \gamma(t)}{\varepsilon} \right\| \leq \frac{S(t+\varepsilon) - S(t)}{\varepsilon} \leq \frac{1}{\varepsilon} \int_t^{t+\varepsilon} \|\dot{\gamma}\| \, dt$$

(Check that this is also valid for $\varepsilon < 0$.) As ε tends to 0, both the left hand side and the right hand side tend to $\dot{\gamma}(t)$, and we conclude that the function $S(t)$ is differentiable with $\frac{dS}{dt} = \|\dot{\gamma}\|$. Since $S(a) = 0$, it follows that

$$\widetilde{L}(\gamma) = S(b) = \int_a^b \|\dot{\gamma}\| \, dt = L(\gamma).$$

□

Remark 1.10. Curves for which the supremum of lengths of inscribed polygons exist are called *rectifiable.* The Peano curve, the Hilbert curve, and the Koch curve are not rectifiable. It can be shown that every Lipschitz curve is rectifiable.

Definition 1.11. *Roll a circle along a line. The curve traced by a point on the circle is called a* cycloid*.*

Let the line be the x-axis, the circle have radius 1, at $t = 0$ the circle be tangent to the x-axis at the origin, and the point which we follow lie at the origin at this moment. If the circle rolls with a constant speed, then the corresponding cycloid has the parametrization

$$\gamma(t) = (t - \sin t, 1 - \cos t)$$

Exercise 1.12. Check the above equation. Compute the velocity vectors at $t = 0$ and near $t = 0$. Sketch the cycloid.

Exercise 1.13. Compute the length of the cycloid arc $t \in [0, 2\pi]$. When you ride a bicycle, what is the ratio between the distance traveled by you and the distance traveled by a point on a tire?

1.3 Area enclosed by a smooth planar curve

In this section we speak only about plane curves.

A curve $\gamma\colon [a,b] \to \mathbb{R}^2$ is called *closed*, if $\gamma(a) = \gamma(b)$. A closed curve is called *simple*, if it is not self-intersecting: $\gamma(t) = \gamma(t') \Rightarrow t = t'$ or $\{t, t'\} = \{a, b\}$.

Denote by $A(\mathrm{int}(\gamma))$ the area of the region bounded by a simple closed plane curve γ. Let us derive a formula for $A(\mathrm{int}(\gamma))$ heuristically. Consider a subdivision $a < t_1 < t_2 < \cdots < t_n = b$. The area of the polygon P_T with the vertices $\gamma(t_1), \ldots, \gamma(t_n)$ equals

$$\sum_{i=1}^{n} \mathrm{area}(O\gamma(t_i)\gamma(t_{i+1})) = \frac{1}{2}\sum_{i=1}^{n} O\gamma(t_i) \times O\gamma(t_{i+1}) = \frac{1}{2}\sum_{i=1}^{n}(x(t_i)y(t_{i+1}) - x(t_{i+1})y(t_i)$$

Here $v \times w$ denotes the planar cross-product, which is half the (signed) area of the triangle spanned by the vectors v and w. A different way to write the area of P_T is

$$\frac{1}{2}\sum_{i=1}^{n} \gamma(t_i) \times (\gamma(t_{i+1}) - \gamma(t_i))$$

Intuitively, as the subdivision becomes finer, the above sum should tend to the integral of $\gamma \times \dot\gamma$. Also the areas of inscribed polygons should tend to the area of the region inside γ. We don't prove either of these statements. Instead, we use a different method to prove the formula that we have just conjectured.

Theorem 1.14. *Let $\gamma\colon [a,b] \to \mathbb{R}^2$ be a simple closed curve. If γ is oriented counter-clockwise, then the area of the region bounded by γ equals*

$$\mathrm{area}(\mathrm{int}(\gamma)) = \frac{1}{2}\int_a^b (\gamma \times \dot\gamma)\, dt = \frac{1}{2}\int_a^b (x\dot y - \dot x y)\, dt \qquad (1.2)$$

If γ is oriented clockwise, then the above integral equals minus the area bounded by γ.

Proof. Recall the *Green theorem*:

$$\int_{\mathrm{int}(\gamma)} \left(\frac{\partial g}{\partial x} - \frac{\partial f}{\partial y}\right) dxdy = \int_\gamma f\, dx + g\, dy$$

Put $f(x,y) = -y$ and $g(x,y) = x$. Then the left hand side equals twice the area, and the right hand side equals

$$\int_\gamma (-y)\, dx + x\, dy = \int_a^b (-y\dot x + x\dot y)\, dt$$

□

Exercise 1.15. Show that if the curve γ is not closed, then formula (1.2) gives the (signed) area of the region bounded by γ and the segments $O\gamma(a)$ and $O\gamma(b)$.

What is the geometric meaning of formula (1.2) in the case when γ has self-intersections?

Exercise 1.16. Compute the area under an arc of the cycloid.

2 Curvature

2.1 Regular curves

Even a smooth curve doesn't always "look smooth". We see it on the example of a cycloid that has sharp points at $t = 2k\pi$ (a *cusp*). Another typical example is the semicubical parabola $y = x^{2/3}$ which has a smooth parametrization $\gamma(t) = (t^3, t^2)$ but exhibits a cusp at the point $(0, 0)$.

Exercise 2.1. Find a smooth parametrization of the curve $y = |x|$.

The problem with these points is that $\dot{\gamma}$ vanishes.

Definition 2.2. *A point $\gamma(t)$ on a parametrized curve γ is called a* regular point *if $\dot{\gamma}(t) \neq 0$; otherwise $\gamma(t)$ is called a* singular point*. A curve is called* regular *if all of its points are regular.*

There is a whole branch of mathematics that studies singularities of smooth maps; for an introduction see [Brö75]. We will focus on regular curves (or regular arcs of curves).

Exercise 2.3. The curve

$$\gamma(t) = (\cos^3 t, \sin^3 t)$$

is called *astroid*. Find singular and regular points of the astroid. Sketch the curve.

Exercise 2.4. Compute the length of the astroid and the area enclosed by it.

A parametrized curve possesses an oriented tangent at each of its regular points. The equation of the tangent at $\gamma(t_0)$ is

$$\ell(t) = \gamma(t_0) + t\dot{\gamma}(t_0)$$

(Note that at a cusp point there is also a tangent, but it has no orientation compatible with the time parameter on the curve.)

2.2 Arc-length parametrization

Definition 2.5. *Let $\gamma\colon I \to \mathbb{R}^n$ be a smooth curve. A* reparametrization *of γ is a curve*

$$\tilde{\gamma}\colon J \to \mathbb{R}^n, \quad \text{where} \quad \tilde{\gamma} = \gamma \circ \varphi$$

and $\varphi\colon J \to I$ is a diffeomorphism (that is, a differentiable bijective map with differentiable inverse).

Clearly, if $\tilde{\gamma}$ is a reparametrization of γ, then γ is a reparametrization of $\tilde{\gamma}$.

Example 2.6. Take $\gamma(t) = (\cos t, \sin t)$ and apply the reparametrization map

$$t = \varphi(s) = \frac{\pi}{2} - s$$

We obtain $\tilde{\gamma}(s) = \gamma(\varphi(s)) = (\sin s, \cos s)$.

Example 2.7. The curves $\gamma(t) = (t, t^2)$ and $\tilde{\gamma}(s) = (s^3, s^6)$ have the same image: the parabola $y = x^2$. However, they are not reparametrizations of each other, as the substitution $t = \varphi(s) = s^3$ is not a diffeomorphism. Indeed, the inverse map $\varphi^{-1}(t) = \sqrt[3]{s}$ has no derivative at 0.

By the inverse function theorem

$$\left.\frac{d\varphi^{-1}}{dt}\right|_{t=\varphi(s)} = \frac{1}{\frac{d\varphi}{ds}}$$

the derivative of a diffeomorphism $\varphi\colon J \to I$ does not vanish. Conversely, every C^1-bijection $J \to I$ with non-zero derivative is a C^1-diffeomorphism. (With some effort, one can also derive an analog of the inverse function theorem for C^∞-functions.)

Lemma 2.8. *A reparametrization of a regular curve is a regular curve.*

Proof. Let $\tilde{\gamma}(s) = \gamma(\varphi(s))$ be a reparametrized curve. We have

$$\frac{d\tilde{\gamma}}{ds} = \left.\frac{d\gamma}{dt}\right|_{t=\varphi(s)} \frac{d\varphi}{ds} = \dot{\gamma}(\varphi(s))\frac{d\varphi}{ds}$$

Since φ is a diffeomorphism, $\frac{d\varphi}{ds} \neq 0$, so that $\dot{\gamma} \neq 0 \Rightarrow \frac{d\tilde{\gamma}}{ds} \neq 0$. □

Definition 2.9. *A curve $\gamma\colon I \to \mathbb{R}^n$ is called a* unit-speed curve*, if $\|\dot{\gamma}(t)\| = 1$ for all $t \in I$.*

Theorem 2.10. *Every regular curve has a unit-speed reparametrization.*

Proof. Let $\gamma\colon I \to \mathbb{R}^n$ be a regular curve. Fix a point $a \in I$ and consider the function

$$\psi\colon I \to \mathbb{R}, \quad \psi(t) = \int_a^t \|\dot{\gamma}\|\, dt$$

Then ψ is continuously differentiable: $\frac{d\psi}{dt} = \|\dot{\gamma}(t)\|$ and strictly monotone, because $\dot{\gamma}(t) \neq 0$ by the regularity assumption. It follows that ψ is injective, so that if we put $J = \psi(I)$ then there is an inverse function $\varphi\colon J \to I$. The map φ is a diffeomorphism by the inverse function theorem. Besides, $\frac{d\varphi}{ds} = \frac{1}{\dot{\gamma}(\varphi(s))}$.

It remains to check that $\tilde{\gamma} = \gamma \circ \varphi$ is a unit-speed curve:

$$\frac{d\tilde{\gamma}}{ds} = \dot{\gamma}(\varphi(s))\frac{d\varphi}{ds} = 1$$

□

Unit-speed curves are also called *arc-length parametrized* or *naturally parametrized.*

2.3 Curvature and osculating circles

Definition 2.11. *Let γ be a unit-speed curve. Then the norm of its acceleration vector is called the curvature of γ:*

$$\kappa(t) := \|\ddot{\gamma}(t)\|$$

The curvature is defined for any regular curve: just reparametrize the curve to the unit speed and take the norm of the second derivative with respect to the arc-length parameter. (Unit-speed parametrization is unique up to parameter changes of the sort $\pm t + a$, which don't change the norm of the second derivative; therefore the curvature of an arbitrary regular curve is well-defined.) With some effort, one can even derive the following formula for the curvature of a regular (non-constant speed) curve in $\mathbb{R}^2$ or $\mathbb{R}^3$.

Theorem 2.12. *The curvature of a regular curve in $\mathbb{R}^2$ or $\mathbb{R}^3$ equals*

$$\kappa(t) = \frac{\|\dot{\gamma} \times \ddot{\gamma}\|}{\|\dot{\gamma}\|^3}$$

For a proof, see [Pre10].

Theorem 2.13. *The acceleration vector of a unit-speed curve is orthogonal to the curve:*

$$\langle \dot{\gamma}, \ddot{\gamma} \rangle = 0$$

Proof. Differentiate with respect to t the equation $\|\dot{\gamma}\|^2 = 1$:

$$0 = \frac{d}{dt}\langle \dot{\gamma}, \dot{\gamma} \rangle = 2\langle \dot{\gamma}, \ddot{\gamma} \rangle$$

□

Definition 2.14. *Let γ be a unit-speed curve, and let $\kappa(t) \neq 0$ for some t. The unit vector in the direction of the acceleration vector $\ddot{\gamma}(t)$ is called the* principal normal *to γ at t. We denote it by $\nu(t)$.*

Thus for the unit-speed curves the following equation holds:

$$\ddot{\gamma}(t) = \kappa(t)\, \nu(t) \tag{1.3}$$

It encodes Definition 2.11 and Theorem 2.13 at the same time.

Example 2.15. The curvature of a circle of radius R equals $\frac{1}{R}$. Indeed, a unit-speed parametrization is $\gamma(t) = (R\cos\frac{t}{R}, R\sin\frac{t}{R})$, so that we have

$$\dot{\gamma}(t) = (-\sin\frac{t}{R}, \cos\frac{t}{R}), \quad \ddot{\gamma}(t) = (-\frac{1}{R}\cos\frac{t}{R}, -\frac{1}{R}\sin\frac{t}{R})$$

Hence $\kappa(t) = \|\ddot{\gamma}(t)\| = \frac{1}{R}$.

Definition 2.16. *Let $\gamma(t)$ be a point of non-zero curvature. The number $\frac{1}{\kappa(t)}$ is called the* radius of curvature*. The point lying on the line through $\gamma(t)$ perpendicular to γ at the distance $\frac{1}{\kappa(t)}$ from it in the direction of the principal normal is called the* center of curvature*. The circle of radius $\frac{1}{\kappa(t)}$ centered at the center of curvature is called the* osculating circle *of γ at $\gamma(t)$.*

By definition, the osculating circle is tangent to the curve and has the same curvature as γ at the point of tangency. Similarly as the tangent line is the best approximating line, the osculating circle is the best approximating circle, as the following exercise shows.

Exercise 2.17. Let $\alpha\colon I \to \mathbb{R}^n$ and $\beta\colon I \to \mathbb{R}^n$ be two unit-speed curves. Assume that at some $t \in I$ the curves go through the same point, have their the same tangent, the same curvature, and the same principal normal (if the curvature is non-zero). Show that

$$\|\alpha(t+\varepsilon) - \beta(t+\varepsilon)\| = o(\varepsilon^2)$$

(Hint: use the Taylor expansion.)

Osculating circles in the plane have the following nice geometric property.

Theorem 2.18. *Let $\gamma\colon [a,b] \to \mathbb{R}^2$ be a regular curve with a monotone increasing curvature, and let O_t be its osculating circle for some $t \in (a,b)$. Then the arc γ_a^t lies outside O_t, and the arc γ_t^b lies inside O_t.*

Even stronger, the osculating circles are nested: $O_t \subset O_{t'}$ for $t > t'$.

See [GTT13].

Chapter 2

Plane curves

1 Evolutes and wave fronts

1.1 Signed curvature

For plane curves, the curvature can be equipped with a sign: curves "turning left" have a positive curvature, those "turning right" a negative curvature.

Let $\tau = \frac{\dot\gamma}{\|\dot\gamma\|}$ denote the unit tangent vector.

Definition 1.1. *Let γ be a plane curve. The* signed unit normal ν_s *is obtained from τ by rotating it by the angle $\frac{\pi}{2}$ in the positive direction.*

Note that the signed unit normal is defined at every point of a regular curve, while the principal normal only at the points of non-vanishing curvature.

Definition 1.2. *The* signed curvature *of a plane curve γ is defined by*

$$\ddot\gamma = \kappa_s \nu_s$$

Thus we have

$$\kappa_s = \begin{cases} \kappa, & \text{if the curve turns left} \\ -\kappa, & \text{if the curve turns right} \end{cases}$$

The signed curvature can be computed for any (not necessarily unit-speed) parametrization by the formula

$$\kappa_s = \frac{\dot\gamma \times \ddot\gamma}{\|\dot\gamma\|^3},$$

where $v \times w$ is simply the determinant of the matrix with columns v and w.

Exercise 1.3. Derive a formula for the signed curvature of the graph of a function.

The curvature of a plane curve can be given a simple geometric meaning. Let $\varphi(t) \in [0, 2\pi)$ be the angle from the first standard basis vector e_1 to the tangent vector $\tau(t)$.

Theorem 1.4. *The signed curvature of a unit-speed plane curve is the rotation speed of its tangent vector:*

$$\kappa_s = \frac{d}{dt}\varphi$$

Proof. By definition of φ, and since (τ, ν_s) is a positive othonormal basis, we have

$$\tau = e_1 \cos\varphi + e_2 \sin\varphi, \quad \nu_s = -e_1 \sin\varphi + e_2 \cos\varphi$$

By definition of the signed curvature,

$$\kappa_s \nu_s = \dot\tau = (-e_1 \sin\varphi + e_2 \cos\varphi)\frac{d\varphi}{dt},$$

and the theorem follows. □

Strictly speaking, the above theorem may fail at the points where $\varphi(t) = 0$, since φ may jump from 0 to 2π there. Later we will modify the definition of φ to avoid this.

Theorem 1.5. *Let $k\colon I \to \mathbb{R}$ be any smooth function. Then there is a unit-speed curve $\gamma\colon I \to \mathbb{R}$ whose signed curvature is k. Besides, the curve γ is uniquely determined by initial conditions $\gamma(t_0)$, $\dot\gamma(t_0)$ at any point $t_0 \in I$.*

Proof. Put

$$\varphi(t) = \varphi_0 + \int_{t_0}^{t} k(t)\,dt \quad \text{(modulo } 2\pi\text{)},$$

where φ_0 is the angle from e_1 to $\dot\gamma(t_0)$. Then put $\tau(t) = e_1 \cos\varphi(t) + e_2 \sin\varphi(t)$ and

$$\gamma(t) = \gamma(t_0) + \int_{t_0}^{t} \tau(t)\,dt$$

Clearly, γ is a unit-speed curve with $\dot\gamma = \tau$; by Theorem 1.4 it has the signed curvature k.

On the other hand, any curve with the curvature $k(t)$ must have the velocity vector $\tau(t)$ as above. Hence its position vector is as above. □

1.2 Evolutes and involutes

We will need the following statement. The method of proof will also be important later.

Theorem 1.6. *Let γ be a regular plane curve. Then the derivative of its signed unit normal is*

$$\dot\nu_s(t) = -\kappa_s(t)\dot\gamma(t)$$

Proof. Since the vectors $\dot\gamma(t)$ and $\nu_s(t)$ form an orthonormal basis, we have

$$\dot\nu_s(t) = a\dot\gamma(t) + b\nu_s(t), \quad a = \langle \dot\nu_s(t), \dot\gamma(t)\rangle, b = \langle \dot\nu_s(t), \nu_s(t)\rangle$$

Exactly as in the proof of Theorem 2.13, by differentiating $\|\nu_s(t)\|^2 = 1$ we obtain $b = 0$. To compute a, differentiate $\langle \nu_s(t), \dot\gamma(t)\rangle = 0$ to obtain

$$a = \langle \dot\nu_s(t), \dot\gamma(t)\rangle = -\langle \nu_s(t), \ddot\gamma(t)\rangle = -\kappa_s(t)$$

□

By the same argument (or by using that $\kappa_s = \kappa \Leftrightarrow \nu_s = -\nu$), we have $\dot\nu = -\kappa\dot\gamma$ at the points with non-vanishing curvature (at the points with $\kappa = 0$ the function κ has a discontinuity).

Theorem 1.6 also has a simple geometric proof. As $(\dot\gamma(t), \nu_s(t))$ form an orthonormal basis, $\nu(t)$ rotates with the same angular velocity as $\dot\gamma$, and this velocity is by definition $\kappa(t)$. The instantaneous change of $\nu(t)$ is orthogonal to $\nu(t)$, thus parallel to $\dot\gamma(t)$...

Definition 1.7. *Let γ be a regular plane curve with everywhere non-vanishing curvature. The curve traced by the centers of curvature of γ is called the* evolute *of γ.*

Theorem 1.8. *The evolute of γ is regular at t if and only if the curvature of γ is non-stationary at t.*

Proof. The evolute e of γ has the parametrization

$$e(t) = \gamma(t) + \frac{1}{\kappa(t)}\nu(t).$$

(The result does not change if we replace κ and ν with κ_s and ν_s!) Therefore

$$\dot e = \dot\gamma + \frac{1}{\kappa}\dot\nu + \frac{d}{dt}\left(\frac{1}{\kappa}\right) = \frac{d}{dt}\left(\frac{1}{\kappa}\right),$$

where we used Theorem 1.6. Hence $\dot e = 0 \Leftrightarrow \dot\kappa = 0$. □

The following are two nice geometric properties of the evolute.

Theorem 1.9. *The evolute is the envelope of the normals to the curve. That is, every line that intersects the curve orthogonally is tangent to the evolute.*

If e is the evolute of γ, then γ is called an *involute* of e.

Theorem 1.10. *Unwind a string stretched along a regular curve. Then its free end traces an involute of the curve. Every involute can be constructed in this way.*

Note that while every curve has only one evolute, an involute is not unique: it depends on the length of the string.

Exercise 1.11. Prove Theorems 1.9 and 1.10.

Exercise 1.12. Compute and sketch the evolutes of the

- parabola $y = x^2$;
- cycloid $(t - \sin t, 1 - \cos t)$;
- astroid $(\cos^3 t, \sin^3 t)$.

On the history of evolutes and wavefronts: [Arn90, Chapter 3].

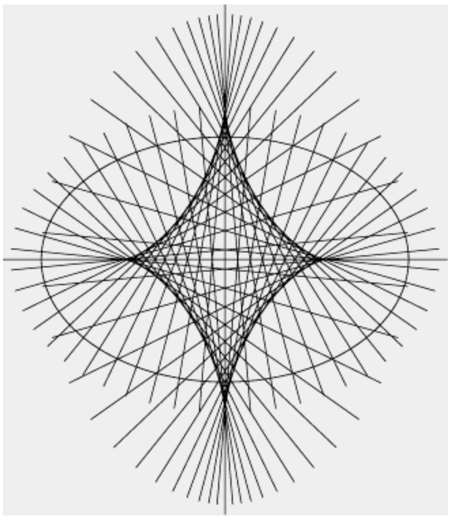

Figure 2.1: The evolute of an ellipse.

1.3 Parallel curves and wave fronts

Definition 1.13. *Let $\gamma\colon I \to \mathbb{R}^2$ be a regular curve, and $\nu_s\colon I \to \mathbb{R}^2$ be its signed unit normal. Choose $\delta \in \mathbb{R}$ The curve*

$$\gamma_\delta = \gamma + \delta\nu_s$$

is called a parallel curve *to γ at distance δ.*

Theorem 1.14. *Singular points of a curve parallel to γ are centers of curvature of γ.*

Proof. We have

$$\dot\gamma_\delta = \dot\gamma + \delta\dot\nu_s = (1 - \delta\kappa_s)\dot\gamma$$

(In particular, the velocity vector of γ_δ at t is parallel to the velocity vector of γ at t. This also implies that being parallel is a symmetric and transitive relation.) Thus γ_δ is singular at t if and only if $\delta = \frac{1}{\kappa_s(t)}$. The corresponding point of γ_δ is

$$\gamma_\delta(t) = \gamma(t) + \frac{1}{\kappa_s(t)} \cdot \nu_s(t),$$

which is the center of curvature of γ at t. □

The family of curves parallel to γ is called a *wave front.* Theorem 1.14 implies the following.

Corollary 1.15. *The evolute of a curve is the set of singularities of a corresponding wave front.*

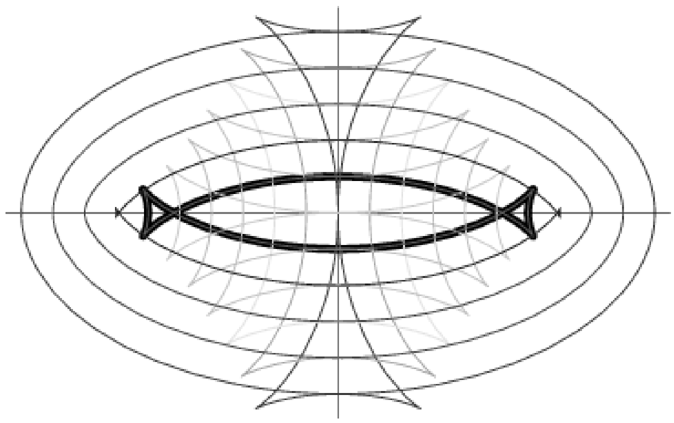

Figure 2.2: The wave front of an ellipse.

See Figure 2.2.

Corollary 1.16. *Parallel curves have the same evolutes.*

Exercise 1.17. Show that the signed curvature of a parallel curve at non-singular points is equal to

$$\frac{\kappa_s}{1 - \delta\kappa_s}$$

1.4 Distance functions

Let $\gamma\colon [a,b] \to \mathbb{R}^2$ be a simple regular curve. For $\varepsilon > 0$ small enough, all parallel curves γ_δ with $|\delta| < \varepsilon$ are also simple and pairwise disjoint:

$$\gamma_{\delta_1}(t_1) = \gamma_{\delta_2}(t_2) \Leftrightarrow \delta_1 = \delta_2, t_1 = t_2$$

(we leave this fact without proof). The union of the images of all these curves is a *tubular neighborhood* of $\gamma(I)$. Notation: $N_\varepsilon(\gamma)$.

A necessary (but not sufficient) condition for the images of γ_δ to be disjoint is $\varepsilon < \frac{1}{\kappa(t)}$ for all t.

Definition 1.18. *Let $p = \gamma_\delta(t) \in N_\varepsilon(\gamma)$. Then (t,δ) are called the* normal coordinates *of p.*

The coordinate lines $t = \text{const}$ are straight line intervals intersecting γ orthogonally; the coordinate lines $\delta = \text{const}$ are curves parallel to γ.

Theorem 1.19. *If $\varepsilon > 0$ is small enough, then the normal coordinate map $\delta\colon N_\varepsilon(\gamma) \to \mathbb{R}$ associates to every point its distance to $\gamma(I)$:*

$$p = \gamma_\delta(t) \in N_\varepsilon(\gamma) \Leftrightarrow \delta = \min_{x\in\gamma(I)} \|p - x\|$$

The gradient of the map δ *has norm* 1 *everywhere:*

$$\|\nabla\delta\| = 1$$

Proof. If $p = \gamma_\delta(t) = \gamma(t) + \delta\nu_s(t)$, then $\|p - \gamma(t)\| = \delta$, thus $\delta(p) \geq \operatorname{dist}(p, \gamma)$. On the other hand, the circle of radius δ centered at p has with $\gamma(I)$ only the point $\gamma(t)$ in common. (Locally this follows from the fact that $\kappa < \frac{1}{|\delta|}$ everywhere, and is related to the behaviour of osculating circles. Globally this can be ensured by an appropriate choice of ε.) Hence $\delta(p) = \operatorname{dist}(p, \gamma)$.

For every differentiable function $F\colon \mathbb{R}^2 \to \mathbb{R}$ we have

$$\|\nabla F\| = \max_{\|v\|=1} \left|\frac{\partial F}{\partial v}\right|$$

The derivative of δ in the direction of the oriented normal equals 1:

$$\frac{\partial\delta}{\partial\nu(t)} = 1.$$

Hence, $\|\nabla\delta\| \geq 1$. Assume there is a unit vector v such that $\left|\frac{\partial\delta}{\partial v}\right| > 1$. Then for a sufficiently small $\alpha > 0$ we have

$$\delta(p + \alpha v) > \delta(p) + \alpha$$

On the other hand, by the triangle inequality we have $\operatorname{dist}(p + \alpha v, \gamma) \leq \operatorname{dist}(p, \gamma) + \alpha$. This contradiction shows that all directional derivatives of δ are smaller than 1 in the absolute value. Hence $\|\nabla\delta\| = 1$. □

The condition $\|\nabla F(p)\| = 1$ for all p is very strong.

Theorem 1.20. *Let* $F\colon U \to \mathbb{R}$ *be a function on a planar domain with* $\|\nabla F(p)\| = 1$ *for all* $p \in U$. *Then locally the level sets of* F *are parallel curves.*

This statement is only local because the domain U may not have the shape of a tubular neighborhood.

First we need two lemmas about functions with gradient norm 1.

Lemma 1.21. *Assume that* $\|\nabla F\| = 1$ *everywhere in a convex domain* W. *Then for any* $p, q \in W$ *we have*

$$|F(p) - F(q)| \leq \|p - q\|$$

Proof. Let $v = \frac{p-q}{\|p-q\|}$ be the unit vector pointing from q to p. Connect p to q by a straight line segment $l(t) = q + tv$, $t \in [0, \|p - q\|]$. Then we have

$$F(p) - F(q) = \int_0^{\|p-q\|} \frac{\partial F}{\partial v}(l(t))\, dt$$

Since $\left|\frac{\partial F}{\partial v}\right| \leq 1$, the statement follows. □

Lemma 1.22. *Every gradient curve of a function $F\colon U \to \mathbb{R}$ with $\|\nabla F\| = 1$ is a straight line.*

A *gradient curve* of F is a curve λ such that $\dot\lambda(t) = \nabla F(\lambda(t))$ for all t. This is a curve that always follows the direction of the steepest ascent.

Proof. Let $\lambda\colon [a,b] \to U$ be a gradient curve of F. By the chain rule we have

$$\frac{d}{dt}F(\lambda(t)) = \langle \nabla F, \dot\lambda\rangle = \|\nabla F\|^2 = 1.$$

Together with Lemma 1.21 this implies

$$\|\lambda(b) - \lambda(a)\| \geq |F(\lambda(b)) - F(\lambda(a)) = b - a$$

On the other hand, the length of λ equals

$$\int_a^b \|\dot\lambda\|\, dt = b - a$$

But the length of a curve does not exceed the distance between its endpoints, and if the length is equal to distance, then this curve is a straight line. □

Proof of Theorem 1.19. Since $\nabla F \neq 0$, the level sets are regular curves. The lines perpendicular to the level curves are the gradient curves of F, by the lemma above. Since the restriction of F to each gradient curve measures the length of the curve (as $\|\nabla F\| = 1$), moving along each normal by distance δ brings us to another level set of F. Hence the level sets are parallel curves. □

Functions satisfying $\|\nabla F(p)\| = 1$ for all p are called *distance functions.*

2 Rotation index and convex curves

2.1 The total signed curvature

Smooth closed curve $\gamma\colon [a,b] \to \mathbb{R}^2$:

$$\frac{d^k\gamma}{dt^k}(a) = \frac{d^k\gamma}{dt^k}(b) \quad \text{for all } k \geq 0$$

(if we only assume $\gamma(a) = \gamma(b)$, then the curve may have a "knick").

Recall the definition of the function $\varphi\colon [a,b] \to [0, 2\pi)$. Note that φ may be discontinuous at the points where it takes value 0.

Lemma 2.1. *There is a continuous map $\tilde\varphi\colon I/\sim \to \mathbb{R}$ such that $\tilde\varphi(t) \equiv \varphi(t) (\mod 2\pi)$.*

For this new function we have

$$\kappa_s = \frac{d\tilde\varphi}{dt}$$

Theorem 2.2. *For every smooth closed curve its total signed curvature is an integer multiple of* 2π*:*

$$\int_\gamma \kappa_s \, dt = 2\pi\iota(\gamma), \quad \iota(\gamma) \in \mathbb{Z},$$

where t *is the arc-length parameter on the curve.*

The number $\iota(\gamma)$ is called the *rotation index* of γ.

Proof. We have

$$\int_\gamma \kappa_s \, dt = \int_a^b \frac{d\tilde{\varphi}}{dt} \, dt = \tilde{\varphi}(b) - \tilde{\varphi}(a) = 2\pi k,$$

since $\varphi(a) = \varphi(b)$ and $\tilde{\varphi}$ differs from φ by an integer multiple of 2π. □

2.2 Rotation of simple closed curves

Theorem 2.3. *The rotation index of every simple smooth closed curve is equal to* ± 1*.*

Proof. Proof by H. Hopf [Hop35], see [Hsi97]. □

2.3 Convex curves

Definition 2.4. *A curve is called* convex*, if it lies on one side of each of its tangent lines.*

Lemma 2.5. *The signed curvature of a convex curve does not change its sign.*

Proof. Assume that κ_s is negative on $(t_0 - \varepsilon, t_0)$ and positive on $(t_0, t_0 + \varepsilon)$. Then the arc $\gamma_{t_0-\varepsilon}^{t_0}$ lies on the right of the oriented tangent at t_0, and the arc $\gamma_{t_0}^{t_0+\varepsilon}$ on the left of this tangent. (Compare this with the behavior of the osculating circles.)

In general, the sign does not change at a point, but there is an interval (t_0, t_1) where κ vanishes, and the signs before t_0 and after t_1 differ. In this case $\gamma_{t_0}^{t_1}$ is a straight line segment and short arcs before t_0 and after t_1 lie on different sides from this line. □

In the following we assume that the curve does not "repeat itself" (like the double circle $\gamma\colon [0, 4\pi) \to \mathbb{R}^2$, $\gamma(t) = (\cos t, \sin t)$).

Theorem 2.6. *Every convex curve is simple.*

Proof. If the tangents at an intersection point differ, then each of them cuts the curve. If the tangents coincide, a tricky argument shows that the curve arcs must coincide in a neighborhood of the point. It follows that if a convex curve is not simple, it must repeat itself. □

Theorem 2.7. *The rotation index of a convex curve equals* ± 1*.*

Proof. Follows from Theorems 2.6 and 2.2. □

2.4 Convex curves in polar coordinates

Radial graph $r = r(t)$. Parametrization $(r(t)\cos t, r(t)\sin t)$.

Formulas for the arclength and the area.

Among all curves of a given diameter the circle encloses the largest area (proof mentioned in the Littlewood book).

3 Global geometry of plane curves

3.1 Support function

Definition 3.1. *Let $X \subset \mathbb{R}^2$ be a compact set. A* support line *of X is an oriented line that has at least one point in common with X and such that X lies on the left from this line.*

The support function *of X associates to every unit vector $v \in \mathbb{R}^2$ the signed distance from the origin to the support line of X with the exterior normal v:*

$$h_X(v) = \max_{x \in X} \langle v, x \rangle$$

If X is a convex positively oriented curve, then all its support lines are its tangents, compatibly oriented.

We will consider only positively oriented convex curves. (Note that this is equivalent to the requirement $\kappa_s \geq 0$.)

Definition 3.2. *A smooth convex curve is called* strictly convex*, if it has only one point in common with each of its tangents.*

Exercise 3.3. Show that $\kappa_s > 0$ implies strict convexity, but is not vice versa.

Strictly convex curves have a special parametrization: to every unit vector we can associate the point that has this vector as the exterior unit normal.

Definition 3.4. *The* Gauss parametrization *of a smooth strictly convex curve is a map*

$$\gamma \colon [0, 2\pi) \to \mathbb{R}^2$$

such that the exterior unit normal at t has direction t.

Note that the exterior unit normal is always opposite to the signed unit normal.

We now modify our definition of the support function and denote by $h(t)$ the value at the vector $(\cos t, \sin t)$.

Theorem 3.5. *The Gauss parametrization of a smooth strictly convex curve is related to the support function by the formula*

$$\gamma(t) = h \begin{pmatrix} \cos t \\ \sin t \end{pmatrix} + \dot{h} \begin{pmatrix} -\sin t \\ \cos t \end{pmatrix}$$

Proof. Denote $v(t) = (\cos t, \sin t)$ and $w(t) = (-\sin t, \cos t)$. This is an orthonormal basis, and $\dot v = w$. Let $\gamma(t) = a(t)v(t) + b(t)w(t)$. Use $\langle \gamma(t), v(t)\rangle = h(t)$. First, this implies $a(t) = h(t)$. Second, its differentiation yields

$$\langle \dot\gamma, v\rangle + \langle \gamma, \dot v\rangle = \dot h$$

The first term vanishes, since the normal is orthogonal to the tangent. Hence $b(t) = \langle \gamma, \dot v\rangle = \dot h$. □

3.2 The average width

Define the width of a compact set X in the direction of a unit vector v as the distance between the support lines with normals v and $-v$. In terms of the support function:

$$W_X(v) = h(v) + h(-v)$$

Equivalently, the width in direction v is the length of the projection to the line parallel to v.

Again, we will use t as the argument of the width function.

Theorem 3.6 (Cauchy-Crofton formula)**.** *The length of a convex closed curve is proportional to the integral of its projection lengths:*

$$L(\gamma) = \frac{1}{2}\int_0^{2\pi} W_\gamma(t)\,dt$$

Corollary 3.7. *All figures of constant width W have perimeter πW.*

Corollary 3.8. *The average width of a convex closed curve is proportional to its length:*

$$\frac{1}{2\pi}\int_0^{2\pi} W_\gamma(t)\,dt = \frac{1}{\pi}L(\gamma)$$

The formula holds for non-smooth convex curves as well, because they can be approximated by smooth convex ones.

Exercise 3.9. Find the average width of the unit square.

Proof of Theorem 3.6. Compute the length using the Gauss parametrization. First we need to compute the norm of the velocity vector. We have

$$\dot\gamma = \dot h \cdot v + h \cdot \dot v + \ddot h \cdot \dot w + \dot h \cdot \dot w = (h + \ddot h)w$$

because $\dot w = v$. Since v looks left both from $\dot\gamma$ and from w, we have $h + \ddot h > 0$, so that $\|\dot\gamma\| = h + \ddot h$.

$$L(\gamma) = \int_0^{2\pi}(h + \ddot h)\,dt = \int_0^{2\pi} h\,dt = \frac{1}{2}\int_0^{2\pi}(h(t) + h(-t))\,dt$$

□

For convex polygons the Cauchy-Crofton formula can be proved by a different elegant argument.

Exercise 3.10. Prove directly that the length of a segment is quarter of its average width. Derive the Cauchy-Crofton formula for convex polygons.

3.3 Isoperimetric inequality

Theorem 3.11. *Among all simple closed curves of given length the circle bounds the largest area. In other words, if A is the area bounded by a simple closed curve γ of length L, then*

$$L^2 - 4\pi A \geq 0,$$

and the equality holds only if γ is a circle.

We will prove the isoperimetric inequality under assumption that γ is a smooth convex curve.

Lemma 3.12. *The area of a smooth convex curve with support function h is given by*

$$A = \frac{1}{2}\int_0^{2\pi} (h^2 - (\dot{h})^2)\, dt$$

Proof. In the Gauss parametrization we have

$$\gamma = hv + \dot{h}w, \quad \dot{\gamma} = (h + \ddot{h})w$$

Hence

$$\gamma \times \dot{\gamma} = h(h + \ddot{h})$$

It follows that

$$A = \frac{1}{2}\int_0^{2\pi} (h^2 + h\ddot{h})\, dt = \frac{1}{2}\int_0^{2\pi} (h^2 - (\dot{h})^2)\, dt,$$

where integration by parts was used. □

Lemma 3.13 (Wirtinger). *Let $f\colon [0, 2\pi) \to \mathbb{R}$ be a smooth function with zero average:*

$$\int_0^{2\pi} f\, dt = 0$$

Then

$$\int_0^{2\pi} f^2\, dt \leq \int_0^{2\pi} (\dot{f})^2\, dt,$$

where the equality holds if and only if $f(t) = a\cos t + b\sin t$ for some $a, b \in \mathbb{R}$.

Proof. Wirtinger's inequality can be reformulated as a statement about the eigenvalues of the Laplace operator. We have

$$\int_0^{2\pi} (\dot{f})^2\, dt = -\int_0^{2\pi} \ddot{f} \cdot f\, dt = -\langle \Delta f, f \rangle_{L^2}$$

The assumption $\int_0^{2\pi} f\, dt = 0$ means that f is L^2-orthogonal to the space of constant functions, the kernel of the Laplace operator. Since the non-zero eigenvalues of Δ are $-n^2$, $n \in \mathbb{Z}$ (Fourier expansion of f), we have

$$-\langle \Delta f, f \rangle_{L^2} \geq \|f\|^2_{L^2},$$

which proves the inequality. The equality takes place if and only if f belongs to the (-1)-eigenspace, that is $f(t) = a\cos t + b \sin t$. □

Proof of Theorem 3.11. Let h be the support function of the curve γ. By scaling we can achieve $L = 2\pi$, that is $\int_0^{2\pi} h\, dt = 2\pi$. Put $f(t) = h(t) - 1$. Then $L = 2\pi \Rightarrow \int_0^{2\pi} f\, dt = 0$. Hence Wirtinger's lemma implies

$$A = \frac{1}{2}\int_0^{2\pi} (h^2 - \dot{h}^2)\, dt = \frac{1}{2}\int_0^{2\pi} (f^2 - \dot{f}^2)\, dt + \int_0^{2\pi} f\, dt + \pi \leq \pi$$

Equality holds if and only if $h(t) = 1 + a\cos t + b\sin t$, which is the support function of the unit circle centered at (a, b). □

3.4 Steiner formulas

Let γ be a smooth closed convex curve, and $\delta > 0$. The parallel curve $\gamma_{-\delta}$ lies in the exterior of γ, and we will call it the *exterior parallel curve.*

Lemma 3.14. *If h is the support function of γ, then the support function of the exterior parallel curve at distance δ equals $h + \delta$.*

Proof. If the maximum of $\langle x, v\rangle$ is attained at $x = \gamma(t)$, then v is the exterior unit normal of γ at t. Then $\gamma(t) + \delta v$ is a point on the parallel curve, and it can easily be seen that it realizes the maximum of the scalar product with v over the parallel curve. □

Theorem 3.15. *Let L and A be the length and the area of a smooth closed convex curve γ. Then the length and the area enclosed by the exterior parallel curve at distance δ are equal*

$$L_\delta = L + 2\pi\delta, \qquad A_\delta = A + L\delta + \pi\delta^2$$

Proof. Using the lemma above and the formulas for the length and area in terms of the support function we compute

$$L_\delta = \int_0^{2\pi} (h + \delta)\, dt = L + 2\pi\delta$$

$$A_\delta = \frac{1}{2}\int_0^{2\pi} \left((h+\delta)^2 - (\dot{h})^2\right)\, dt = A + \delta\int_0^{2\pi} h\, dt + \frac{\delta^2}{2}\int_0^{2\pi} dt = A + L\delta + \pi\delta^2$$

□

Exercise 3.16. Prove Steiner formulas for convex polygons.

3.5 Gauss map

The Gauss parametrization of a smooth convex closed curve can be viewed as a map $\mathbb{S}^1 \to C \subset \mathbb{R}^2$. Its inverse is called the Gauss map.

In fact, while the Gauss parametrization is defined only for strictly convex curves, the Gauss map makes sense for non-strictly convex curves as well.

Definition 3.17. *Let $C \subset \mathbb{R}^2$ be a smooth convex closed curve. The* Gauss map

$$\Gamma\colon C \to \mathbb{S}^1$$

associates to every point $p \in C$ the exterior unit normal to C at p.

Theorem 3.18. *The Jacobian of the Gauss map is equal to the absolute curvature of the curve.*

Here by the Jacobian we mean the length scaling factor. Both C and $\mathbb{S}^1$ are equipped with a measure (the angular measure on $\mathbb{S}^1$ and the $\|\dot\gamma\| dt$ measure on C, which is independent of the parametrization γ). The measure of the Gauss image of an arc divided by the measure of the arc tends to a limit as the arc length tends to zero.

Proof. This is a reformulation of Theorem 1.4:

$$\frac{d\varphi}{d\ell} = \kappa_s \Rightarrow \left|\frac{d\varphi}{d\ell}\right| = \kappa$$

Here we denote by ℓ an arc-length parameter on C. □

As a consequence, the integral of any function over curve C can be rewritten as the integral over $\mathbb{S}^1$:

$$\int_C f\, d\ell = \int_0^{2\pi} f\kappa^{-1}\, dt$$

For example, the total curvature of a convex curve can now be computed as

$$\int_C \kappa\, d\ell = \int_0^{2\pi} dt = 2\pi$$

Theorem 3.19. *The radius of curvature of a smooth convex closed curve can be computed from its support function by the formula*

$$R = h + \ddot{h}$$

Proof. Indeed, as we found out in the proof of Theorem 3.6, the Jacobian of the Gauss parametrization equals $h + \ddot{h}$. Since the Jacobian of the inverse is the inverse of the Jacobian, and the radius of curvature is the inverse of the curvature, Theorem 3.18 implies $R = h + \ddot{h}$. □

3.6 Minkowski formulas

Theorem 3.20. *For every smooth convex closed curve $C \subset \mathbb{R}^2$ the following identities hold:*

$$\int_C \kappa\nu \, d\ell = 0, \quad \int_C \nu \, d\ell = 0$$

Proof. Use the Gauss parametrization of C. Then the identities become

$$\int_0^{2\pi} v(t)\, dt = 0, \quad \int_0^{2\pi} \kappa^{-1}(t) v(t)\, dt = 0$$

The first one is obvious. For the second one remember that $\kappa^{-1} = h + \ddot{h}$ and apply integration by parts:

$$\int_0^{2\pi} \ddot{h} v \, dt = -\int_0^{2\pi} \dot{h}\dot{v} \, dt = \int_0^{2\pi} h\ddot{v} \, dt$$

Since $\ddot{v} = -v$, the result follows. □

Remark 3.21. The second Minkowski formula has a physical interpretation. It says that the vector of the total pressure exerted on the perimeter of a planar balloon is a zero vector.

3.7 Four-vertex theorem

Definition 3.22. *A* vertex *of a smooth curve is a point of a local extremum of the curvature.*

An ellipse has four vertices. A circle has infinitely many.

Theorem 3.23 (Four-vertex theorem)**.** *Every simple closed smooth curve has at least four vertices.*

Note that the simplicity assumption is essential. There are non-simple closed curves with only two vertices (one is the global minimum, the other is the global maximum of curvature).

We will prove the four-vertex theorem for convex curves only.

Lemma 3.24. *For every smooth convex closed curve parametrized with the unit speed the following identity holds:*

$$\int_I \dot{\kappa}\gamma \, dt = 0,$$

Proof. Apply integration by parts:

$$\int_I \dot{\kappa}\gamma \, dt = -\int_I \kappa\dot{\gamma} \, dt = 0$$

The last integral is equal to zero since $\int_I \kappa\nu \, dt = 0$ by the first Minkowski formula, and because the unit tangent vector $\dot{\gamma}$ is obtained by $90°$-rotation from the unit normal ν. □

Proof of the four vertex theorem. The above lemma implies that for every vector $c \in \mathbb{R}^2$ we have

$$\int_I \dot{\kappa}\langle\gamma, c\rangle\, dt = 0 \tag{2.1}$$

Assume the contrary: there is a curve that has only two vertices P (the global maximum of the curvature) and Q (the global minimum of the curvature). The line PQ cuts the curve in two arcs $\vec{PQ}$ and $\vec{QP}$. Choose the coordinate origin on the line PQ and a vector c orthogonal to PQ and looking towards the arc $\vec{QP}$. Then on the arc $\vec{QP}$ we have $\dot{\kappa} \geq 0$ and $\langle\gamma, c\rangle > 0$, and on the arc $\vec{PQ}$ we have $\dot{\kappa} \leq 0$ and $\langle\gamma, c\rangle < 0$. It follows that the integrand in (2.1) is everywhere non-negative (and somewhere positive), which is a contradiction. □

An alternative proof of the four-vertex theorem is based on the following.

Theorem 3.25 (Sturm). *A periodic function free of the harmonics of order $< k$ changes the sign at least $2k$ times.*

Indeed, the radius of curvature in terms of the support function $h + \ddot{h}$ contains no harmonics of order 1; hence its derivative contains no harmonics of order < 2. Sturm theorem implies that the derivative of the curvature changes the sign at least four times, hence the curvature has at least four extrema.

The proof of the Sturm theorem is based on the same idea as the above proof of the four-vertex theorem. If a function changes its sign at most $2k - 2$ times, then there is a trigonometric polynomial of degree $< k$ that changes its sign at the same points. Then the L^2-product of the function and the polynomial is positive; on the other hand a function without harmonics of order $< k$ is L^2-orthogonal to trigonometric polynomials of degree $< k$.

3.8 Steiner point

Definition 3.26. *The* Steiner point *of a smooth convex closed curve is defined by means of the Gauss parametrization as*

$$\mathrm{St}(\gamma) = \frac{1}{\pi} \int_0^{2\pi} h(t) v(t)\, dt$$

where, as usual, $v(t) = (\cos t, \sin t)$.

Don't confuse it with other points (associated to triangles) also carrying the name of Steiner!

Exercise 3.27. Show that the Steiner point is the center of mass for a mass distributed along the curve with the density proportional to the curvature. Because of this the Steiner point is sometimes called the *curvature centroid*.

In particular, the Steiner point is equivariant with respect to rigid motions: if a curve is translated or rotated, its Steiner point is subject to the same transformation.

Exercise 3.28. Show that the Steiner point possesses the following extremal property. Roll the curve along a straight line. For every point inside the curve consider its (periodic) trajectory and the area under (a period of) this trajectory. The Steiner point minimizes this area.

This is the original definition given by Steiner [Ste40].

Chapter 3

Space curves

1 Torsion

1.1 Frenet formulas

Suppose we have a space curve $\gamma\colon I \to \mathbb{R}^3$ with nowhere vanishing curvature. This means that, if γ is of unit-speed (which we assume in this section), then

$$\ddot{\gamma} = \kappa\nu \neq 0$$

The vectors $\dot{\gamma}$ and ν can be completed to a positively oriented orthonormal basis

$$e_1(t) = \dot{\gamma}, \quad e_2(t) = \nu, \quad e_3(t) = e_1(t) \times e_2(t)$$

Definition 1.1. *The vectors $e_1(t), e_2(t), e_3(t)$ (the unit tangent, the normal, and the binormal) applied to the point $\gamma(t)$ form a moving frame, called the* Frenet frame *of the curve γ.*

Let us compute the change of the vectors $e_i(t)$ as the point moves along the curve. We already know

$$\dot{e}_1 = \kappa e_2$$

Determine now the coefficients in

$$\dot{e}_2 = ae_1 + be_3, \quad \dot{e}_3 = ce_1 + de_2$$

(recall that $\langle \dot{e}_2, e_2 \rangle = 0$ because of $\|e_2(t)\| = 1$). We have

$$a = \langle \dot{e}_2, e_1 \rangle = -\langle e_2, \dot{e}_1 \rangle = -\kappa$$

Definition 1.2. *The coefficient at e_3 in the decomposition of $\dot{e}_2$ with respect to the Frenet frame is called the* torsion *of the curve γ:*

$$\dot{e}_2 = -\kappa e_1 + \tau e_3$$

Further, we compute

$$c = \langle \dot{e}_3, e_1 \rangle = -\langle e_3, \dot{e}_1 \rangle = 0$$
$$d = \langle \dot{e}_3, e_2 \rangle = -\langle e_3, \dot{e}_2 \rangle = -\tau$$

Altogether, the formulas

$$\begin{aligned} \dot{e}_1 &= \kappa e_2 \\ \dot{e}_2 &= -\kappa e_1 + \tau e_3 \\ \dot{e}_3 &= -\tau e_2 \end{aligned}$$

are called the *Frenet formulas.*

Remark 1.3. The Frenet formulas can be written as

$$\begin{pmatrix} \dot{e}_1 \\ \dot{e}_2 \\ \dot{e}_3 \end{pmatrix} = \begin{pmatrix} 0 & \kappa & 0 \\ -\kappa & 0 & \tau \\ 0 & -\tau & 0 \end{pmatrix} \begin{pmatrix} e_1 \\ e_2 \\ e_3 \end{pmatrix}$$

The 3×3-matrix in this formula is skew-symmetric. This holds for any moving frame and is due to the fact that the skew-symmetric matrices form the tangent space to the orthogonal group at the point id.

Example 1.4. The curvature and the torsion of the circular helix

$$\gamma(t) = (a \cos t, a \sin t, bt)$$

1.2 Osculating plane and osculating sphere

The plane spanned by the tangent and the normal of a curve γ at the point t_0 is called the *osculating plane* at t_0. This is the plane that is the closest to γ in a neighborhood of t_0. To find how close it is, let us consider the Taylor expansion of γ near t_0.

Theorem 1.5. *Assume that $\gamma(0) = 0$ and put $\kappa_0 = \kappa(0)$, $\tau_0 = \tau(0)$. Then the coordinates of $\gamma(t)$ in the Frenet frame of the point $\gamma(0)$ have the form*

$$\begin{aligned} x_1(t) &= t - \frac{\kappa_0^2}{6} t^3 + o(t^3) \\ x_2(t) &= \frac{\kappa_0}{2} t^2 + \frac{\kappa'(0)}{6} t^3 + o(t^3) \\ x_3(t) &= \frac{\kappa_0 \tau_0}{6} t^3 + o(t^3) \end{aligned}$$

Projections of the curve to the planes spanned by the Frenet basis vectors (in the case $\kappa_0 \neq 0$, $\tau_0 \neq 0$).

If $\kappa_0 \neq 0$ and $\tau_0 \neq 0$, then the osculating plane cuts the curve locally in two parts.

The tangent line approximates the curve up to terms of order 2 (the t^2 term vanishes if $\kappa = 0$). The osculating circle approximates the curve up to terms of order 3 (when does the t^3 term vanish?). The osculating plane approximates the curve up to terms of order 3 (the t^3 vanishes if $\kappa\tau = 0$).

Definition 1.6. *The osculating sphere of the curve γ at the point $\gamma(t)$ is the sphere that approximates γ near t up to terms of order 4 at least.*

Theorem 1.7. *The center of the osculating sphere at the point $\gamma(t)$ is the point $\gamma(t) + \frac{1}{\kappa(t)} e_2(t) + \frac{1}{\tau(t)} \left(\frac{1}{\kappa(t)}\right)^{\cdot} e_3(t)$.*

Proof. Substitute the formulas of Theorem 1.5 into the equation of the sphere with center (a_1, a_2, a_3) passing through the origin:

$$x_1^2 + x_2^2 + x_3^2 - 2a_1x_1 - 2a_2x_2 - 2a_3x_3 = 0$$

The requirement that the coefficients at t, t^2, and t^3 vanish yields a system of linear equations on a_1, a_2, a_3. By solving it we obtain the coordinates of the center of the osculating sphere with respect to the Frenet frame $\left(0, \frac{1}{\tau}, \frac{1}{\kappa}\left(\frac{1}{\tau}\right)^{\cdot}\right)$. □

Exercise 1.8. Let $c(t)$ be the curve formed by the centers of the osculating spheres.

1. Show that the tangent to c at t is parallel to the binormal of γ at t. (Compare this with the situation in the plane, where the tangents to the evolute are parallel to the normals of the curve.)

2. Show that the singular points of c correspond to the points of γ where

$$\frac{\tau}{\kappa} + \left(\frac{1}{\tau}\left(\frac{1}{\kappa}\right)^{\cdot}\right)^{\cdot} = 0 \tag{3.1}$$

3. Show that the curve γ lies on a sphere if and only if the above equation holds everywhere on γ.

1.3 Existence and uniqueness theorems

The initial position of the Frenet frame, the curvature (non-vanishing), and the torsion, both given as functions of the arclength, determine the curve uniquely.

2 Global geometry of space curves

2.1 Curves on the sphere

The result of the last exercise immediately implies that for every closed space curve contained in a sphere we have

$$\int_\gamma \frac{\tau}{\kappa}\, d\ell = 0$$

There is a more elegant vanishing theorem.

Theorem 2.1 (Geppert)**.** *The total torsion of every closed spherical curve equals zero:*

$$\int_\gamma \tau\, d\ell = 0$$

In order to prove this, let us look at the spherical curves more closely. It suffices to study the curves on the unit sphere centered at the origin: $\|\gamma(t)\| = 1$ for all t. Introduce the notation $e_0(t) = \gamma(t)$. In addition to the Frenet formulas we then have $\dot{e}_0 = e_1$.

Lemma 2.2. *The curvature of a curve on the unit sphere is at least* 1*.*

Proof. We have $\kappa = \|\dot{e}_1\| = \|\ddot{e}_0\|$. On the other hand,

$$0 = \langle e_0, \dot{e}_0 \rangle^{\cdot} = 1 + \langle e_0, \ddot{e}_0 \rangle$$

Therefore

$$\kappa = \|\ddot{e}_0\| \geq |\langle e_0, \ddot{e}_0 \rangle| = 1$$

□

Let $\theta(t)$ be the angle between the osculating plane of the curve and the great circle through $\gamma(t)$ in direction $\dot{\gamma}(t)$. We take $\theta \in (-\frac{\pi}{2}, \frac{\pi}{2})$ so that it is positive if e_0, e_1, e_2 form a positively oriented basis, as on Figure 3.1. Note that e_0, e_2, e_3 are coplanar (since all of them are orthogonal to e_1), and that

$$e_0 = -\cos\theta \cdot e_2 + \sin\theta \cdot e_3$$

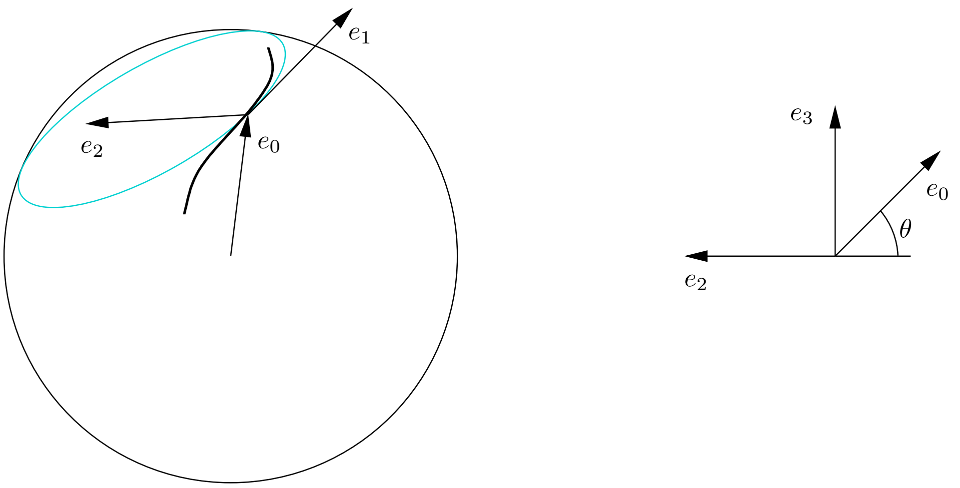

Figure 3.1: Curve on a sphere.

Theorem 2.3. *In terms of the angle θ just introduced the curvature and the torsion of a spherical curve can be expressed as*

$$\kappa = \frac{1}{\cos\theta}, \quad \tau = \dot{\theta}$$

Proof. From the argument in the previous lemma and from $\dot{e}_1 = \kappa e_2$ we have

$$-1 = \langle e_0, \dot{e}_1 \rangle = \kappa \langle e_0, e_2 \rangle = -\kappa \cos\theta$$

Further, by differentiating the equation $-\cos\theta = \langle e_0, e_2 \rangle$ we obtain

$$\dot{\theta} \sin\theta = \langle \dot{e}_0, e_2 \rangle + \langle e_0, \dot{e}_2 \rangle = \langle e_0, -\kappa e_1 + \tau e_3 \rangle = \tau \sin\theta$$

□

Now we can reprove the first theorem of this section and prove Geppert's theorem:

$$\int_C \frac{\tau}{\kappa}\, d\ell = \int_0^L \dot{\theta} \cos\theta\, dt = \int_0^L (\sin\theta)^{\cdot}\, dt = 0$$

$$\int_C \tau\, d\ell = \int_0^L \dot{\theta}\, dt = 0$$

Also, it follows that $\kappa = \frac{1}{\cos}\theta, \tau = \dot{\theta}$ provides a general solution of the differential equation (3.1).

Interestingly enough, the torsion is the speed of the binormal indicatrix, which is itself a spherical curve. The binormal indicatrix is the spherical analog of the evolute. Thus, Geppert's theorem is a spherical analog of the fact that the signed length of the evolute vanishes.

The dual curves (?).

For more about curves on the sphere see [Arn95].

2.2 Fenchel theorem

Theorem 2.4 (Fenchel)**.** *The total curvature of a space curve does not exceed* 2π*. It equals* 2π *only for convex plane curves.*

As a preparation for the proof, we will need a definition, two lemmas, and a theorem.

Definition 2.5. *Let* γ *be a unit-speed space curve. Then the curve on the unit sphere traced by the vector* $\dot{\gamma}$ *is called the* tangent indicatrix *of* γ*.*

For example, if γ lies in a plane, its tangent indicatrix is contained in the great circle parallel to this plane. In this case the indicatrix will backtrack at the places where the signed curvature changes its sign; it will be a simple curve if and only if γ is convex.

Lemma 2.6. *The total curvature is the length of the tangent indicatrix.*

Proof. Indeed,

$$\int_C \kappa\, dt = \int_0^L \|\ddot{\gamma}\|\, dt = \int_0^L \|\dot{e}_1\|\, dt = L(e_1)$$

and the curve $e_1 : [0, L] \to \mathbb{S}^2$ is the tangent indicatrix. □

Lemma 2.7. *The tangent indicatrix intersects each great circle at least twice.*

Proof. Every great circle can be represented as

$$v^\circ = v^\perp \cap \mathbb{S}^2$$

for some non-zero vector $v \in \mathbb{R}^3$. Now

$$e_1(t) \in v^\circ \Leftrightarrow \langle e_1(t), v\rangle = 0 \Leftrightarrow \langle \gamma(t), v\rangle^{\cdot} = 0$$

But the function $t \mapsto \langle \gamma(t), v\rangle$ has at least two critical points: the global maximum and the global minimum. Therefore the curve e_1 intersects v° at least twice. □

Theorem 2.8 (Crofton)**.** *Let C be a smooth curve of length L on the unit sphere. Let M be the measure of the oriented great circles intersecting C, each counted with a multiplicity which is the number of its points of intersection with C. Then*

$$M = 4L$$

The measure on the set of oriented great circles is induced from the standard measure on the sphere through the bijection between oriented great circles and points on the sphere: to every equator there correspond two poles, and if the equator is oriented, there is a consistent way to choose one of these poles.

Proof. For a formal proof, see [Hsi97]. We explain here why the theorem holds for arcs of great circles.

An oriented great circle S intersects an arc pq if and only if its pole S° lies between the polars p° and q° of p and q. The space between p° and q° consists of two "lunes", each of the angle L, where L is the distance between p and q. The area of each lune is $2L$. □

Compare this with the Cauchy-Crofton formula in the plane, which can be interpreted in terms of the measure of the set of lines intersecting a convex curve.

Proof of the Fenchel theorem. Since the tangent indicatrix intersects every oriented great circle at least twice, its length is at least 2π. Hence the total curvature is at least 2π. The equality takes place only if the tangent indicatrix is a great circles, traced without backtracking. This means that γ is a convex plane curve. □

2.3 Weyl tube formula

Theorem 2.9. *Let γ be a simple curve of length L in $\mathbb{R}^2$ or $\mathbb{R}^3$. Then the volume of the tubular ε-neighborhood of γ equals*

$$\begin{cases} 2\varepsilon L & \text{in the plane} \\ \pi\varepsilon^2 L & \text{in the space} \end{cases}$$

An analog holds for curves in $\mathbb{R}^n$ with $n > 3$. The length is multiplied with the volume of the ball of radius ε in $\mathbb{R}^n$.

Proof. Parametrize the tubular neighborhood by the rectangle/cylinder and integrate the Jacobian.

In the 3-dimensional case the normal-binormal parametrization is well-defined and continuous only if the curvature does not vanish. In the general situation, however, one can take a parametrization defined by any moving orthonormal frame $(e_1, \bar{e}_2, \bar{e}_3)$. □

For convex curves in the plane, the Weyl tube formula also follows from the Steiner formula. Just subtract $A_{-\delta}$ from A_δ.

Chapter 4

Surfaces: the first fundamental form

1 Definitions and examples

1.1 What is a surface?

Similar to curves, there are three convenient ways to represent a surface in $\mathbb{R}^3$:

- as the graph of a function $f\colon U \to \mathbb{R}$ with $U \subset \mathbb{R}^2$;
- as a level set $F(x, y, z) = 0$;
- as the image of a map $\sigma\colon U \to \mathbb{R}^3$ with $U \subset \mathbb{R}^2$.

Again, the third way is the most flexible one. But we encounter difficulties when trying to represent some topologically non-trivial surfaces. In the case of curves we could represent a closed curve as the image of a, say, periodic map; in the case of surfaces it is hard to parametrize the sphere by a "good" map. This leads us to use a collection of *local parametrizations* in order to represent a surface.

For simplicity, we will avoid dealing with self-intersecting surfaces (although can later allow self-intersections). The following definition explains, what an *embedded surface without boundary* is.

Definition 1.1. *A subset $S \subset \mathbb{R}^3$ is a* surface *if, for every point $p \in S$ there is an open set $W \subset \mathbb{R}^3$ containing p and a homeomorphism $\Phi\colon W \to W'$ onto an open subset of $\mathbb{R}^3$ such that $W' \cap \mathbb{R}^2 = \Phi(W \cap S)$.*

Informally speaking, the map Φ "flattens out" the surface in a neighborhood of p. Note that we don't require the set $W' \cap \mathbb{R}^2$ to be connected.

Definition 1.2. *Let $U \subset \mathbb{R}^2$ be an open subset, and $\sigma\colon U \to \mathbb{R}^3$ be a continuous map whose image is contained in the surface S. The pair (U, σ) is called a* local parametrization *of S or a* surface patch.

The definition of a surface implies that for every point on a surface there is a surface patch covering its neighborhood. One may suggest this as a "simplified" definition of a surface: a subset $S \subset \mathbb{R}^3$ such that for every $p \in S$ there is a neighborhood of p in S homeomorphic to an open subset of $\mathbb{R}^2$. But this is not equivalent to the definition given above. Counterexample: Alexander's horned sphere.

For smooth surfaces there is no such a problem, so we may use the weak definition as well.

1.2 Examples

A collection of local parametrizations is called an *atlas*.

Example 1.3. Three atlases for the sphere.

- The spherical coordinates

$$(\theta, \varphi) \mapsto (\cos\theta\cos\varphi, \cos\theta\sin\varphi, \sin\theta)$$

  allow to cover the unit sphere by the rectangle $[-\pi/2, \pi/2] \times [0, 2\pi]$. However, this is not a homeomorphism (examine $\theta = \pm\pi/2$), and the rectangle is closed. We get a surface patch if we restrict the map to the open rectangle. This gives a parametrization of the sphere minus half of a great circle (a meridian). We can complete the atlas by appropriately rotating the mapped region so as to cover the complement of a great half-circle disjoint from the unmapped meridian.

- The stereographic projection

$$\mathbb{R}^2 \to \mathbb{S}^2, \quad (x, y) \mapsto \left(\frac{2x}{1 + x^2 + y^2}, \frac{2y}{1 + x^2 + y^2}, \frac{-1 + x^2 + y^2}{1 + x^2 + y^2}\right)$$

  is a homeomorphism onto the complement of the northern pole $(0, 0, 1)$. It can be completed to an atlas by a projection onto the complement of the southern pole.

- The map

$$\{(x, y) \mid x^2 + y^2 < 1\} \to \mathbb{S}^2, \quad (x, y) \mapsto (x, y, \sqrt{1 - x^2 - y^2})$$

  is a part of an atlas that consists of 6 local parametrizations.

Example 1.4. A *surface of revolution* is the surface obtained by rotating a plane curve around a line lying in the same plane. If the curve and the line are disjoint, then we obtain an embedded surface. Local parametrizations can be computed from a parametrization of the curve. If $\gamma(t) = (f(t), 0, g(t))$, then the surface consists of the points

$$\sigma(t, \varphi) = (f(t)\cos\varphi, f(t)\sin\varphi, g(t)),$$

and can be covered by two patches $(0, 2\pi) \times I$ and $(-\pi, \pi) \times I$ (assuming I is an open interval and γ is a simple curve).

Torus.

Exercise 1.5. It is easy to find an atlas of four maps on the torus. Does there exist an atlas of three maps?

Example 1.6. Quadric surfaces. Ellipsoid, hyperboloids of one and of two sheets, hyperbolic paraboloid.

Exercise 1.7. Show that the hyperboloid of one sheet $x^2+y^2-z^2 = 1$ and the hyperbolic paraboloid $x^2 - y^2 = z$ are *doubly ruled*, that is every point lies on two different lines contained in the surface.

1.3 Smooth surfaces

Up to now, we spoke about topological surfaces. For example, the boundary of a convex polyhedron is a surface in the sense of our definition (find an atlas of 4 maps on the surface of the tetrahedron).

Definition 1.8. *A surface patch $\sigma\colon U \to \mathbb{R}^3$ is called* regular *if it is smooth and the vectors σ_u and σ_v are linearly independent at all points $(u, v) \in U$.*

A *smooth surface* is a surface that can be parametrized by regular surface patches.
Reparametrizations and transition maps.
Tangent space and normals.
Confocal quadrics.

2 Measuring lengths and areas

This section closely follows Pressley [Pre10].

2.1 Lengths of curves on surfaces

Let $\sigma\colon U \to \mathbb{R}^3$ be a regular surface patch. Any space curve within $\sigma(U)$ is the image of a curve within U:

$$\gamma(t) = \sigma(u(t), v(t)), \quad t \in [a, b]$$

Using the known formula for the length of a space curve and the chain rule, we compute

$$\|\dot\gamma\|^2 = \langle \sigma_u\dot u + \sigma_v\dot v, \sigma_u\dot u + \sigma_v\dot v\rangle = \|\sigma_u\|^2\dot u^2 + 2\langle\sigma_u, \sigma_v\rangle\dot u\dot v + \|\sigma_v\|^2\dot v^2$$

$$L(\gamma) = \int_a^b \|\dot\gamma\|\, dt = \int_a^b \sqrt{E\dot u^2 + 2F\dot u\dot v + G\dot v^2}\, dt,$$

where

$$E = \|\sigma_u\|^2, \quad F = \langle\sigma_u, \sigma_v\rangle, \quad G = \|\sigma_v\|^2$$

The expression

$$ds^2 = Edu^2 + 2Fdudv + Gdv^2$$

is called the *first fundamental form* of the surface S with respect to the local parametrization σ. This is a symmetric $(0,2)$-tensor field (a modern notation would be $Edu \otimes du + F(du \otimes dv + dv \otimes du) + Gdv \otimes du)$ on U. We will use the notation

$$\mathrm{I}_\sigma(X,Y) = Ex_1x_2 + F(x_1y_2 + x_2y_1) + Gy_1y_2 = X^\top \begin{pmatrix} E & F \\ F & G \end{pmatrix} Y,$$

where $X = (x_1, x_2), Y = (y_1, y_2) \in \mathbb{R}^2 = TU$. To every point of U this associates a positive definite symmetric bilinear form, which is the scalar product in $\mathbb{R}^3$ composed with the linear map $d\sigma$:

$$\mathrm{I}_\sigma(X,Y) = \langle d\sigma(X), d\sigma(Y)\rangle$$

Example 2.1.

- The first fundamental form of the sphere in the spherical coordinates:

$$d\theta^2 + \cos^2\theta d\varphi^2$$

- More generally, the surface of revolution $(f(t)\cos\varphi, f(t)\sin\varphi, g(t))$ generated by a unit-speed parametrized curve $(f(t), g(t))$ has (with respect to the usual parametrization) the first fundamental form

$$dt^2 + f(t)d\varphi^2$$

The first fundamental form changes under reparametrization $\tilde{\sigma} = \sigma \circ \varphi$ as follows:

$$\mathrm{I}_{\tilde{\sigma}}(X,Y) = \langle d\tilde{\sigma}(X), d\tilde{\sigma}(Y)\rangle = \mathrm{I}(d\varphi(X), d\varphi(Y))$$

In other words,

$$\begin{pmatrix} \tilde{E} & \tilde{F} \\ \tilde{F} & \tilde{G} \end{pmatrix} = J^\top \begin{pmatrix} E & F \\ F & G \end{pmatrix} J,$$

where J is the Jacobi matrix of the map φ.

2.2 Isometries

2.3 Conformal maps

2.4 Surface area

Chapter 5

Curvature of surfaces

1 The second fundamental form

1.1 Deviation from the tangent plane

Compute the distance of the surface from one of its tangent planes. Given a surface patch σ, take a point $\sigma(u,v)$, denote by ν the unit normal at this point. Then

$$\text{dist} = \langle \sigma(u+\delta u, v+\delta v) - \sigma(u,v), \nu \rangle$$

From the Taylor expansion,

$$\text{dist} = \frac{1}{2}\left(L(\delta u)^2 + 2M\delta u \delta v + N(\delta v)^2\right) + o((\delta u)^2 + (\delta v)^2),$$

where

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

The expression

$$II_\sigma = Ldu^2 + 2Mdudv + Ndv^2$$

is called the *second fundamental form* of the surface with respect to a local parametrization σ.

Lemma 1.1. *We have*

$$II_\sigma(X,Y) = -\langle d\sigma(X), d\nu(Y) \rangle = -\langle d\sigma(Y), d\nu(X) \rangle$$

Proof. We have to show that

$$L = -\langle \sigma_u, \nu_u \rangle, \quad M = -\langle \sigma_u, \nu_v \rangle = -\langle \sigma_v, \nu_u \rangle, \quad N = -\langle \sigma_v, \nu_v \rangle$$

But this follows from differentiating the identities $\langle \sigma_u, \nu \rangle = 0 = \langle \sigma_v, \nu \rangle$. □

Lemma 1.2. *Let $\tilde{\sigma} = \sigma \circ \varphi$ be a reparametrization of the surface, and let J_φ be the Jacobi matrix of φ. Then*

$$\begin{pmatrix} \tilde{L} & \tilde{M} \\ \tilde{M} & \tilde{N} \end{pmatrix} = \pm J^\top \begin{pmatrix} L & M \\ M & N \end{pmatrix} J,$$

where we take the plus sign if $\det J_\varphi > 0$*, and the minus sign if* $\det J_\varphi < 0$.

Example 1.3. Compute the first fundamental form of a surface of revolution

$$\sigma(t, \varphi) = (f(t)\cos\varphi, f(t)\sin\varphi, g(t))$$

under assumption that the profile curve is unit-speed parametrized: $(\dot{f})^2 + (\dot{g})^2 = 1$.

We obtain

$$II_\sigma = (\dot{f}\ddot{g} - \ddot{f}\dot{g})dt^2 + f\dot{g}d\varphi^2$$

In particular, if $(f(t), g(t)) = (1, t)$ (the surface is a cylinder), then $II_\sigma = d\varphi^2$. If $(f(t), g(t)) = (\cos t, \sin t)$ (the surface is a unit sphere), then $II_\sigma = dt^2 + \cos^2 t d\varphi^2 = I_\sigma$.

1.2 The curvature of curves on a surface

Let $\gamma(t) = \sigma(u(t), v(t))$ be a unit-speed curve on the surface. The velocity vector $\dot{\gamma}$ lies in the tangent plane. The acceleration vector $\ddot{\gamma}$ is orthogonal to $\dot{\gamma}$ but not necessarily orthogonal to the tangent plane. Let us decompose it into a normal and a tangential component. It is convenient to use the orthonormal frame $(\nu, \dot{\gamma}, \nu \times \dot{\gamma})$. The acceleration vector will lie in the plane spanned by ν and $\nu \times \dot{\gamma}$.

Definition 1.4. *Let*

$$\ddot{\gamma} = \kappa_n \nu + \kappa_g (\nu \times \dot{\gamma}).$$

Then κ_n is called the normal curvature*, and κ_g is called the* geodesic curvature *of γ.*

Note that the absolute curvature of γ is given by

$$\kappa = \|\ddot{\gamma}\| = \sqrt{\kappa_n^2 + \kappa_g^2}$$

Lemma 1.5. *The normal curvature of a unit-speed curve on a surface is given by*

$$\kappa_n = L\dot{u}^2 + 2M\dot{u}\dot{v} + N\dot{v}^2 = II_\sigma(\dot{\gamma}, \dot{\gamma})$$

Proof. From the definition of κ_n, we have

$$\begin{aligned} \kappa_n = \langle \ddot{\gamma}, \nu \rangle &= \left\langle \nu, \frac{d}{dt}(\sigma_u \dot{u} + \sigma_v \dot{v}) \right\rangle \\ &= \langle \nu, \sigma_u \ddot{u} + \sigma_v \ddot{v} + (\sigma_{uu}\dot{u} + \sigma_{uv}\dot{v})\dot{u} + (\sigma_{vu}\dot{u} + \sigma_{vv}\dot{v})\dot{v} \rangle = L\dot{u}^2 + 2M\dot{u}\dot{v} + N\dot{v}^2 \end{aligned}$$

□

Thus, the normal curvature of a curve on a surface depends only on the tangent vector of the curve, that is on the first-order data.

Theorem 1.6 (Meusnier)**.** *The curvature of a curve on a surface depends only on the osculating plane of the curve. Namely, if θ is the angle between the osculating plane of the curve and the tangent plane of the surface, then*

$$\kappa \sin \theta = \kappa_n.$$

In other words, the osculating circles of all curves with the same tangent form a sphere (whose radius is the inverse of the normal curvature in the direction of the tangent).

Proof. The angle between the normal to the surface and the acceleration vector equals $\frac{\pi}{2} - \theta$. Hence

$$\kappa_n = \langle \ddot{\gamma}, \nu \rangle = \|\ddot{\gamma}\| \|\nu\| \cos(\frac{\pi}{2} - \theta) = \kappa \sin \theta.$$

□

Compare this with a particular case of curves on the unit sphere.

Investigate what happens on a surface of revolution, if the tangent line is orthogonal to the rotation axis.

2 Principal curvatures

2.1 Definition

Definition 2.1. *The* principal curvatures *of a surface patch σ are the roots of the equation*

$$\det(II_\sigma - \kappa I_\sigma) = 0$$

Reparametrization does not change the principal curvatures.

The roots are real (principal axes theorem).

2.2 Curvature of surfaces of revolution

The principal directions and principal curvatures of a surface of revolution.

2.3 Curvature lines and triply orthogonal systems

2.4 The shape operator and the Weingarten matrix

Definition 2.2. *The* shape operator

$$S \colon T_pM \to T_pM, \quad S(X) = -\nabla_X \nu$$

The eigenvectors and the eigenvalues of the shape operator are the principal curvature directions and the principal curvatures.

3 The Gaussian curvature and the mean curvature

3.1 Definitions and formulas

3.2 The third fundamental form

3.3 The Gauss map and the Gauss-Bonnet theorem for convex surfaces

3.4 Covariant differentiation

Definition 3.1. *Let $f\colon M \to \mathbb{R}$. The* gradient *of f at $p \in M$ is a vector $\nabla^M f(p) \in T_pM$ such that*

$$\frac{\partial f}{\partial X}(p) = \langle \nabla^M f(p), X\rangle \quad \text{for all } X \in T_pM$$

Lemma 3.2. *For any smooth extension $\widetilde{f}\colon \mathbb{R}^3 \to \mathbb{R}$ of the function f we have*

$$\nabla^M f(p) = \top(\nabla \widetilde{f}).$$

Here $\top\colon \mathbb{R}^3 \to T_pM$ is the orthogonal projection; it produces the tangential component of a vector:

$$\top(Y) = Y - \langle Y, \nu\rangle\nu$$

Definition 3.3. *Let $Y\colon M \to \mathbb{R}^3$ be a vector field along the surface M, and let $X \in T_pM$. The* covariant derivative *of Y in the direction X is*

$$\nabla^M_X Y = \top(\nabla_X \widetilde{Y}).$$

On the right hand side we are taking the componentwise directional derivative of an arbitrary smooth extension $\widetilde{Y}$ of Y to $\mathbb{R}^3$.

Lemma 3.4. *The following Leibniz-type formulas hold:*

$$\nabla^M_X(fY) = (\nabla^M_X f)Y + f \cdot \nabla^M_X Y$$
$$\nabla^M_X\langle Y, Z\rangle = \langle \nabla^M_X Y, Z\rangle + \langle Y, \nabla^M_X Z\rangle$$

Theorem 3.5. *The covariant differentiation and the second fundamental form are related through the following formula:*

$$\nabla^M_X Y = \nabla_X Y - II(X, Y) \cdot \nu$$

Proof. By definition,

$$\nabla^M_X Y = \nabla_X Y - \langle \nabla_X Y, \nu\rangle\nu$$

On the other hand,

$$\langle \nabla_X Y, \nu\rangle = -\langle \nabla_X \nu, Y\rangle = II(X, Y)$$

because of $\langle Y, \nu\rangle = 0$. The theorem follows. □

3.5 Hessian and Laplacian

Definition 3.6. *The* Hessian operator *of a function* $f\colon M \to \mathbb{R}$ *is defined as*

$$\mathrm{Hess}^M f(X) = \nabla^M_X(\nabla f)$$

The Hessian quadratic form *is defined as*

$$\mathrm{Hess}^M f(X,Y) = \langle \nabla^M_X(\nabla f), Y\rangle$$

Lemma 3.7. *The Hessian quadratic form is symmetric:* $\mathrm{Hess}^M f(X,Y) = \mathrm{Hess}^M f(Y,X)$.

The definition is motivated by the matrix of second derivatives of a function $f\colon \mathbb{R}^3 \to \mathbb{R}$. There we have

$$\mathrm{Hess}\, f(X,Y) = \langle \nabla_X(\nabla f), Y\rangle = X^\top (D^2 f) Y,$$

where $D^2 f = \left(\frac{\partial f^2}{\partial x_i \partial x_j}\right)$.

Theorem 3.8. *The Hessian and the Laplacian on the surface are related to the shape operator and the mean curvature through the following formulas:*

$$\mathrm{Hess}^M f(X) = \top(\mathrm{Hess}\, f(X)) + f_\nu \cdot S(X)$$
$$\Delta^M f = \Delta f - f_{\nu\nu} + f_\nu \cdot 2H$$

Proof.

$$\begin{aligned}\nabla^M_X(\nabla^M f) &= \nabla_X(\nabla^M f) - \langle \nabla_X(\nabla^M f), \nu\rangle \nu \\ &= \nabla_X(\nabla f - \langle \nabla f, \nu\rangle \nu) - c_1 \nu = \nabla_X \nabla f - \langle \nabla f, \nu\rangle \nabla_X \nu - c_2 \nu\end{aligned}$$

Since the left hand side belongs to T_pM, the $c_2\nu$ summand on the right hand side is the normal component of $\nabla_X \nabla f$. Thus we get

$$\mathrm{Hess}^M f(X) = \top(\mathrm{Hess}\, f(X)) - \langle \nabla f, \nu\rangle \nabla_X \nu = \top(\mathrm{Hess}\, f(X)) + \nabla_\nu f \cdot S(X)$$

Taking the trace of both sides, we obtain

$$\Delta^M f = \mathrm{tr}(\top \circ \mathrm{Hess}\, f) + \nabla_\nu f \cdot 2H$$

By computing the matrix of Hess f in an orthonormal basis (e_1, e_2, ν), one sees that

$$\mathrm{tr}(\top \circ \mathrm{Hess}\, f) = f_{e_1e_1} + f_{e_2e_2} = \mathrm{tr}(\mathrm{Hess}\, f) - f_{\nu\nu}$$

□

There are two special choices of a function f where the formulas of the above Theorem yield interesting results.

Corollary 3.9. *Let f be the distance function from M. Then $\Delta f = 2H$.*

Proof. Indeed, since $f_M = 0$, we have $\nabla^M f = 0$, and hence $\Delta^M = 0$. Also we have $f_\nu = 1$ and $f_{\nu\nu} = 0$. □

Corollary 3.10. *Consider the vector-valued function $i\colon M \to \mathbb{R}^3$, $i(p) = p$. The componentwise Laplacian of i equals the unit normal scaled by twice the mean curvature:*

$$\Delta^M i = 2H \cdot \nu$$

Proof. Each component of i is the restriction to M of a linear function on $\mathbb{R}^3$. If $f(x) = \langle a, x\rangle$, then we have

$$\nabla f = a, \quad \Delta f = 0, \quad f_\nu = \langle a, \nu\rangle, \quad f_{\nu\nu} = 0$$

Hence $\Delta^M f = \langle a, 2H\nu\rangle$. It follows that the Laplacian of the components of i consists of the components of $2H\nu$. □

4 Parallel surfaces

4.1 Curvature of parallel surfaces

Let $M \subset \mathbb{R}^3$ be an orientable smooth surface, and ν be a field of unit normals along M. By a *parallel surface* at distance $\delta \in \mathbb{R}$ we mean

$$M^\delta = \{p + \delta\nu \mid p \in M\}$$

We will show that for δ small enough M^δ is smooth and will compute its curvatures. Note that a parallel surface M^δ comes together with a natural map $M \to M^\delta$.

Theorem 4.1. *1. Let $C = \max_p\{|\kappa_1(p)|, |\kappa_2(p)|\}$ and assume that $|\delta| < \frac{1}{C}$. Then the surface M^δ is smooth.*

2. The vector $\nu(p)$ is a unit normal to M^δ at the point $p + \delta\nu$. That is, the tangent planes to M and M^δ at corresponding points are parallel.

3. The principal curvature directions of M^δ are parallel to the principal curvature directions of M at the corresponding point, and are given by

$$\kappa_i^\delta = \frac{\kappa_i}{1 - \delta\kappa_i}$$

Proof. Choose a surface patch $\sigma\colon U \to M$ compatibly oriented with ν (that is, the basis $(\sigma_u, \sigma_v, \nu)$ is positively oriented). This gives rise to a surface patch $\sigma^\delta = \sigma + \delta\nu$ of M^δ, and we have

$$\sigma_u^\delta = \sigma_u + \delta\nu_u, \quad \sigma_v^\delta = \sigma_v + \delta\nu_v$$

Since $\nu_u, \nu_v \in T_pM$, this shows already that $\nu^\delta = \nu$, provided that M^δ is smooth.

Now let σ be chosen so that $\sigma_u(p)$ and $\sigma_v(p)$ are principal directions at p (this can be achieved by a linear change of variables). This means

$$\nu_u = -\kappa_1 \sigma_u, \quad \nu_v = -\kappa_2 \sigma_v$$

at the point p. Therefore

$$\sigma_u^\delta = (1 - \delta\kappa_1)\sigma_u, \quad \sigma_v^\delta = (1 - \delta\kappa_2)\sigma_v$$

Hence, if $|\delta|$ is smaller than the maximum curvature radius, then the map σ^δ is regular, and the smooth surface M^δ has the same normals as M. This proves the first and the second part of the theorem.

For the third part, since $\nu_u^\delta = \nu_u$, we have from the above

$$\nu_u^\delta = -\frac{\kappa_1}{1 - \delta\kappa_1}\sigma_u^\delta, \quad \nu_v^\delta = -\frac{\kappa_2}{1 - \delta\kappa_2}\sigma_v^\delta$$

It follows that σ_u and σ_v are principal curvature directions, and the principal curvatures are as stated in the theorem. □

4.2 Delaunay surfaces

4.3 Weyl tube formula

Lemma 4.2. *The area of a parallel surface at distance δ is given by*

$$\text{area}(M^\delta) = \text{area}(M) - 2\delta \int_M H \text{ darea} + \delta^2 \int_M K \text{ darea}.$$

Here we assume that δ is small enough, so that M^δ is a smooth surface.

Proof. By Theorem 4.1 we have

$$\sigma_u^\delta \times \sigma_v^\delta = (1 - \delta\kappa_1)(1 - \delta\kappa_2)\sigma_u \times \sigma_v = (1 - 2\delta H + \delta^2 K)\sigma_u \times \sigma_v$$

Integrating over the parametrization domain U yields the stated formula. □

The domain between the parallel surfaces M^ε and $M^{-\varepsilon}$ is called a *tubular neighborhood* of M and denoted by $N_\varepsilon(M)$.

Theorem 4.3 (Weyl's tube formula)**.** *The volume of the tubular ε-neighborhood of a smooth surface M equals*

$$\text{vol}(N_\varepsilon(M)) = 2\varepsilon \,\text{area}(M) + \frac{2}{3}\varepsilon^3 \int_M K \text{ darea}$$

Proof. Consider the map

$$\Phi\colon U \times [-\varepsilon, \varepsilon] \to \mathbb{R}^3, \quad \Phi(u, v, \delta) = \sigma^\delta(u, v) = \sigma(u, v) + \delta\nu(u, v),$$

whose image is $N_\varepsilon(M)$. The Jacobi matrix of Φ is

$$(\sigma_u + \delta\nu_u, \sigma_v + \delta\nu_v, \nu),$$

so that

$$\det \Phi = \|(\sigma_u + t\nu_u) \times (\sigma_v + t\nu_v)\|$$

It follows that

$$\mathrm{vol}(N_\varepsilon(M)) = \int_{U\times[-\varepsilon,\varepsilon]} \|\sigma_u^\delta \times \sigma_v^\delta\| \, dudvd\delta = \int_{-\varepsilon}^{\varepsilon} \mathrm{area}(M^\delta) \, d\delta$$

Integration of the formula from Lemma 4.2 yields the formula of this theorem. □

4.4 Steiner formula

Let $M \subset \mathbb{R}^3$ be a smooth convex surface. Let ν be the interior unit normal field, so that the principal curvatures of M (and hence the mean curvature) are positive. Let $\varepsilon > 0$, and denote by $\overline{M}$ and $\overline{M^\varepsilon}$ the bodies bounded by M, and the exterior parallel surface $M^{-\varepsilon}$.

Theorem 4.4 (Steiner)**.** *The volume enclosed by a parallel surface is given by*

$$\mathrm{vol}(\overline{M^\varepsilon}) = \mathrm{vol}(\overline{M}) + \varepsilon \, \mathrm{area}(M) + \varepsilon^2 \int_M H \ \mathrm{darea} + \frac{4\pi}{3}\varepsilon^3$$

Proof. Similar to the proof of the Weyl tube formula we have

$$\mathrm{vol}(\overline{M^\varepsilon}) = \mathrm{vol}(\overline{M}) + \int_0^\varepsilon \mathrm{area}(M^{-\delta}) \, d\delta$$

Integrating the formula from Lemma 4.2 yields the result. □

5 Differential geometry of convex bodies

5.1 The support function

Similarly to the case of curves in the plane, we define the support function of a set $M \subset \mathbb{R}^3$ as

$$h\colon \mathbb{R}^3 \to \mathbb{R}, \quad h(v) = \max_{x \in M} \langle x, v \rangle$$

Clearly, h is positively homogeneous: $h(\lambda v) = \lambda h(v)$ for $\lambda \geq 0$. Therefore, the restriction of h to the unit sphere $\mathbb{S}^2 \subset \mathbb{R}^3$ gives the full information about it. Also, if M is a strictly smooth convex body, then $\max_{x \in M} \langle x, v \rangle$ is attained at the point with the unit normal positively proportional to v. It follows that we can write

$$h(v) = \langle \Gamma^{-1}(v/\|v\|), v \rangle, \tag{5.1}$$

where $\Gamma^{-1} \colon \mathbb{S}^2 \to M$ is the inverse of the Gauss map.

However, it is often convenient to view h as a function on $\mathbb{R}^3$. Then it is subadditive:

$$h(v + w) \leq h(v) + h(w),$$

which together with the positive homogeneity implies that h is convex.

Lemma 5.1. *A strictly convex surface is determined by its support function. Moreover, the Gauss parametrization of the surface can be written in terms of the support function as*

$$\Gamma^{-1}(v) = \operatorname{grad} h(v), \quad \|v\| = 1$$

Proof. Differentiate equation (5.1) in the direction of a vector $X \in \mathbb{R}^3$ using the shorthand notation $\Gamma^{-1}(v/\|v\|) = p(v)$:

$$\frac{\partial h(v)}{\partial X} = \langle \frac{\partial p(v)}{\partial X}, v \rangle + \langle p(v), X \rangle = \langle p(v), X \rangle,$$

because $\frac{\partial p(v)}{\partial X} \in T_{p(v)} M$ and hence is orthogonal to v. It follows that $\operatorname{grad} h(v) = p(v) = \Gamma^{-1}(v)$ for $\|v\| = 1$. □

Note that one can define the support function on any surface whose Gauss map is injective. This holds, for example, for a sufficiently small neighborhood of any point with non-vanishing Gaussian curvature.

5.2 Volume in terms of the support function

Lemma 5.2. *The volume of a body with smooth boundary is given by*

$$\operatorname{vol}(\overline{M}) = \frac{1}{3} \int_M h \operatorname{darea}$$

Proof. Consider the function $f \colon \mathbb{R}^3 \to \mathbb{R}$, $f(x) = \frac{\|x\|^2}{2}$. We have $\operatorname{grad} f(x) = x$, and $\Delta f = 3$. Therefore by the Stokes theorem we have

$$3 \operatorname{vol}(\overline{M}) = \int_{\overline{M}} \Delta f(x) \operatorname{darea} = \int_M \langle \operatorname{grad} f(p), \nu(p) \rangle \operatorname{darea} = \int_M h(p) \operatorname{darea}.$$

□

A similar method can be used to prove one of the Minkowski formulas.

Theorem 5.3. *For every body with smooth boundary M we have*

$$\int_M \nu(p)\ \mathrm{darea} = 0$$

Proof. Consider this time a linear function $f(x) = \langle a, x\rangle$ on $\mathbb{R}^3$. We have $\operatorname{grad} f(x) = a$ for all x, and $\Delta f = 0$. Therefore

$$0 = \int_{\overline{M}} \Delta f(x)\ \mathrm{darea} = \int_M \langle \operatorname{grad} f(p), \nu(p)\rangle\ \mathrm{darea} = \left\langle a, \int_M \nu(p)\ \mathrm{darea}\right\rangle$$

Since this holds for any $a \in \mathbb{R}^3$, it follows that the integral vanishes. □

A physical interpretation: the total pressure on the surface of a balloon is a zero vector.

5.3 Minkowski formulas

Theorem 5.4. *For every smooth convex surface M the following holds:*

$$\operatorname{area}(M) = \int_M hH\ \mathrm{darea}, \qquad \int_M H\ \mathrm{darea} = \int_M hK\ \mathrm{darea}$$

Proof.

$$\begin{aligned}\operatorname{vol}(\overline{M^\varepsilon}) &= \frac{1}{3}\int_{M^\varepsilon} h^\varepsilon\ \mathrm{darea} = \frac{1}{3}\int_M (h+\varepsilon)(1+\varepsilon\kappa_1)(1+\varepsilon\kappa_2)\ \mathrm{darea}\\ &= \frac{1}{3}\int_M \big(h + \varepsilon + \varepsilon h(\kappa_1+\kappa_2) + \varepsilon^2(\kappa_1+\kappa_2) + \varepsilon^2 h\kappa_1\kappa_2 + \varepsilon^3\kappa_1\kappa_2\big)\ \mathrm{darea}\end{aligned}$$
$$= \operatorname{vol}(\overline{M}) + \frac{\varepsilon}{3}\operatorname{area}(M) + \frac{2\varepsilon}{3}\int_M hH\ \mathrm{darea} + \frac{2\varepsilon^2}{3}\int_M H\ \mathrm{darea} + \frac{\varepsilon^2}{3}\int_M hK\ \mathrm{darea} + \frac{\varepsilon^3}{3}4\pi$$

Comparing this with the Steiner formula for $\operatorname{vol}(\overline{M^\varepsilon})$ we obtain the formulas of the theorem. □

Corollary 5.5. *The average width of a convex body with smooth boundary is $\frac{1}{2\pi}$ of the total mean curvature of its boundary:*

$$\frac{1}{4\pi}\int_{\mathbb{S}^2} w_v(M)\ \mathrm{darea} = \frac{1}{2\pi}\int_M H\ \mathrm{darea},$$

where $w_v(M) = h_v(M) + h_{-v}(M)$ is the width of M in the direction v.

Proof. Rewrite the first integral using the Gauss parametrization of M:

$$\frac{1}{2\pi}\int_{\mathbb{S}^2} h(v)\ \mathrm{darea} = \frac{1}{2\pi}\int_M hK\ \mathrm{darea}$$

and apply the second Minkowski formula. □

5.4 Spherical harmonics

5.5 Spherical Laplacian and the Poincaré-Wirtinger inequality

5.6 Support function and curvature

5.7 On the isoperimetric inequality

6 Minimal surfaces

6.1 Divergence, rate of the volume change, and the mean curvature

Let $X\colon \mathbb{R}^3 \to \mathbb{R}^3$ be a vector field in $\mathbb{R}^3$. The divergence of X is defined as

$$\operatorname{div} X = \operatorname{tr} \nabla X = \sum_{i=1}^{3} \frac{\partial X_i}{\partial x_i}$$

For every vector field there is an associated flow $\varphi_t\colon \mathbb{R}^3 \to \mathbb{R}^3$, $\varphi_0 = \mathrm{id}$, $\frac{d\varphi_t(p)}{dt} = X(\varphi_t(p))$.

Lemma 6.1. *The divergence of a vector field is equal to the rate of volume change under the flow associated to the vector field:*

$$\operatorname{div} X(p) = \left.\frac{d}{dt}\right|_{t=0} \det d\varphi_t(p)$$

Proof. We have

$$\varphi_t(p) = p + tX(p) + o(t)$$

Therefore

$$d\varphi_t(p) = \mathrm{id} + t\nabla X(p) + o(t),$$

which implies

$$\det d\varphi_t = 1 + t \operatorname{tr}(\nabla X) + o(t),$$

and the lemma is proved. □

Remark 6.2. Instead of the flow φ_t in the above lemma one can take any smooth family of diffeomorphisms φ_t such that $\left.\frac{d\varphi_t(p)}{dt}\right|_{t=0}(p) = X(p)$.

Now let $M \subset \mathbb{R}^3$ be a smooth surface, and $\varphi_t\colon M \to \mathbb{R}^3$ a smooth family of smooth maps such that $\varphi_0(p) = p$ for all $p \in M$. Our goal is to determine how the area of $\varphi_t(M)$ changes with t. The differential $d\varphi_t$ maps the plane T_pM to the plane $T_{\varphi_t(p)}\varphi_t(M)$, and the infinitesimal area change is equal to the determinant of this map (with respect to some orthogonal bases in both planes). At the same time, this determinant is the square root of the Gram determinant

$$\sqrt{\det(d\varphi_t^\top d\varphi_t)}.$$

It turns out that the infinitesimal rate of area change (that is the derivative of the above determinant) can be expressed in terms of the divergence of the velocity vector field, this time the divergence with respect to the surface M.

The initial velocity of φ_t defines a vector field

$$X\colon M \to \mathbb{R}^3, \quad X(p) = \left.\frac{d\varphi_t(p)}{dt}\right|_{t=0}(p)$$

(which is not necessarily tangent to M). Define the divergence of X with respect to M as

$$\operatorname{div}^M X = \operatorname{tr}(\nabla^M X) = \langle e_1, \nabla^M_{e_1} X\rangle + \langle e_2, \nabla^M_{e_2} X\rangle$$

where (e_1, e_2) is any orthonormal basis of T_pM.

Lemma 6.3. *For any smooth deformation of a surface the rate of the infinitesimal area change is equal to the divergence of the initial velocity vector field along the surface:*

$$\left.\frac{d}{dt}\right|_{t=0} \sqrt{\det(d\varphi_t^\top d\varphi_t)} = \operatorname{div}^M X$$

Proof. As in the proof of the previous lemma, we have

$$d\varphi_t(p) = \operatorname{id} + t\nabla X(p) + o(t),$$

which implies that

$$d\varphi_t(p)^\top d\varphi_t(p) = \big(\delta_{ij} + t(\langle e_i, \nabla_{e_j} X\rangle + \langle e_j, \nabla_{e_i} X\rangle) + o(t)\big)^2_{i,j=1}$$

Hence

$$\sqrt{\det(d\varphi_t^\top d\varphi_t)} = \sqrt{1 + 2t\operatorname{div}^M X + o(t)} = 1 + t\operatorname{div}^M X + o(t)$$

□

Theorem 6.4. *The first variation of the area under a smooth deformation of a surface is given by*

$$\left.\frac{d}{dt}\right|_{t=0} \operatorname{area}(\varphi_t(M)) = -2\int_M H\langle X, \nu\rangle \operatorname{darea} + \int_{\partial M} \langle X, \nu_{\partial M}\rangle\, d\ell$$

Here $\nu_{\partial M}$ is the exterior unit normal to ∂M, tangent to M.

Proof. By Lemma 6.3 we have

$$\left.\frac{d}{dt}\right|_{t=0} \operatorname{area}(\varphi_t(M)) = \int_M \operatorname{div}^M X \operatorname{darea} = \int_M \operatorname{div}^M(X_\top) \operatorname{darea} + \int_M \operatorname{div}^M(X_\perp) \operatorname{darea},$$

where $X_\top$ and $X_\perp$ are the tangential and the normal components of X. For the tangential component we have

$$\int_M \operatorname{div}^M(X_\top) \operatorname{darea} = \int_{\partial M} \langle X_\top, \nu_{\partial M}\rangle \operatorname{darea} = \int_{\partial M} \langle X, \nu_{\partial M}\rangle \operatorname{darea},$$

and for the normal component $X^\perp = f\nu$ we have

$$\operatorname{div}^M(f\nu) = \langle e_1, \nabla_{e_1}(f\nu)\rangle + \langle e_2, \nabla_{e_2}(f\nu)\rangle = f\langle e_1, \nabla_{e_1}\nu\rangle + f\langle e_2, \nabla_{e_2}\nu\rangle = -f\operatorname{tr} II = -2H\langle X, \nu\rangle$$

□

6.2 Definition and examples of minimal surfaces

If we consider only deformations vanishing on ∂M, then

$$\frac{d}{dt}\bigg|_{t=0} \operatorname{area}(\varphi_t(M)) = -2\int_M H\langle X, \nu\rangle \,\mathrm{darea}$$

It follows that $H = 0$ is a necessary and sufficient condition for the area of M to have zero derivative under all deformations vanishing on the boundary.

Definition 6.5. *A surface $M \subset \mathbb{R}^3$ is called* minimal*, if its mean curvature vanishes everywhere, that is, if M is a critical point of the area functional.*

Example 6.6. The catenoid $\sigma(t,\theta) = (\cosh t\cos\theta, \cosh t\sin\theta, t)$. The Scherk surface $z = \log\frac{\cos y}{\cos x}$.

6.3 Conformal harmonic parametrizations

The following theorem gives an important characterization of minimal surfaces. In particular, it will be used in finding the Weierstrass representation of minimal surfaces.

Theorem 6.7. *If $\sigma\colon U \to \mathbb{R}^3$ is conformal and (componentwise) harmonic, then $\sigma(U)$ is a minimal surface. Conversely, every minimal surface can be locally parametrized by conformal harmonic surface patches.*

Example 6.8. The Enneper minimal surface is

$$\sigma(u,v) = \left(u - \frac{u^3}{3} + uv^2, v - \frac{v^3}{3} + u^2v, u^2 - v^2\right)$$

The components of σ are harmonic polynomials. Also we have

$$\|\sigma_u\| = \|\sigma_v\| = \sqrt{1 + u^2 + v^2}, \quad \langle\sigma_u, \sigma_v\rangle = 0$$

Thus σ is conformal and harmonic, and its image is a minimal surface.

A first approximation to Theorem 6.7 is the following

Lemma 6.9. *A surface $M \subset \mathbb{R}^3$ is minimal if and only if the inclusion map $i\colon M \to \mathbb{R}^3$ is harmonic: $\Delta^M i = 0$.*

Proof. Follows directly from Corollary 3.10. □

Two more lemmas are needed.

Lemma 6.10. *Let $\sigma\colon U \to M$ be a conformal surface patch, and let $f\colon M \to \mathbb{R}$. Then*

$$\Delta^M f = 0 \Leftrightarrow \Delta(f\circ\sigma) = 0.$$

This lemma will be proved in the next subsection.

Lemma 6.11. *Every surface has an atlas consisting of conformal surface patches.*

The proof is omitted.

Proof of Theorem 6.7. Assume that $\sigma\colon U \to \mathbb{R}^3$ is conformal and componentwise harmonic: $\Delta(\mathrm{pr}_\alpha \circ \sigma) = 0$ for $\alpha = 1, 2, 3$. Put $M = \sigma(U)$. By Lemma 6.10 we have $\Delta^M \mathrm{pr}_\alpha = 0$. Hence by Lemma 6.9 the surface M has zero mean curvature.

In the opposite direction, assume $M \subset \mathbb{R}^3$ is minimal. Then, reversing the above argument, we see that every conformal surface patch of M is harmonic. By Lemma 6.11 there is a conformal surface patch $\sigma\colon U \to M$ around every point. Thus M has a conformal harmonic atlas. □

6.4 Laplacian and conformal maps

Our goal is to prove Lemma 6.10. Let Δ^σ be an operator on $C^\infty(U)$ corresponding to the surface Laplacian:

$$\Delta^\sigma f = \Delta^M (f \circ \sigma^{-1}) \circ f \quad \text{for every } f\colon U \to \mathbb{R}$$

We want to show that for a conformal parametrization σ we have $\Delta f = 0 \Leftrightarrow \Delta^\sigma f = 0$.

Lemma 6.12. *Let $I_\sigma = \lambda(du^2 + dv^2)$. Then we have*

$$\nabla^\sigma f = \lambda^{-1} \nabla f$$
$$\operatorname{div}^\sigma X = \operatorname{div} X + \lambda^{-1} \frac{\partial \lambda}{\partial X}$$

Proof. The first identity follows from

$$\langle \nabla f, X \rangle = I_\sigma(\nabla^\sigma f, X) = \lambda \langle \nabla^\sigma f, X \rangle$$

for all $X \in T_pU = \mathbb{R}^2$.

For the second one, note that the area element of I_σ is λ times the area element of the euclidean metric. Therefore the infinitesimal area change under φ_t with respect to I_σ is

$$\frac{\lambda(\varphi_t(p))}{\lambda(p)} \det d\varphi_t$$

□

Theorem 6.13. *If $I_\sigma = \lambda(du^2 + dv^2)$, then*

$$\Delta^\sigma f = \lambda^{-1} \Delta f$$

Proof. □

6.5 Adjoint minimal surface

From Theorem 6.7 we know that every minimal surface has a conformal harmonic parametrization $\sigma\colon U \to \mathbb{R}^3$. From now on we will assume that U is simply connected.

The harmonicity means that every component of

$$\sigma(u,v) = (x(u,v), y(u,v), z(u,v))$$

has zero Laplacian. Therefore the differential form $x_u dv - x_v du$ is closed; by integrating it (here we use that U is simply-connected) we obtain a function $x^*\colon U \to \mathbb{R}$ such that

$$x^*_u = -x_v, \quad x^*_v = x_u$$

It follows that x^* is also harmonic; it is called the harmonic function *conjugate to* x.

Note that $x + ix^*\colon U \to \mathbb{C}$ is a holomorphic function, with respect to the complex coordinate $u + iv$. We will return to the tools from the complex analysis later.

Definition 6.14. *Let $\sigma = (x, y, z)$ be a conformal harmonic parametrization of a minimal surface. The map $\sigma^*\colon U \to \mathbb{R}^3$, $\sigma^* = (x^*, y^*, z^*)$, whose components are the harmonic functions conjugate to the components of σ, is called the* adjoint *of σ.*

Theorem 6.15. *Let σ^* be the adjoint of a minimal surface σ. Then it has the following properties.*

1. *It is also a minimal surface.*
2. *The map $\sigma(u,v) \mapsto \sigma^*(u,v)$ is an isometry between the surface and its adjoint.*
3. *The tangent planes to $\sigma(U)$ and $\sigma^*(U)$ at the corresponding points are parallel.*

Proof. The components of σ^* are harmonic functions by construction. The first fundamental form of σ^* has the coefficients

$$\|\sigma^*_u\|^2 = \|\sigma_v\|^2 = \|\sigma_u\|^2 = \|\sigma^*_v\|^2, \quad \langle \sigma^*_u, \sigma^*_v \rangle = -\langle \sigma_v, \sigma_u \rangle = 0,$$

hence σ^* is conformal, hence $\sigma^*(U)$ is a minimal surface.

The above calculation of the coefficients of the first fundamental form shows that $\sigma(u,v) \mapsto \sigma^*(u,v)$ is an isometry. Also obviously $\operatorname{span}\{\sigma^*_u, \sigma^*_v\} = \operatorname{span}\{\sigma_u, \sigma_v\}$. □

Example 6.16. The adjoint of the catenoid

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

is the helicoid

$$\sigma^*(u,v) = (\sinh u \sin v, -\sinh u \cos v, v)$$

Exercise 6.17. Find the adjoint of the Enneper surface

$$\sigma(u,v) = \left(u - \frac{u^3}{3} + uv^2, v - \frac{v^3}{3} + u^2 v, u^2 - v^2\right)$$

Show that the adjoint is actually congruent to the Enneper surface.

6.6 Associate family of minimal surfaces

Definition 6.18. *Let $\sigma\colon U \to \mathbb{R}^3$ be a conformal harmonic parametrization of a minimal surface, and let $\sigma^*\colon U \to \mathbb{R}$ be its adjoint. The* associate family *of σ is the family of maps*

$$\sigma_\theta\colon U \to \mathbb{R}^3, \quad \sigma_\theta = \sigma\cos\theta + \sigma^*\sin\theta, \quad \theta \in [0, 2\pi)$$

In particular, $\sigma_{\frac{\pi}{2}} = \sigma^*$.

Lemma 6.19. *All surfaces of the associate family are minimal surfaces isometric to σ. Their tangent planes are parallel to the tangent planes to σ at the corresponding points.*

Proof. It can be checked by a direct computation, using the already known properties of σ^*, that σ_θ is conformal and harmonic, and has the same first fundamental form as σ.

Alternatively, note that

$$\sigma_\theta = \mathrm{Re}(e^{-i\theta}F),$$

where $F = \sigma + i\sigma^*$. The components of $e^{-i\theta}F$ are holomorphic functions, therefore σ_θ is harmonic. Conformality and isometry can be proved in a similar way, see the next section.

The tangent planes are parallel, because $(\sigma_\theta)_u$ and $(\sigma_\theta)_v$ are linear combinations of $\sigma_u, \sigma_v, \sigma^*_u = -\sigma_v$, and $\sigma^*_v = \sigma_u$. □

6.7 Minimal surfaces and isotropic holomorphic curves in $\mathbb{C}^3$

The harmonicity of σ is equivalent to $\sigma = \mathrm{Re}(F)$ for a holomorphic curve $F\colon U \to \mathbb{C}^3$. How to express the conformality of σ in terms of F?

Extend the scalar product on $\mathbb{R}^3$ to a scalar product on $\mathbb{C}^3$:

$$\langle z, w\rangle = z_1w_1 + z_2w_2 + z_3w_3$$

(note: this is not the hermitian inner product!). If $z = x + iy$ with $x, y \in \mathbb{R}^3$, then we have

$$\|z\|^2 = \|x\|^2 - \|y\|^2 + 2i\langle x, y\rangle$$

A vector $z \in \mathbb{C}^3$ is called *isotropic*, if $\|z\|^2 = 0$.

Lemma 6.20. *A harmonic map $\sigma\colon U \to \mathbb{R}^3$ is conformal if and only if the corresponding holomorphic curve*

$$F = \sigma + i\sigma^*$$

has isotropic derivative:

$$\|F'\|^2 = 0$$

Proof. We have

$$F' = \sigma_u - i\sigma_v$$

Hence

$$\|F'\|^2 = 0 \Leftrightarrow \|\sigma_u\|^2 - \|\sigma_v\|^2 = 0 \text{ and } \langle\sigma_u, \sigma_v\rangle = 0$$

□

Since U is simply connected, every holomorphic map $F\colon U \to \mathbb{C}^3$ is a componentwise integral of a holomorphic map $f\colon U \to \mathbb{C}^3$. Thus we arrive to the following conclusion.

Lemma 6.21. *Let $f = (f_1, f_2, f_3)\colon U \to \mathbb{C}^3$ be an isotropic holomorphic curve: $f_1^2 + f_2^2 + f_3^2 = 0$. Then the map*

$$\sigma(w) = \sigma_0 + \mathrm{Re} \int_{w_0}^{w} f(w)\, dw, \quad w = u + iv, \sigma_0 \in \mathbb{R}^3$$

determines a minimal surface, and

$$\sigma(w^*) = \sigma_0^* + \mathrm{Im} \int_{w_0}^{w} f(w)\, dw, \quad \sigma_0^* \in \mathbb{R}^3$$

its adjoint. Conversely, every simply connected minimal surface has a parametrization σ that arises from some isotropic holomorphic curve in this way.

6.8 Enneper-Weierstrass representation

There are different ways to solve the equation $f_1^2+f_2^2+f_3^2 = 0$. For example, we can take f_1 and f_2 to be arbitrary holomorphic functions and define $f_3 = \sqrt{f_1^2 + f_2^2}$ (choosing some branch of the square root). The following parametrization of the set of solutions has a nice geometric meaning, as will be shown later.

Lemma 6.22. *Let $\varphi\colon U \to \mathbb{C}$ be a holomorphic function, and $\psi\colon U \to \mathbb{C} \cup \{\infty\}$ be a meromorphic function such that $\varphi \not\equiv 0$ and every point which is a pole of order n for ψ is a zero of order $\geq 2n$ for φ. Then the functions*

$$f_1 = \frac{1}{2}\varphi(1 - \psi^2), \quad f_2 = \frac{i}{2}\varphi(1 + \psi^2), \quad f_3 = \varphi\psi$$

are holomorphic in U and satisfy $f_1^2 + f_2^2 + f_3^2 = 0$. Conversely, every $f = (f_1, f_2, f_3)$ such that $f_1^2 + f_2^2 + f_3^2 = 0$, $f \not\equiv 0$, $f_1 - if_2 \not\equiv 0$ can be written in this form.

Proof. The first part is obvious. For the second part, put

$$\varphi = f_1 - if_2, \quad \psi = \frac{f_3}{f_1 - if_2}.$$

□

Note that $f_1 = if_2$ would imply $f_3 = 0$, so that the surface would lie in a plane.

This yields the *Enneper-Weierstrass representation* of non-planar minimal surfaces. Take any φ and ψ satisfying the above conditions and put

$$\begin{aligned} x(w) &= x_0 + \mathrm{Re} \int_{w_0}^{w} \frac{1}{2}\varphi(1 - \psi^2)\, dw \\ y(w) &= y_0 + \mathrm{Re} \int_{w_0}^{w} \frac{i}{2}\varphi(1 + \psi^2)\, dw \\ z(w) &= z_0 + \mathrm{Re} \int_{w_0}^{w} \varphi\psi\, dw \end{aligned}$$

Theorem 6.23. *In the Enneper-Weierstrass representation the map* $\psi \colon U \to \mathbb{C} \cup \{0\}$ *is the composition of the Gauss map with the stereographic projection:*

$$\psi = \pi \circ \Gamma \circ \sigma, \quad U \xrightarrow{\sigma} M \xrightarrow{\Gamma} \mathbb{S}^2 \xrightarrow{\pi} \mathbb{C} \cup \{0\}$$

Proof. We have

$$\sigma_u - i\sigma_v = F = \left(\frac{1}{2}\varphi(1-\psi^2), \frac{i}{2}\varphi(1+\psi^2), \varphi\psi\right)$$

From this we compute

$$\Gamma(\sigma(w)) = \frac{1}{1+|\psi|^2}(2\operatorname{Re}\psi, 2\operatorname{Im}\psi, |\psi|^2 - 1)$$

Now the theorem follows from the formula for the inverse of the stereographic projection:

$$\pi^{-1}(x, y) = \frac{1}{1+x^2+y^2}(2x, 2y, x^2+y^2-1)$$

□

Chapter 6

Geodesics

1 Definition and variational properties

1.1 Curves of vanishing geodesic curvature

Let $\gamma\colon I \to M \subset \mathbb{R}^3$ be a smooth curve on a smooth surface M. Decompose its acceleration vector into the tangential and normal components:

$$\ddot{\gamma} = \ddot{\gamma}_\top + \ddot{\gamma}_\perp.$$

By definition of the geodesic and normal curvatures (see Section 1.2, the geodesic curvature equals the norm of the tangential component of the acceleration vector:

$$\kappa_g = \|\ddot{\gamma}_\top\|$$

Definition 1.1. *A curve on a surface is called a (parametrized)* geodesic*, if its geodesic curvature vanishes, that is, if its acceleration vector is orthogonal to the surface.*

Also recall (although we will not need this) that if $\ddot{\gamma} = c\nu$, then $c = II(\dot{\gamma}, \dot{\gamma})$.

Lemma 1.2. *Geodesics have constant speed.*

Proof.

$$\frac{d}{dt}\|\dot{\gamma}\|^2 = 2\langle\dot{\gamma}, \ddot{\gamma}\rangle = 2\langle\dot{\gamma}, \ddot{\gamma}_\top\rangle = 0$$

□

Let $t = \varphi(s)$ be a parameter change along the geodesic. The curve $\tilde{\gamma} = \gamma \circ \varphi$ will be called a *reparametrized geodesic.*

Lemma 1.3. *A curve is a reparametrized geodesic if and only if the tangential component of its acceleration vector is parallel to the velocity vector.*

Proof. For a reparametrization φ of a curve γ we have

$$\frac{d^2\tilde{\gamma}}{ds^2} = \frac{d^2\varphi}{ds^2}\dot{\gamma} + \left(\frac{d\varphi}{ds}\right)^2 \ddot{\gamma}.$$

Therefore, if γ is a geodesic, then the tangential component of $\ddot{\gamma}$ vanishes, and the tangential component of $\frac{d^2\tilde{\gamma}}{ds^2}$ is parallel to $\frac{d\tilde{\gamma}}{ds}$. Conversely, if the tangential component of $\frac{d^2\tilde{\gamma}}{ds^2}$ is parallel to $\frac{d\tilde{\gamma}}{ds}$, then the function $\psi = \varphi^{-1}$ can be found by solving the differential equation of the form

$$\frac{d^2\psi}{dt^2} \cdot f + \left(\frac{d\psi}{dt}\right)^2 \cdot g = 0.$$

□

Thus, you are moving along a geodesic if and only if your acceleration has only a normal component and a component collinear with the direction of your motion.

Note that the only reparametrizations of a geodesic that leave the acceleration vector orthogonal to the surface are the linear ones: $s = ct$, $c = \text{const}$.

In general, it is a difficult problem to describe geodesics on a surface explicitely (and we will see in the next section, why). Below are some special cases, where we can assert that a given curve is a geodesic.

Lemma 1.4. *Any straight line segment on a surface is a geodesic.*

Proof. Indeed, for a parametrization of the form $\gamma(t) = a+bt$ we have $\ddot{\gamma} = 0$, in particular the tangential component of $\ddot{\gamma}$ is zero. □

Lemma 1.5. *Any normal section of a surface is a geodesic.*

Proof. Let $L \subset \mathbb{R}^3$ be a plane orthogonal to M at every intersection point. Then L is the osculating plane of the intersection curve. It follows that for every parametrization γ of the intersection curve the acceleration vector $\ddot{\gamma}$ is contained in L. Hence the tangential component of $\ddot{\gamma}$ is parallel to $\dot{\gamma}$, and γ is a reparametrized geodesic. □

Example 1.6. The straight lines on the hyperboloid of one sheet and on the hyperbolic paraboloid are geodesics.

The meridians of surfaces of revolution are geodesics. The parallels at the critical values of the radius are geodesics.

The big circles on the sphere are geodesics.

1.2 Geodesic equations and their consequences

Theorem 1.7. *Let $\sigma\colon U \to M$ be a regular surface patch. A curve $\gamma(t) = \sigma(u(t), v(t))$ is a geodesic if and only if the functions $u(t)$ and $v(t)$ satisfy a system of differential equations*

$$\ddot{u} + P(\dot{u}, \dot{v}) = 0, \quad \ddot{v} + Q(\dot{u}, \dot{v}) = 0,$$

where P and Q are symmetric bilinear forms whose coefficients depend on the coefficients of the first fundamental form and their first derivatives.

Lemma 1.8. *A curve $\gamma = \sigma(u(t), v(t))$ is a geodesic if and only if*

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

Proof. By definition, γ is a geodesic if and only if the vector $\ddot{\gamma}$ is orthogonal to M, that is

$$\langle\ddot{\gamma}, \sigma_u\rangle = 0 = \langle\ddot{\gamma}, \sigma_v\rangle$$

We have

$$\begin{aligned}\langle\ddot{\gamma}, \sigma_u\rangle &= \frac{d}{dt}\langle\dot{\gamma}, \sigma_u\rangle - \left\langle\dot{\gamma}, \frac{d}{dt}\sigma_u\right\rangle = \frac{d}{dt}(\dot{u}\sigma_u + \dot{v}\sigma_v, \sigma_u) - \langle\dot{u}\sigma_u + \dot{v}\sigma_v, \dot{u}\sigma_{uu} + \dot{v}\sigma_{uv}\rangle \\ &= \frac{d}{dt}(E\dot{u} + F\dot{v}) - (\langle\sigma_u, \sigma_{uu}\rangle\dot{u}^2 + \langle\sigma_u, \sigma_{uv}\rangle\dot{u}\dot{v} + \langle\sigma_v, \sigma_{uu}\rangle\dot{u}\dot{v} + \langle\sigma_v, \sigma_{uv}\rangle\dot{v}^2)\end{aligned}$$

On the other hand,

$$E_u = \frac{\partial}{\partial u}\langle\sigma_u, \sigma_u\rangle = 2\langle\sigma_u, \sigma_{uu}\rangle$$
$$F_u = \frac{\partial}{\partial u}\langle\sigma_u, \sigma_v\rangle = \langle\sigma_{uu}, \sigma_v\rangle + \langle\sigma_u, \sigma_{uv}\rangle$$
$$G_u = \frac{\partial}{\partial u}\langle\sigma_v, \sigma_v\rangle = 2\langle\sigma_v, \sigma_{uv}\rangle$$

Substituting this into the above formula proves the first equation of the lemma. The second equation is proved in a similar way. □

Proof of the theorem. Transform the equations of the lemma. We have

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

This yields the equations

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

Solving this as a linear system with respect to $\ddot{u}$ and $\ddot{v}$, we express $\ddot{u}$ and $\ddot{v}$ as functions of $\dot{u}$, $\dot{v}$, namely as quadratic forms with coefficients depending on $E, F, G, E_u, \ldots$. □

The equations in Theorem 1.7 are non-linear and there is no universal method to solve them. However, the existence and uniqueness theorem from the theory of ordinary differential equations implies the following.

Corollary 1.9. *Let $p \in M$ be a point on a smooth surface, and $X \in T_pM$ be a tangent vector at p. Then there is a unique geodesic starting at p with the initial velocity X.*

Interestingly enough, equations of Theorem 1.7 involve only the coefficients of the first fundamental form, although the definition of a geodesic uses the way the surface is situated in the space. This implies the following.

Corollary 1.10. *An isometry between two surfaces takes the geodesics on one surface to the geodesics on the other.*

Example 1.11. Geodesics on a developable surface become straight lines, when the surface is developed onto the plane. In particular, every geodesic on a circular cylinder is either a meridian, or a parallel, or a helix.

The meridians and the shortest parallel on the catenoid are geodesics by Lemma 1.5. The isometry between the catenoid and the helicoid maps these curves to the straight lines (the axis and the generatrices perpendicular to the axis), which are geodesics by Lemma 1.4. Less trivially, the images of these curve on every surface from the associate family are also geodesics.

1.3 The energy and the length functionals

Let $\gamma\colon [a,b] \to M$ be a curve on the surface M. A smooth deformation of γ within M is a family of curves $\gamma^s\colon [a,b] \to M$, $s \in (-\varepsilon, \varepsilon)$, such that $\gamma^0(t) = \gamma(t)$, and such that $[a,b] \times (-\varepsilon, \varepsilon) \to M \subset \mathbb{R}^3$ is a smooth map.

An infinitesimal deformation of γ is a vector field $X\colon [a,b] \to TM$ such that $X(t) \in T_{\gamma(t)}M$. This is also called a vector field along γ. The derivative (with respect to s) of a smooth deformation is an infinitesimal deformation. Conversely, every infinitesimal deformation is the derivative of some (actually of infinitely many) smooth deformation.

Definition 1.12. *Let $\gamma\colon [a,b] \to \mathbb{R}^n$ be a smooth curve. The* energy *of γ is the integral*

$$E(\gamma) = \frac{1}{2}\int_a^b \|\dot\gamma\|^2\,dt$$

Theorem 1.13. *Let γ^s be a smooth deformation of a curve $\gamma\colon [a,b] \to M$, and let X be a vector field along γ, which is the derivative of γ^s. Then we have*

$$\left.\frac{d}{ds}\right|_{s=0} E(\gamma^s) = \langle \dot\gamma(b), X(b)\rangle - \langle \dot\gamma(a), X(a)\rangle - \int_a^b \langle \ddot\gamma, X\rangle\,dt$$

Proof. We have $\gamma^s(t) = \gamma(t) + sX(t) + o(s)$. Therefore the velocity vector of γ^s satisfies

$$\dot\gamma^s = \dot\gamma + s\dot X + o(s) \Rightarrow \|\dot\gamma^s\|^2 = \|\dot\gamma\|^2 + 2s\langle \dot\gamma, \dot X\rangle + o(s).$$

Differentiating under the integral sign and using integration by parts we obtain

$$\left.\frac{d}{ds}\right|_{s=0} E(\gamma^s) = \int_a^b \langle \dot\gamma, \dot X\rangle\,dt = \langle \dot\gamma, X\rangle|_a^b - \int_a^b \langle \ddot\gamma, X\rangle\,dt$$

and the theorem is proved. □

Corollary 1.14. *A curve $\gamma\colon [a,b] \to M \subset \mathbb{R}^3$ is a (constant speed parametrized) geodesic if and only if it is a critical point of the energy functional with respect to deformations with fixed endpoints.*

Proof. If $\gamma^s(a) = \gamma(a)$ and $\gamma^s(b) = \gamma(b)$, then the formula from Theorem 1.13 becomes

$$\frac{d}{ds}\bigg|_{s=0} E(\gamma^s) = -\int_a^b \langle \ddot{\gamma}, X \rangle \, dt$$

If γ is a geodesic, then $\ddot{\gamma}$ is orthogonal to M, and the integrand on the right hand side vanishes. Hence, geodesics are critical points. Conversely, if a curve is a critical point of the energy functional, then the integrand on the right hand side must vanish for all vector fields X. This implies that $\ddot{\gamma}$ is orthogonal to M. □

As usual, by $L(\gamma)$ we denote the length of γ.

Theorem 1.15. *Let γ^s be a smooth deformation of a curve $\gamma\colon [a,b] \to M$, and let X be a vector field along γ, which is the derivative of γ^s. Then we have*

$$\frac{d}{ds}\bigg|_{s=0} L(\gamma^s) = \langle \tau(b), X(b) \rangle - \langle \tau(a), X(a) \rangle - \int_a^b \kappa \|\dot{\gamma}\| \langle \nu_\gamma, X \rangle \, dt,$$

where $\tau(t) = \frac{\dot{\gamma}(t)}{\|\dot{\gamma}(t)\|}$ is the unit tangent vector to γ, κ is the curvature of γ, and ν_γ is the unit normal to γ.

Proof. We compute

$$\frac{d}{ds}\bigg|_{s=0} \|\dot{\gamma}^s\| = \langle \tau, \dot{X} \rangle$$

Integration by parts yields

$$\frac{d}{ds}\bigg|_{s=0} L(\gamma^s) = \langle \tau, \dot{X} \rangle \bigg|_a^b - \int_a^b \langle \dot{\tau}, X \rangle.$$

By definition of the curvature of curves, $\frac{\dot{\tau}}{\|\dot{\gamma}\|} = \kappa \nu_\gamma$, and the theorem follows. □

Corollary 1.16. *A curve $\gamma\colon [a,b] \to M \subset \mathbb{R}^3$ is a reparametrized geodesic if and only if it is a critical point of the length functional with respect to deformations with fixed endpoints.*

Proof. If γ is a reparametrized geodesic, then the osculating plane of γ is perpendicular to M at every point. This implies that ν_γ is orthogonal to M, and therefore the integrand in the first variation formula vanishes.

Conversely, if the integrand vanishes for all X tangent to M, then the osculating planes are perpendicular to M. Since $\ddot{\gamma}$ lies in the osculating plane, its tangential component is parallel to $\dot{\gamma}$. Hence γ is a reparametrized geodesic. □

2 Geodesic coordinates

2.1 Geodesic polar coordinates

Choose a point $p_0 \in M$ and a unit vector $e_0 \in T_{p_0}M$. For every $\theta \in [0, 2\pi)$ denote by e_θ the unit vector that forms the angle θ with e_0 (we need to orient M in order to choose the direction in which θ increases). Let $\gamma^\theta \colon [0, R] \to M$ be the unit speed geodesic with the initial conditions

$$\gamma^\theta(0) = p_0, \quad \dot\gamma^\theta(0) = e_\theta.$$

This defines the map

$$\sigma \colon [0, 2\pi) \times [0, R] \to M, \quad \sigma(r, \theta) = \gamma^\theta(r)$$

Theorem 2.1. *For every (r_0, θ_0) with $r_0 > 0$ sufficiently small there is an open set $U \ni (r_0, \theta_0)$ such that $\sigma \colon U \to M$ is a regular surface patch.*

The first fundamental form of σ has the form

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

for some positive smooth function $G(r, \theta)$.

Proof. The first part follows from the inverse function theorem provided that the vectors $\sigma_r(r_0, \theta_0)$ and $\sigma_\theta(r_0, \theta_0)$ are linearly independent. For small r_0 these vectors are close to the coordinate vectors of the polar coordinate system in the plane $T_{p_0}M$.

For the second part note first that σ_r is the velocity vector of a unit speed geodesic γ^θ, hence $\|\sigma_r\|^2 = 1$. Further, since all curves $\gamma^\theta \colon [0, r_0] \to M$ have the same energy $\frac{r_0}{2}$, the first variation formula for the energy yields

$$0 = \langle \frac{d}{dr}\gamma^\theta, \frac{d}{d\theta}\gamma^\theta \rangle = \langle \sigma_r, \sigma_\theta \rangle$$

(we consider the deformation $\gamma^{\theta+s}$ of the curve γ^θ). Thus the coefficients of the first fundamental form satisfy $E = 1$, $F = 0$, and the theorem is proved. □

The coordinates (r, θ) are called the *geodesic polar coordinates* on M. Sending a point with polar coordinates (r, θ) on $T_{p_0}M$ to the point with the same geodesic polar coordinates on M defines a map

$$T_{p_0}M \to M,$$

called the *exponential map.*

2.2 Local length minimization

Let $\gamma \colon [a, b] \to M$ be a geodesic. In general, γ is not the shortest curve between $\gamma(a)$ and $\gamma(b)$. Example: two points on the cylinder not lying on the same meridian can be joined by infinitely many geodesics, so most of them are not the shortest paths. However, we prove in this section that geodesics are locally the shortest.

Theorem 2.2. *Let $p_0 \in M$, and let $R > 0$ be such that the map $\sigma\colon (0, R] \times [0, 2\pi) \to M$ defined in the previous section is injective. Then for every $(r_0, \theta_0) \in (0, R] \times [0, 2\pi)$ the geodesic*

$$\gamma\colon [0, r_0] \to M, \quad \gamma(t) = \sigma(t, \theta_0)$$

is the shortest among all curves joining p_0 with $\sigma(r_0, \theta_0)$.

Proof. The geodesic γ has length r_0. Let us show that every other curve δ from p_0 to p has length greater than r_0. Assume first that δ does not leave the set $W = \sigma\left((0, R) \times [0, 2\pi)\right)$. Then

$$\delta(t) = \sigma(r(t), \theta(t)), \quad r(a) = 0, r(b) = r_0$$

We have

$$\dot{\delta} = \dot{r}\sigma_r + \dot{\theta}\sigma_\theta \Rightarrow \|\dot{\delta}\|^2 = \dot{r}^2 + G(r, \theta)\dot{\theta}^2 \geq \dot{r}^2.$$

Therefore

$$L(\delta) = \int_a^b \|\dot{\delta}\|\, dt \geq \int_a^b \|\dot{r}\|\, dt = r_0$$

The equality takes place only if $\dot{\theta} = 0$, that is only if δ is a geodesic. Now, if δ leaves the set W, then it has already length R at the moment when its r-coordinate attains the value R, in particular its total length is bigger than r_0. □

Corollary 2.3. *Let $\gamma\colon [a, b] \to M$ be a geodesic. Then for all $t_0 < t_1$ sufficiently close to each other, the arc of γ between t_0 and t_1 is the shortest among all curves joining $\gamma(t_0)$ with $\gamma(t_1)$.*

Proof. Put $p_0 = \gamma(t_0)$ and use the theorem. □

2.3 Geodesic normal coordinates

Let $\delta\colon [-S, S] \to M$ be an arbitrary smooth curve on the surface M. Choose a field of unit normals ν_δ to δ tangent to M. For every $s \in [-S, S]$ construct a geodesic $\gamma^s\colon [-T, T] \to M$ with the initial conditions

$$\gamma^s(0) = \delta(s), \quad \dot{\gamma}^s(0) = \nu_\delta.$$

Then the map

$$\sigma\colon [-S, S] \times [-T, T] \to M, \quad \sigma(s, t) = \gamma^s(t)$$

has a non-degenerate Jacobi matrix at the point $(0, 0)$ (indeed, the vectors $\sigma_s(0, 0) = \frac{d}{ds}\big|_{s=0} \delta$ and $\sigma_t(0, 0) = \dot{\gamma}^0(0) = \nu_\delta(0)$ are non-zero and orthogonal to each other). Hence for S and T sufficiently small it is a diffeomorphism onto the image, that is a regular surface patch for M.

Definition 2.4. *The coordinates (s, t) on M defined above are called* geodesic normal coordinates *with respect to the curve δ.*

Theorem 2.5. *The first fundamental form with respect to geodesic normal coordinates has the form*

$$dt^2 + G(s,t)ds^2$$

for some positive smooth function $G(s,t)$.

Proof. Similar to the proof of Theorem 2.1. □

This theorem implies that every coordinate curve $s = \text{const}$ is perpendicular to the geodesics $t = \text{const}$. Hence, the geodesic normal coordinates with respect to the curve $s = \text{const}$ are the same, differing only by a constant shift in s.

Example 2.6. For the surface of revolution $(f(t)\cos\theta, f(t)\sin\theta, g(t))$ with the profile curve $(f(t), g(t))$ of unit speed, the first fundamental form is

$$dt^2 + f^2(t)d\theta^2$$

The coordinates (θ, t) are geodesic normal coordinates with respect to the parallel $\theta = 0$.

Exercise 2.7. Prove a converse of the above theorem: if the first fundamental form of a surface patch $\sigma(s,t)$ has the form $dt^2 + G(s,t)ds^2$, then (s,t) are geodesic polar coordinates with respect to the curve $s = 0$.

Exercise 2.8. Assume that a surface patch $\sigma(s,t)$ has the first fundamental form $dt^2 + f^2(t)d\theta^2$. Does the image of σ necessarily lie on a surface of revolution? Is the image of σ necessarily isometric to a surface of revolution?

3 Geodesics on special surfaces

3.1 Surfaces of revolution

Theorem 3.1 (Clairaut)**.** *Let γ be a geodesic on a surface of revolution, let f denote the distance from a point on the surface to the axis of rotation, and let ψ be the angle between $\dot\gamma$ and the meridians of the surface. Then $f\sin\psi$ is constant along the geodesic.*

Conversely, if $f\sin\psi$ is constant along some curve γ, and no arc of this curve is contained in a parallel, then γ is a geodesic.

Proof. The surface can be parametrized as $\sigma(u,v) = (f(u)\cos v, f(u)\sin v, g(u))$ with $\dot f^2 + \dot g^2 = 1$.

If γ has unit speed and forms the angle ψ with the meridians, then its velocity vector has the form

$$\dot\gamma = \cos\psi\sigma_u + f^{-1}\sin\psi\sigma_v,$$

that is, we have $\dot u = \cos\psi$, $\dot v = f^{-1}\sin\psi$.

Take the second of the geodesic equations:

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

Since $E = 1$, $F = 0$, $G = f(u)$, we obtain

$$\frac{d}{dt}(f^2\dot{v}) = 0.$$

Substituting the formula for $\dot{v}$ from the previous paragraph, we get

$$\frac{d}{dt}(f \sin\psi) = 0,$$

which proves the first part of the theorem.

The second part follows from the existence and uniqueness of geodesics with given initial conditions. □

Physical interpretation: angular momentum.
Qualitative behavior of geodesics.

3.2 Distance functions

Definition 3.2. *A function $t\colon M \to \mathbb{R}$ on a smooth surface M is called a* distance function *if its gradient has norm* 1 *everywhere:* $\|\nabla^M t\| = 1$.

Theorem 3.3. *The gradient curves of a distance functions together with its level curves form a geodesic system of coordinates. Namely, extend a diffeomorphism $s\colon C \to I$ between a level curve and an interval to a function $s\colon U \to \mathbb{R}$ by the flow of t. Then in the (s, t)-coordinates the metric has the form*

$$dt^2 + G(s,t)ds^2$$

In particular, the gradient curves of t are geodesics.

Proof. In the new coordinates (s, t), the first fundamental form is

$$\|\tilde{\sigma}_t\|^2 dt^2 + 2\langle\tilde{\sigma}_t, \tilde{\sigma}_s\rangle dtds + \|\tilde{\sigma}_s\|^2 ds^2$$

□

A single distance function produces a one-parameter family of geodesics, which are pairwise disjoint. The set of all (unparametrized or unit-speed parametrized) geodesics depends on two parameters. Therefore, in order to describe all geodesics, we would need a one-parameter family of distance functions. It turns out that this even allows to write down an equation of unparametrized geodesics.

Theorem 3.4. *Let $W\colon U \times J \to \mathbb{R}$ be a one-parameter family of distance functions, that is for every $\beta_0 \in J$ we have $\|\nabla W|_{\beta=\beta_0}\|^2 = 1$ or, equivalently*

$$\|dW|_{\beta=\beta_0}\|^2 = E^*\left(\frac{\partial W}{\partial u}\right)^2 + 2F^*\frac{\partial W}{\partial u}\frac{\partial W}{\partial u} + G^*\left(\frac{\partial W}{\partial v}\right)^2 = 1.$$

Then the equation

$$\frac{\partial W}{\partial \beta} = b \tag{6.1}$$

describes a two-parameter family (with parameters b and β) of geodesics.

Proof. Let us show that the level curves of the partial derivative $\left.\frac{\partial W}{\partial \beta}\right|_{\beta=\beta_0}$ are the gradient lines of the distance function $W|_{\beta=\beta_0}$. For this it suffices to prove

$$\left\langle \nabla \frac{\partial W}{\partial \beta}, \nabla W \right\rangle = 0.$$

This follows from the permutability of partial derivatives:

$$\nabla \frac{\partial W}{\partial \beta} = \frac{\partial}{\partial \beta} \nabla W$$

and from

$$\|\nabla W|_{\beta=\beta_0}\| = 1 \Rightarrow \frac{\partial}{\partial \beta} \|\nabla W\|^2 = 0.$$

□

Under certain non-degeneracy assumptions on the function W, equation 6.1 describes all geodesic arcs in a neighborhood of a point. However, it describes geodesic implicitely, without a constant speed parametrization.

Theorem 3.5. *Let $W : U \times I \times J \to \mathbb{R}$ be a two-parameter family of functions on U such that $\|\nabla W|_{(\alpha_0,\beta_0)}\| = 2\alpha_0$ for all $\alpha_0 \in I, \beta_0 \in J$. Then the system of equations*

$$\begin{cases} \left.\frac{\partial W}{\partial \beta}\right|_{\alpha=\frac{1}{2}} = b \\ \left.\frac{\partial W}{\partial \alpha}\right|_{\alpha=\frac{1}{2}} = t \end{cases} \tag{6.2}$$

describes a two-parameter family of geodesics (depending on parameters b and β) together with a unit-speed parametrization of each geodesic by a parameter t.

Proof. By the inverse function theorem, equations (6.2) can be resolved with respect to u and v, which yields

$$u = u(t, b, \beta), \quad v = v(t, b, \beta)$$

This gives a two-parameter family of curves $\gamma(t)$. Since $\gamma(t)$ satisfies the first equation of the system, from Theorem 3.4 we know that $\gamma(t)$ parametrizes a geodesic, that is

$$\dot{\gamma} = \lambda \cdot \nabla W|_{(\frac{1}{2},\beta_0)}$$

with λ depending on t. Since $\|\nabla W|_{(\frac{1}{2},\beta_0)}\| = 1$, we have to show that $\lambda = 1$. We have

$$2 = \frac{\partial}{\partial \alpha} \|\nabla W\|^2 = 2 \left\langle \nabla W, \nabla \frac{\partial W}{\partial \alpha} \right\rangle = 2\langle \nabla W, \nabla t\rangle \Rightarrow \langle \nabla W, \nabla t\rangle = 1.$$

Now observe that $\dot{\gamma}$ is the projection to γ of ∇t. Hence we have

$$1 = \langle \nabla W, \nabla t\rangle = \langle \nabla W, \dot{\gamma}\rangle = \lambda \|\nabla W\|^2 = \lambda,$$

and the theorem is proved. □

3.3 Geodesics on Liouville surfaces

Definition 3.6. *A surface patch* $\sigma\colon U \to \mathbb{R}^3$ *is called a* Liouville *parametrization, if its first fundamental form is of the kind*

$$I_\sigma = (U - V)(du^2 + dv^2), \tag{6.3}$$

where $U = U(u)$ *and* $V = V(v)$ *are functions of one variable.*

Not every surface allows a Liouville parametrization. It seems to be an open problem to describe in some explicit way the metrics that can be reparametrized to a Liouville form.

We are more interested not in the coordinates u, v on a Liouville surface, but rather in the coordinate lines (therefore it is more appropriate to speak of *Liouville nets*). The coordinate lines don't change if we perform a change of variables $u = \varphi(u')$, $v = \psi(v')$. The metric then takes the form

$$I_{\tilde\sigma} = (U - V)\left(U_1^2 du'^2 + V_1^2 dv'^2\right),$$

where U and $U_1 = \frac{\partial\varphi}{\partial u'}$ are now functions of u', while V and $V_1 = \frac{\partial\psi}{\partial v'}$ are functions of v'. Conversely, every metric of this more general form can be transformed to the more special form (6.3) by means of an independent change in each variable, that is without changing the coordinate lines.

Note that the coordinate lines in a Liouville net intersect at the right angle. Besides, the special Liouville parametrization is conformal.

Theorem 3.7. *The unit-speed geodesics of the metric* (6.3) *are given (up to time shift and time reversal) by*

$$\int \frac{du}{\sqrt{U - 2\beta}} \pm \int \frac{dv}{\sqrt{2\beta - V}} = b$$
$$\int \frac{U\,du}{\sqrt{U - 2\beta}} \pm \int \frac{V\,dv}{\sqrt{2\beta - V}} = t,$$

where β *and* b *are independent parameters, and the* $\pm$ *sign in the second equation coincides with the* $\pm$ *sign in the second equation.*

Proof. For a parameter $\alpha > 0$ we find a one-parameter family of functions $W(u, v)$ such that

$$\|\nabla W\|^2 = 2\alpha,$$

and then apply Theorem 3.5. In terms of the norm of the differential we have the equation

$$\frac{1}{U - V}\left(\left(\frac{\partial W}{\partial u}\right)^2 + \left(\frac{\partial W}{\partial v}\right)^2\right) = 2\alpha$$

Rewrite this as

$$-\left(\frac{\partial W}{\partial u}\right)^2 + 2\alpha U = \left(\frac{\partial W}{\partial v}\right)^2 + 2\alpha V.$$

We will find a one-parameter family of solutions if we make the assumption that both sides are equal to a constant, which we denote by 2β. This yields a system of two equations

$$\left(\frac{\partial W}{\partial u}\right)^2 = 2(\alpha U - \beta)$$
$$\left(\frac{\partial W}{\partial v}\right)^2 = 2(\beta - \alpha V).$$

Its general solution is given by

$$W = \pm \int \sqrt{2(\alpha U - \beta)}\, du \pm \int \sqrt{2(\beta - \alpha V)}\, dv + C,$$

but since we are interested only in partial derivatives of W and mod out the time reversal, the constant C does not matter, and the sign before the first integral can be set to the plus.

Now the theorem follows by a direct application of Theorem 3.5. □

Theorem 3.8 (Blaschke)**.** *In every coordinate quadrilateral of a Liouville net the two diagonals have the same length. Conversely, if all coordinate quadrilaterals of a surface parametrization possess this property, then the metric in these coordinates has the (general) Liouville form.*

Proof. We prove only the first part of this theorem, but not the converse.

□

3.4 Some theorems on ellipses and ellipsoids

Chapter 7

Fundamental equations of the surface theory

1 Theorema Egregium

1.1 Extrinsic and intrinsic objects

Let $M \subset \mathbb{R}^3$ be a smooth surface. Objects that are completely determined by the lengths of curves lying on M are called *intrinsic*; those that depend on the relative situation of points of M inside $\mathbb{R}^3$ are called *extrinsic*. In other words, extrinsic are those objects that change when the surface is bent without stretching or shrinking.

Lengths of curves on M are determined by the first fundamental form; conversely, the first fundamental form at every point is determined by lengths of short curves issued from this point. The second fundamental form can change under bending (example: bending a plane to a cylinder), therefore the quantities defined via the second fundamental form are a priori extrinsic.

Example 1.1. Angles between curves, areas of domains are intrinsic. Geodesic curvature is intrinsic as well. ∇f, $\operatorname{Hess} f$, Δf, $\operatorname{div} X$

Normal curvature is extrinsic. Therefore also extrinsic are the principal curvatures and principal curvature directions. The mean curvature is extrinsic, as the example of the plane and the cylinder shows.

The Gauss curvature $K = \frac{\det II}{\det I}$ is a priori extrinsic. The following theorem comes as a surprise.

Theorem 1.2 (Gauss)**.** *The Gauss curvature is extrinsic, that is completely determined by the first fundamental form of the surface.*

1.2 The intrinsic nature of the covariant derivative

Lemma 1.3. $\nabla_X Y$ *is intrinsic.*

Proof. It suffices to prove this for some basis vector fields, for example for the coordinate vector fields σ_u, σ_v of some surface patch σ. Use the short-hand notation $\nabla_{\sigma_u} = \nabla_u$. Then we have

$$\begin{aligned}\nabla_u \sigma_u &= T(\widetilde{\nabla}_u(\sigma_u)) = T(\sigma_{uu})\\ \nabla_u \sigma_v &= T(\sigma_{uv}) = \nabla_v \sigma_u\\ \nabla_v \sigma_v &= T(\sigma_{vv})\end{aligned}$$

The Leibniz rule for covariant derivatives yields

$$\begin{aligned}\langle \nabla_u \sigma_u, \sigma_u\rangle &= \frac{1}{2}\nabla_u\langle \sigma_u, \sigma_u\rangle = \frac{1}{2}E_u\\ \langle \nabla_v \sigma_u, \sigma_u\rangle &= \frac{1}{2}E_v\\ \langle \nabla_v \sigma_v, \sigma_v\rangle &= \frac{1}{2}G_v\\ \langle \nabla_u \sigma_v, \sigma_v\rangle &= \frac{1}{2}G_u\\ \langle \nabla_u \sigma_u, \sigma_v\rangle &= \nabla_u\langle \sigma_u, \sigma_v\rangle - \langle \sigma_u, \nabla_u\sigma_v\rangle = F_u - \frac{1}{2}E_v\\ \langle \nabla_v \sigma_v, \sigma_u\rangle &= F_v - \frac{1}{2}G_u\end{aligned}$$

If $X = a\sigma_u + b\sigma_v$, then

$$\begin{pmatrix}\langle X, \sigma_u\rangle\\ \langle X, \sigma_v\rangle\end{pmatrix} = \begin{pmatrix}E & F\\ F & G\end{pmatrix}\begin{pmatrix}a\\ b\end{pmatrix} \Rightarrow \begin{pmatrix}a\\ b\end{pmatrix} = \frac{1}{EG - F^2}\begin{pmatrix}G & -F\\ -F & E\end{pmatrix}\begin{pmatrix}\langle X, \sigma_u\rangle\\ \langle X, \sigma_v\rangle\end{pmatrix}$$

This allows to obtain the coordinates of the vectors $\nabla_u\sigma_u, \nabla_u\sigma_v, \nabla_v\sigma_v$ in the (σ_u, σ_v)-basis as (complicated) expressions involving only the components of the first fundamental form and their first derivatives. □

We will write the explicit expressions in the special case $F = 0$.

$$\begin{aligned}\nabla_u \sigma_u &= \frac{1}{2}\left(\frac{E_u}{E}\sigma_u - \frac{E_v}{G}\sigma_v\right)\\ \nabla_u \sigma_v = \nabla_v \sigma_u &= \frac{1}{2}\left(\frac{E_v}{E}\sigma_u + \frac{G_u}{G}\sigma_v\right)\\ \nabla_v \sigma_v &= \frac{1}{2}\left(-\frac{G_u}{E}\sigma_u + \frac{G_v}{G}\sigma_v\right)\end{aligned}$$

1.3 Linear and differential operators

Let $C^\infty(M)$ denote the space of C^∞-functions $M \to \mathbb{R}$, and $\mathcal{X}(M)$ denote the space of C^∞ vector fields on M.

Let $p \in M$. We say that a map $A_p\colon C^\infty(M) \to V$ or $A_p\colon \mathcal{X}(M) \to V$ to some real vector space V is $C^\infty(M)$-linear, if

$$A_p(x+y) = A_p(x) + A_p(y), \quad A_p(fx) = f(p)A_p(x) \quad \text{for all } x, y$$

A C^∞-linear map depends only on the value of the function or vector field at the point p. That is, we have $A_p(f) = cf(p)$ and $A_p(X) = \langle C, X(p)\rangle$.

Linear differential operators of the first (second) order depend linearly on the value at a point and on the first (and second) order derivatives at this point.

Example 1.4. $\nabla_X Y$ depends linearly on X and is a differential operator of first order in Y. This follows for example from its definition

$$\nabla_X Y = \widetilde{\nabla}_X Y - II(X,Y)\cdot \nu$$

or from its properties

$$\nabla_{fX} Y = f\nabla_X Y, \quad \nabla_X(fY) = \nabla_X f Y + f\nabla_X Y$$

$f \mapsto \nabla_X f$ is a first-order differential operator on functions. Conversely, every first-order differential operator on functions has the form $f \mapsto cf(p) + \nabla_X f|_p$.

Sometimes a differential operator has a smaller order than it seems it should have.

Lemma 1.5. *For every two vector fields the operator*

$$f \mapsto \nabla_X \nabla_Y f - \nabla_Y \nabla_X f$$

is a differential operator of the first order. Namely, it associates to f its derivative in the direction of the vector $\nabla_X Y - \nabla_Y X$.

The vector field

$$[X,Y] := \nabla_X Y - \nabla_Y X$$

is called the *commutator* of the vector fields X and Y.

Proof. Need to show

$$\nabla_X \nabla_Y f - \nabla_Y \nabla_X f = \nabla_{\nabla_X Y - \nabla_Y X} f$$

This follows from

$$\operatorname{Hess} f(X,Y) = \langle \nabla_X \nabla f, Y\rangle = \nabla_X \langle \nabla f, Y\rangle - \langle \nabla f, \nabla_X Y\rangle = \nabla_X(\nabla_Y f) - \nabla_{\nabla_X Y} f$$

and the symmetry of the Hessian, Lemma 3.7. □

1.4 The curvature tensor

Introduce the notation

$$\nabla^2_{X,Y} = \nabla_X \nabla_Y - \nabla_{\nabla_X Y}$$

We have seen that $\nabla^2 f = \operatorname{Hess} f$, and that this second covariant derivative of functions is symmetric with respect to the directions of derivation. Let us now look at the second covariant derivative of a vector field:

$$\nabla^2_{X,Y} Z = \nabla_X \nabla_Y Z - \nabla_{\nabla_X Y} Z$$

It is linear both in X and in Y, by the same argument as in the case of functions. It will be no more symmetric in X and Y, but its antisymmetrization produces an interesting object.

Lemma 1.6. *The map*

$$(X, Y, Z) \mapsto \nabla^2_{X,Y} Z - \nabla^2_{Y,X} Z$$

is $C^\infty(M)$-linear in each argument.

Proof. Only linearity with respect to Z needs to be checked. This can be done by a direct computation. □

Definition 1.7. *The tensor*

$$R(X, Y)Z := \nabla^2_{X,Y} Z - \nabla^2_{Y,X} Z = \nabla_X \nabla_Y Z - \nabla_Y \nabla_X Z - \nabla_{[X,Y]} Z$$

is called the curvature tensor*.*

1.5 Proof of the Theorema Egregium

Lemma 1.8. *Let X and Y be vector fields tangent to M. Then for every their extension to $\mathbb{R}^3$ their $\mathbb{R}^3$-commutator coincides with their M-commutator:*

$$\widetilde{\nabla}_X Y - \widetilde{\nabla}_Y X = \nabla_X Y - \nabla_Y X$$

Proof. Follows from $\nabla_X Y = \widetilde{\nabla}_X Y - II(X, Y) \cdot \nu$ and from the symmetry of the second fundamental form. □

Lemma 1.9.

$$\langle R(X, Y)Y, X\rangle = II(X, X)II(Y, Y) - II(X, Y)^2$$

Proof. We have

$$\begin{aligned}\nabla_X \nabla_Y Z &= \nabla_X (\widetilde{\nabla}_Y Z - II(Y, Z) \cdot \nu) \\ &= \nabla_X (\widetilde{\nabla}_Y Z) - \nabla_X (II(Y, Z)) \cdot \nu - II(Y, Z) \nabla_X \nu \\ &= \widetilde{\nabla}_X \widetilde{\nabla}_Y Z - II(Y, Z) \nabla_X \nu + c \cdot \nu\end{aligned}$$

It follows that

$$\nabla_X\nabla_Y Z - \nabla_Y\nabla_X Z = (\widetilde{\nabla}_X\widetilde{\nabla}_Y Z - \widetilde{\nabla}_Y\widetilde{\nabla}_X Z) + II(X,Z)\nabla_Y\nu - II(Y,Z)\nabla_X\nu + c\cdot\nu$$

Because of the previous lemma,

$$\begin{aligned} R(X,Y)Z &= (\widetilde{\nabla}_X\widetilde{\nabla}_Y Z - \widetilde{\nabla}_Y\widetilde{\nabla}_X Z - \widetilde{\nabla}_{[X,Y]}Z) + II(X,Z)\nabla_Y\nu - II(Y,Z)\nabla_X\nu + c\cdot\nu \\ &= II(X,Z)\nabla_Y\nu - II(Y,Z)\nabla_X\nu \end{aligned}$$

Hence

$$\begin{aligned} \langle R(X,Y)Y,X\rangle &= \langle II(X,Y)\nabla_Y\nu - II(Y,Y)\nabla_X\nu, X\rangle \\ &= -II(X,Y)II(X,Y) + II(Y,Y)II(X,X) \end{aligned}$$

□

Proof of Theorem 1.2. We have, for any two linearly independent vectors X, Y

$$K = \frac{\det II}{\det I} = \frac{II(X,X)II(Y,Y) - II(X,Y)^2}{I(X,X)I(Y,Y) - I(X,Y)^2} = \frac{\langle R(X,Y)Y,X\rangle}{I(X,X)I(Y,Y) - I(X,Y)^2}$$

The right hand side is intrinsic, because the covariant derivation is intrinsic. □

2 Applications of Theorema Egregium

Bibliography

[Arn90] V. I. Arnol′d. *Huygens and Barrow, Newton and Hooke*. Birkhäuser Verlag, Basel, 1990. Pioneers in mathematical analysis and catastrophe theory from evolvents to quasicrystals, Translated from the Russian by Eric J. F. Primrose.

[Arn95] V. I. Arnol′d. The geometry of spherical curves and quaternion algebra. *Uspekhi Mat. Nauk*, 50(1(301)):3–68, 1995.

[Brö75] Th. Bröcker. *Differentiable germs and catastrophes*. Cambridge University Press, Cambridge-New York-Melbourne, 1975. Translated from the German, last chapter and bibliography by L. Lander, London Mathematical Society Lecture Note Series, No. 17.

[GTT13] Étienne Ghys, Sergei Tabachnikov, and Vladlen Timorin. Osculating curves: around the Tait-Kneser theorem. *Math. Intelligencer*, 35(1):61–66, 2013.

[Hop35] Heinz Hopf. Über die Drehung der Tangenten und Sehnen ebener Kurven. *Compos. Math.*, 2:50–62, 1935.

[Hsi97] Chuan-Chih Hsiung. *A first course in differential geometry*. International Press, Cambridge, MA, 1997.

[MSW02] David Mumford, Caroline Series, and David Wright. *Indra's pearls*. Cambridge University Press, New York, 2002. The vision of Felix Klein.

[Pre10] Andrew Pressley. *Elementary differential geometry*. Springer Undergraduate Mathematics Series. Springer-Verlag London, Ltd., London, second edition, 2010.

[Ste40] J. Steiner. Von dem Krümmungs-Schwerpuncte ebener Curven. *J. Reine Angew. Math.*, 21:33–63, 101–133, 1840.