Differential Geometry

Ivan Izmestiev

TU Vienna, Spring 2021

Contents

Preface

This intends to be a slight modification of the notes for the DiffGeo course from the Spring semester 2020. Sections marked with an asterisk are not being read in 2021, but are left here for the curious reader. The author thanks Florian Bogner and Johannes Gams for spotting several typos and mistakes in the last year version.

1 Curves

1.1 Parametric curves

A curve may be thought of as the trajectory of a moving point. In the real word the point moves in the plane or in the (3-dimensional) space, so we speak about *plane curves* and *space curves*. One can draw a plane curve on the blackboard or on a piece of paper, the moving point being represented by the chalk or by the tip of the pen.

Formally, a moving point is a continuous map from a time interval to $\mathbb{R}^2$ or to $\mathbb{R}^3$. Continuous curves can exhibit strange behavior, for example:

- A constant map: its image is a point.
- Peano curve or Hilbert curve: they fill the square.
- The boundary of the Koch snowflake: a very beautiful injective curve, but of infinite length.

In differential geometry the map γ is assumed to be of class C^1 at least, that is continuously differentiable. This allows to define tangent vectors and compute the length. Second order differentiability allows to define the curvature. For simplicity we will assume the existence of derivatives of all orders, and mean this each time when we call a map *smooth*.

Definition 1.1. *A* smooth parametric curve *in $\mathbb{R}^n$ is a C^∞-map*

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

where $I \subset \mathbb{R}$ is an interval (of any kind: open, half-open, half-infinite, etc.).

Let $t_0 \in I$. The vector

$$\dot\gamma(t_0) = \left.\frac{d\gamma}{dt}\right|_{t=t_0}$$

is called the tangent vector *or the* velocity vector *of γ at t_0.*

Definition 1.2. *If $\dot\gamma(t_0) = 0$, then t_0 is called a* singular point *of the curve γ. If $\dot\gamma(t_0) \neq 0$, then t_0 is called a* regular point *of γ. If all points in I are regular for γ, then γ is called a* regular parametric curve*.*

Example 1.3. Consider the curve

$$\gamma\colon \mathbb{R} \to \mathbb{R}^2, \quad \gamma(t) = (t^2, t^3).$$

One has $\dot\gamma(t) = (2t, 3t^2)$ which vanishes if and only if $t = 0$. Thus $t = 0$ is the only singular point of this curve, see Figure 1.

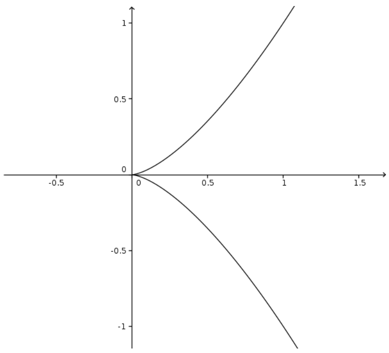

Figure 1: The so-called semicubical parabola. This curve has a cusp at $(0, 0)$.

Singularities have their own theory. Singular curves will appear in this course only episodically.

At a regular point a curve has a tangent. The tangent to γ at t_0 is the line with the parametric equation

$$l(t) = \gamma(t_0) + \dot\gamma(t_0)(t - t_0).$$

The tangent approximates the curve in the first order:

$$\gamma(t) = l(t) + o(t - t_0).$$

Note that the curve in Figure 1 has a tangent at $(0, 0)$ in the geometric sense, but there is no parametrized straight line which would approximate the curve in the first order (the curve "changes its direction").

Remark 1.4. The reader may have noted that we are sometimes mixing the curve (a map) with its image. Ideally one would like to call curves certain subsets of $\mathbb{R}^n$ and use parametrizations only as a tool for studying such subsets. There are some technical difficulties in giving proper definitions. We will say more about that in Section 1.5.

In any case, parametric curves which differ by a "reasonable" parameter change should be considered as representing the same curve. More exactly, let $\varphi\colon J \to I$ be a diffeomorphism between two intervals of the real line, and let $\gamma\colon I \to \mathbb{R}^n$ be a smooth curve. Then the curve

$$\delta := \gamma \circ \varphi\colon J \to \mathbb{R}^n$$

is called a *reparametrization* of γ. The relation "δ is a reparametrization of γ" is reflexive, symmetric, and transitive, which allows to formulate the following definition.

Definition 1.5. *A* smooth curve *in* $\mathbb{R}^n$ *is an equivalence class of smooth parametric curves under the reparametrization equivalence relation.*

Lemma 1.6. *Regularity is preserved under reparametrization. A point* $s_0 \in J$ *is a regular point of* δ *if and only if* $\varphi(s_0) \in I$ *is a regular point of* γ.

Proof. By the chain rule one has

$$\dot{\delta}(s_0) = \dot{\gamma}(\varphi(s_0))\ \varphi'(s_0).$$

Since $\varphi'(s) \neq 0$, one has $\dot{\delta}(s_0) \neq 0$ if and only if $\dot{\gamma}(\varphi(s_0)) \neq 0$. □

This lemma implies that the notion "regular curve" is well-defined.

Remark 1.7. The parametric curves $\gamma(t) = (t, 0)$ and $\gamma(t) = (t^3, 0)$ represent different curves. The first one is regular, the other one is not.

We will often omit the word "parametric" and say "let $\gamma\colon I \to \mathbb{R}^n$ be a smooth/regular curve".

Example 1.8. The tractrix and its different parametrizations.

1.2 Length and natural parametrization

Definition 1.9. *Let* $\gamma\colon I \to \mathbb{R}^n$ *be a smooth parametric curve. Its* length *is defined as*

$$L(\gamma) := \int_I \|\dot{\gamma}\|\, dt.$$

The integral is understood here in the measure-theoretic sense, so that $L(\gamma) > 0$. It can be computed as the Riemann integral, with the integration done in the positive direction along I.

Example 1.10. Compute the length of the logarithmic spiral

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

on the interval $(-\infty, 0]$.

One has

$$\dot{\gamma}(t) = e^t(\cos t - \sin t, \cos t + \sin t),$$

thus $\|\dot{\gamma}(t)\| = \sqrt{2}e^t$ and $L(\gamma) = \sqrt{2}$.

Lemma 1.11. *A reparametrization does not change the length of a curve, so that the length of a smooth curve is well-defined.*

Proof.

$$L(\delta) = \int_J \|\dot{\delta}\|\, ds = \int_J \|(\dot{\gamma} \circ \varphi)\varphi'\|\, dt = \int_I \|\dot{\gamma}\|\, dt = L(\gamma)$$

□

Definition 1.9 has a clear meaning: the distance is the integral of the speed. An equivalent way of measuring the length of a smooth curve is given in the next theorem.

Theorem 1.12. *The length of a smooth curve is the least upper bound of the lengths of inscribed polygons.*

For simplicity assume that $I = [a, b]$. An inscribed polygon is the union of chords

$$[\gamma(a), \gamma(t_1)] \cup [\gamma(t_1), \gamma(t_2)] \cup \cdots \cup [\gamma(t_n), \gamma(b)]$$

where $a < t_1 < t_2 < \cdots t_n < b$.

Lemma 1.13. *An arc of a curve is longer than the spanning cord:*

$$\int_{t_i}^{t_{i+1}} \|\dot{\gamma}(t)\| \, dt \geq \|\gamma(t_{i+1}) - \gamma(t_i)\|$$

Proof. $\int_{t_i}^{t_{i+1}} \|\dot{\gamma}(t)\| \, dt \geq \left\| \int_{t_i}^{t_{i+1}} \dot{\gamma}(t) dt \right\| = \|\gamma(t_{i+1}) - \gamma(t_i)\|$ □

Proof of Theorem 1.12. Lemma 1.13 implies that the length of a smooth curve is bigger than the length of any inscribed polygon. Let us show that by choosing a fine subdivision of $[a, b]$ one can make the length of the inscribed polygon arbitrarily close to the integral of $\|\dot{\gamma}(t)\|$. Since $\dot{\gamma}(t)$ is a continuous (vector-valued) function on a closed interval, for every $\varepsilon > 0$ there is $\delta > 0$ such that if $|t - t'| < \delta$, then $\|\gamma(t) - \gamma(t')\| < \varepsilon$. Choose $t_1, \ldots, t_n$ so that $|t_{i+1} - t_i| < \delta$. Then one has

$$\left\| \int_{t_i}^{t_{i+1}} \dot{\gamma}(t) \, dt - (t_{i+1} - t_i)\dot{\gamma}(t_i) \right\| < (t_{i+1} - t_i)\varepsilon > \left| \int_{t_i}^{t_{i+1}} \|\dot{\gamma}(t)\| \, dt - (t_{i+1} - t_i)\|\dot{\gamma}(t_i)\| \right|$$

or, in other words,

$$\left\| \int_{t_i}^{t_{i+1}} \dot{\gamma}(t) \, dt \right\| \approx (t_{i+1} - t_i)\|\dot{\gamma}(t_i)\| \approx \int_{t_i}^{t_{i+1}} \|\dot{\gamma}(t)\| \, dt,$$

where $\approx$ means "differ by less than $(t_{i+1} - t_i)\varepsilon$". Summing up we obtain that the length of the inscribed polygon and the integral of the norm of the velocity vector differ by less than $(b - a)\varepsilon$. Since ε can be chosen arbitrarily small, the theorem is proved. □

Remark 1.14. A (continuous, not smooth) curve is called rectifiable if the lengths of inscribed polygons are bounded from above. The Peano curve, the Hilbert curve, the boundary of the Koch snowflake are not rectifiable. Every Lipschitz curve is rectifiable.

Definition 1.15. *A parametric curve γ is called a* unit-speed curve *if $\|\dot{\gamma}(t)\| = 1$ for all t.*

Unit-speed curves are also called *arc-length parametrized* or *naturally parametrized.* The length of an arc-length parametrized curve $\gamma\colon [a, b] \to \mathbb{R}^2$ is $b - a$.

Theorem 1.16. *Every regular parametric 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 $\delta = \gamma \circ \varphi$ is a unit-speed curve:

$$\dot\delta(s) = \dot\gamma(\varphi(s))\varphi'(s) = 1$$

□

Theorem 1.16 can be reformulated as saying that every regular curve has a unit-speed parametric representative.

Lemma 1.17. *Any two unit-speed parametrizations differ by time shift and/or time reversal:* $\varphi(s) = \pm t + c$.

Proof. Indeed, if $\|\dot\delta(s)\| = 1 = \|\dot\gamma(\varphi(s))\|$, then $|\varphi'(s)| = 1$, which implies that φ is a linear function with the coefficient ± 1. □

1.3 The local structure of a smooth curve

Theorem 1.18. *Let $\gamma\colon I \to \mathbb{R}^n$ be a regular curve. Then for every $t_0 \in I$ there is an open neighborhood J of t_0 and an open subset $V \subset \mathbb{R}^n$ such that $\gamma(J) \subset V$, and a diffeomorphism $\Phi\colon V \to W$ to another open subset of $\mathbb{R}^n$ such that*

$$\Phi(\gamma(t)) = (t, 0, \ldots, 0) \textit{ for all } t \in J.$$

See Figure 1.3. The map Φ is sometimes called a local straightening of the curve γ. Observe that the set V may contain points $\gamma(t)$ for t outside of the interval J.

Proof. Let $\gamma(t) = (x_1(t), \ldots, x_n(t))$. Without loss of generality, $\dot x_1(t_0) \neq 0$. Consider the map

$$g\colon I \times \mathbb{R}^{n-1} \to \mathbb{R}^n,$$
$$g(t, y) = \gamma(t) + (0, y).$$

The Jacobian matrix of g has the form

$$\begin{pmatrix} \dot x_1 & * \\ 0 & E, \end{pmatrix}$$

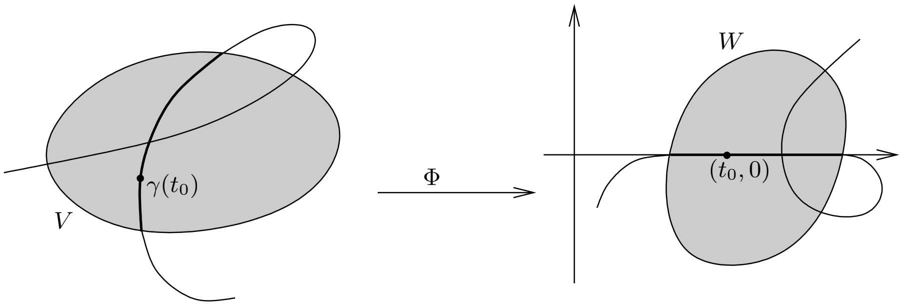

Figure 2: Local straightening of a regular curve.

where E is the identity matrix. The determinant of this matrix at $(t_0, 0)$ does not vanish by assumption. By the inverse function theorem the map g is a local diffeomorphism at $(t_0, 0)$. That is, there exists an open subset $W \subset I \times \mathbb{R}^{n-1}$ containing $(t_0, 0)$ such that the restriction $g|_W$ is a diffeomorphism onto its image $V = g(W)$. One has $\gamma(t_0) = g(t_0, 0) \in V$. Since V is open, the preimage $\gamma^{-1}(V)$ is an open subset of I containing t_0. Let $J \subset \gamma^{-1}(V)$ be an open interval containing t_0. Since $g(t, 0) = \gamma(t)$ and $\Phi = g^{-1}$, one has $\Phi(\gamma(t)) = (t, 0)$. □

1.4 Other ways to describe a curve

Proposition 1.19. *Let $f: I \to \mathbb{R}$ be a smooth function on an interval. Then the curve*

$$\gamma(t) = (t, f(t))$$

(which parametrizes the graph of f) is regular.

Conversely, every regular curve $\gamma: I \to \mathbb{R}^2$ is locally the graph of a function. That is, for every $t_0 \in I$ there is a subinterval $J \subset I$ containing t_0 such that

$$\gamma(J) = \{(x, f(x)) \mid x \in K\} \quad \textit{or} \quad \gamma(J) = \{(g(y), y) \mid y \in L\}$$

for some interval K or L.

Proof. The first part of the proposition is obvious.

The second part follows from the inverse function theorem: if $\dot{x}(t_0) \neq 0$, then the function $x(t)$ can be locally inverted around t_0. Setting $f(x) = y(t(x))$ we make the image of γ to the graph of f. □

Proposition 1.20. *Let $U \subset \mathbb{R}^2$ be an open set, and $F: U \to \mathbb{R}$ a smooth function such that for all $x \in F^{-1}(0)$ the gradient $\operatorname{grad}_x F$ does not vanish. Then for every $x_0 \in F^{-1}(0)$ there is an open subset $V \subset U$ containing x_0 such that $F^{-1}(0) \cap V$ is the image of a regular curve.*

Proof. By the implicit function theorem there is V such that $F^{-1}(0) \cap V$ is the graph of a smooth function (of one of the coordinates). And the graph of a smooth function is the image of a regular curve. □

1.5 Simple curves

Here "simple" does not mean "unsophisticated" but rather "without self-intersections" (nobody knows why).

If one wants to look at curves as sets, then the following definition is a proper one.

Definition 1.21. *A* one-dimensional smooth submanifold *of* $\mathbb{R}^n$ *is a connected subset* $M \subset \mathbb{R}^n$ *such that for every* $p \in M$ *there is an open subset* $V \subset \mathbb{R}^n$ *such that* $p \in V$ *and a diffeomorphism* $\Phi\colon V \to W$ *to another open subset of* $\mathbb{R}^n$ *such that* $\Phi(V \cap M) = W \cap \{x_2 = \ldots = x_n = 0\}$.

See Figure 1.5. The difference from the situation described in Theorem 1.18 is that all of $M \cap V$ is straightened. This would be impossible if M was a curve intersecting itself at p.

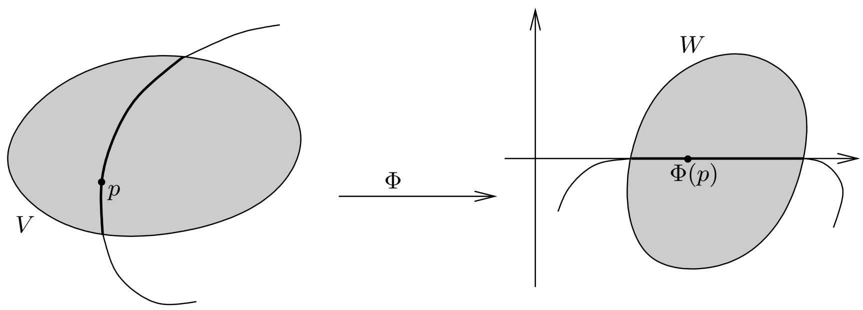

Figure 3: Local straightening of a one-dimensional submanifold.

Let us now give the corresponding "parametric" definitions.

Definition 1.22. *A* simple smooth open curve *is the image of a regular parametric curve* $\gamma\colon I \to \mathbb{R}^n$ *defined on an open interval* I *and such that the map* γ *is a homeomorphism onto its image.*

The condition of γ being a homeomorphism onto its image implies that γ is injective. Example on Figure 4 shows that an injective continuous map is not always a homeomorphism onto its image.

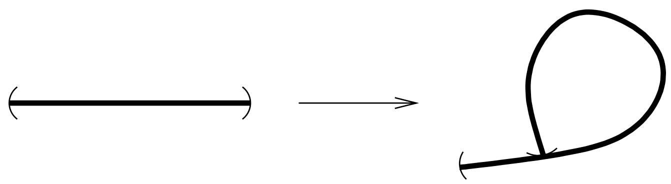

Figure 4: An injective regular map of an interval which is not a simple curve.

Definition 1.23. *A* simple smooth closed curve *is the image of a regular parametric curve* $\gamma\colon \mathbb{R} \to \mathbb{R}^n$ *such that* $\gamma(t_1) = \gamma(t_2)$ *if and only if* $t_1 - t_2 \in \mathbb{Z}$.

Theorem 1.24. *Every simple curve in* $\mathbb{R}^n$ *is a one-dimensional submanifold of* $\mathbb{R}^n$ *and vice versa: every one-dimensional submanifold is a simple curve.*

We do not prove this theorem.

1.6 Curvature of a regular curve

Lecture 2

The curvature is the change in the direction of a curve. Since the direction is given by the velocity vector $\dot{\gamma}$, the curvature should be related to the second derivative $\ddot{\gamma}$. However, the change of parameter influences the derivatives, so we cannot use $\ddot{\gamma}$ unconditionally. In order to feel the curvature, let us move along the curve with the unit speed. The magnitude of the centrifugal force is the curvature.

Definition 1.25. *Let* γ *be a regular curve in* $\mathbb{R}^n$*, and let* δ *be its unit-speed reparametrization. The* curvature *of* γ *at* $\gamma(t_0)$ *is defined as the norm of the second derivative of* δ *at the corresponding point:*

$$\kappa(t_0) = \left\| \frac{d^2\delta}{ds^2}(s_0) \right\|,$$

where $\delta = \gamma \circ \varphi$ *and* $t_0 = \varphi(s_0)$.

Although the natural parameter s on a curve is not unique, any other choice of it is related by $s' = \pm s + c$, which does not change the norm of the second derivative, and can at most revert its direction.

Lemma 1.26. *A curve has zero curvature at every point if and only if it is a line segment.*

Proof.

$$\frac{d^2\delta}{ds^2} = 0 \Leftrightarrow \frac{d\delta}{ds} = \text{const} \Leftrightarrow \delta(s) = a + sv \text{ for some } a, v \in \mathbb{R}^n.$$

□

Example 1.27. The curvature of a circle of radius R is equal to $\frac{1}{R}$. One can give an explicit unit-speed parametrization:

$$\delta(s) = \left(R\cos\frac{s}{R}, R\sin\frac{s}{R} \right).$$

Compute the derivatives:

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

Example 1.28. Let us compute the curvature of a helix

$$\gamma(t) = (a\cos t, a\sin t, bt), \quad a, b > 0.$$

The parametrization has constant speed $\|\dot\gamma(t)\| = \sqrt{a^2+b^2}$. Thus a natural parameter is $s = \frac{t}{\|\dot\gamma\|}$, and the curvature is equal to

$$\kappa = \frac{\|\ddot\gamma\|}{\|\dot\gamma\|^2} = \frac{a}{a^2+b^2}.$$

The following theorem allows to compute the curvature of an arbitrary parametric curve straightforwardly, without computing its unit-speed reparametrization.

For most curves computing an explicit unit-speed parametrization is difficult (it involves an integral, see the proof of Theorem 1.16). The following theorem gives a formula which allows to compute the curvature from any parametrization.

Theorem 1.29. *The curvature of an arbitrary regular curve γ is given by*

$$\kappa = \frac{A(\dot\gamma, \ddot\gamma)}{\|\dot\gamma\|^3},$$

where $A(\dot\gamma, \ddot\gamma)$ is the area of the parallelogramm spanned by the vectors $\dot\gamma$ and $\ddot\gamma$. In particular,

$$A(\dot\gamma, \ddot\gamma) = \begin{cases} |\det(\dot\gamma, \ddot\gamma)|, & \text{for } \gamma\colon I \to \mathbb{R}^2, \\ |\dot\gamma \times \ddot\gamma|, & \text{for } \gamma\colon I \to \mathbb{R}^3. \end{cases}$$

The proof is computational (chain rule and the Gram determinant), and we omit it.

Instead, let us give a physical explanation. The curvature should be proportional to the magnitude of the normal component of the acceleration (the tangential component describes the change in the speed). The proportionality coefficient should depend on the speed only. In order to see why one has to divide by the square of the speed, let us look at the units of measurement (this method is called dimensional analysis). The unit of measurement for curvature should contain no time, since the curvature is independent of the parametrization. The acceleration is measured in m/s^2, and the speed in m/s. Thus in order to eliminate the time we should divide the acceleration by the speed squared. The curvature is then measured in m^{-1}, which is reasonable.

If the curve lies in a plane, then its curvature can be computed as follows.

Theorem 1.30. *Let γ be a regular curve in the plane, and let $\alpha(t)$ be the angle from the positive x-semiaxis to the velocity vector $\dot\gamma(t)$:*

$$\dot\gamma(t) = \|\dot\gamma(t)\|(\cos\alpha(t), \sin\alpha(t)).$$

Then one has

$$\kappa(t) = \frac{|\dot\alpha(t)|}{\|\dot\gamma(t)\|}.$$

Proof. First, let us show that the expression on the right hand side is independent of the choice of parameter. Let $\varphi\colon J \to I$ be a reparametrization, $\delta = \gamma \circ \varphi$ the new curve, and $\beta = \alpha \circ \varphi$ the new angle function. By the chain rule one has

$$\frac{|\beta'|}{\|\delta'\|} = \frac{|\dot{\alpha} \cdot \varphi'|}{\|\dot{\gamma} \cdot \varphi'\|} = \frac{|\dot{\alpha}|}{\|\dot{\gamma}\|}.$$

Next, show that the formula works for a natural parametrization. Let s be a natural parameter. Then one has

$$\begin{aligned} \delta'(s) &= (\cos\beta(s), \sin\beta(s)), \\ \delta''(s) &= \beta'(s)(-\sin\beta(s), \cos\beta(s)). \end{aligned}$$

Thus $\kappa = \|\delta''\| = |\beta'|$, and we are done. □

Example 1.31. (I wanted to talk about this in the first lecture.) The *tractrix* is the path of an object which is initially situated at $(1, 0)$ and is pulled by a puller (tractor) to which it is attached by a rope of unit length and which moves on the y-axis from $(0, 0)$ upwards. At any moment of time the rope is tangent to the trajectory of the pulled object. This allows us to write a differential equation of the tractrix:

$$\frac{dy}{dx} = -\frac{\sqrt{1-x^2}}{x},$$

see Figure 5.

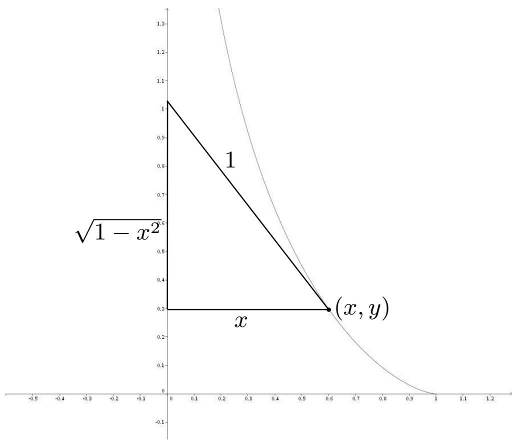

Figure 5: The tractrix.

It follows that

$$y = -\int_1^x \frac{\sqrt{1-x^2}}{x}\, dx = \log\frac{1+\sqrt{1-x^2}}{x} - \sqrt{1-x^2}.$$

This represents the tractrix as the graph of a function. A different parametrization can be obtained by substituting $x = \sin t$:

$$\gamma(t) = \left(\sin t, \log\cot\frac{t}{2} - \cos t\right), \quad t \in (0, \pi/2).$$

(Note that this parameter is the angle between the tangent and the y-axis.)

What we want to do now is to compute the curvature of the tractrix. Let us do it twice: first with the help of Theorem 1.29, then with the help of Theorem 1.30.

Let us use the trigonometric parametrization of the tractrix. A somewhat lengthy computation produces the following formulas for the first and the second derivatives of γ:

$$\dot\gamma(t) = \left(\cos t, -\frac{\cos^2 t}{\sin t}\right), \quad \ddot\gamma(t) = \left(-\sin t, \frac{\cos t(1+\sin^2 t)}{\sin^2 t}\right).$$

From this we compute

$$\det(\dot\gamma, \ddot\gamma) = \frac{\cos^2 t}{\sin^2 t}, \quad \|\dot\gamma\| = \frac{\cos t}{\sin t},$$

which leads to

$$\kappa = \tan t.$$

Let α be the angle at the vertex (x, y) in the right-angled triangle in Figure 5 (it is π minus the angle we meant in Theorem 1.30, but this does not matter). Consider the parametrization of the tractrix by the variable x. One has

$$\alpha = \arccos x \Rightarrow \frac{d\alpha}{dx} = \frac{1}{\sqrt{1-x^2}}.$$

Further,

$$\left\|\frac{d\gamma}{dx}\right\| = \sqrt{1 + \left(\frac{dy}{dx}\right)^2} = \sqrt{1 + \frac{1-x^2}{x^2}} = \frac{1}{x}.$$

It follows that

$$\kappa = \frac{x}{\sqrt{1-x^2}},$$

which coincides with the previously obtained formula.

Still another parametrization is obtained by substituting $x = \frac{1}{\cosh t}$:

$$\gamma(t) = \left(\frac{1}{\cosh t}, t - \tanh t\right), \quad t \in (0, +\infty).$$

1.7 Signed curvature and turning number

Lemma 1.32. *If γ is a constant-speed curve, then $\dot\gamma \perp \ddot\gamma$.*

Proof.

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

□

Let $\gamma: I \to \mathbb{R}^2$ be a unit-speed curve in the plane. For every $t \in I$ denote by $\nu_s(t)$ the unit normal to $\dot{\gamma}(t)$ such that $(\dot{\gamma}(t), \nu_s(t))$ is a positively oriented orthonormal basis. By Lemma 1.32 the vectors $\ddot{\gamma}$ and ν_s are collinear.

Definition 1.33. *The number $\kappa_s(t)$ such that $\ddot{\gamma}(t) = \kappa_s(t)\nu_s(t)$ is called the* signed curvature *of γ at the point $\gamma(t)$.*

Since $\|\ddot{\gamma}(t)\| = \kappa(t)$, one has $\kappa_s(t) = \pm\kappa(t)$. The sign is positive if the curve turns left and negative if it turns right.

If the curve γ is non-unit speed, then one defines the signed curvature using a unit-speed reparametrization. Observe however that changing the direction of motion along the curve changes the sign of the signed curvature. Thus the signed curvature is well-defined for oriented regular curves.

From the proof of Theorem 1.30 one sees that $\kappa_s(t) = \dot{\alpha}(t)/\|\dot{\gamma}(t)\|$ and in particular, for the unit-speed curves

$$\kappa_s(t) = \dot{\alpha}(t).$$

Theorem 1.34. *The total signed curvature of a closed regular curve in the plane is an integer multiple of 2π.*

Proof. A closed regular curve is a periodic map $\gamma: \mathbb{R} \to \mathbb{R}^2$, $\gamma(t+1) = \gamma(t)$. One has

$$\int_0^1 \dot{\alpha}(t)\, dt = \alpha(1) - \alpha(0).$$

Since $\dot{\gamma}(0) = \dot{\gamma}(1)$, one has $\alpha(1) - \alpha(0) = 2k\pi$. □

In the above proof we use the fact that the turning angle $\alpha(t)$ is well-defined up to a multiple of 2π, and that there is a choice of $\alpha(t)$ for each $t \in I$ which makes the map $t \mapsto \alpha(t)$ continuous.

Definition 1.35. *The number k from the above theorem is called the* turning number *of the curve γ.*

This is the number of times the tangent vector turns while we run along the curve. Figure 6 shows examples of curves with the turning number 2 and 0.

Theorem 1.36 (Hopf's Umlaufsatz)**.** *The turning number of a simple closed curve is equal to ± 1.*

A proof can be found in [Hsi97].

The following theorem is sometimes called the fundamental theorem of plane curves.

Theorem 1.37. *Let $f: I \to \mathbb{R}$ be any smooth function. Then there is a unit-speed curve $\gamma: I \to \mathbb{R}^2$ whose signed curvature is f. This curve is unique up to a rigid motion: if $\tilde{\gamma}: [a,b] \to \mathbb{R}$ is another curve whose signed curvature is f, then $\tilde{\gamma} = \Phi \circ \gamma$, where $\Phi: \mathbb{R}^2 \to \mathbb{R}^2$ is an orientation-preserving isometry of $\mathbb{R}^2$.*

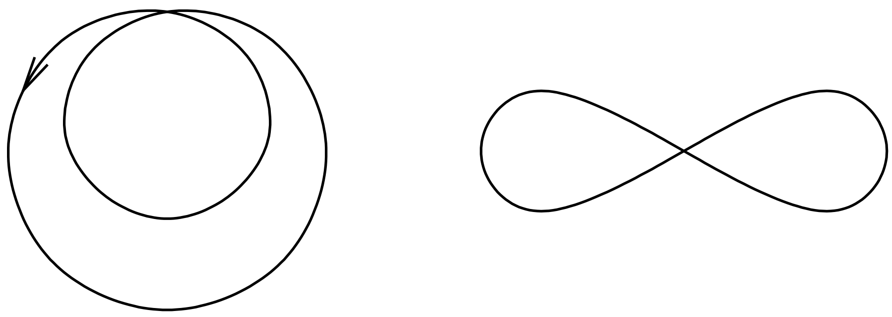

Figure 6: Curves with the turning number 2 and 0.

Proof. A unit-speed curve with the signed curvature $f(t)$ is the solution of the following system of equations.

$$\gamma(t) = \gamma(t_0) + \int_{t_0}^{t} \dot\gamma(t)\,dt, \quad \dot\gamma(t) = (\cos\alpha(t), \sin\alpha(t)), \quad \dot\alpha(t) = f(t)$$

The general solution of the last equation is

$$\alpha(t) = \alpha(t_0) + \int_{t_0}^{t} f(t)\,dt.$$

Substituting it into the second and then the first one, we obtain a curve with the given curvature. Changing the value of $\alpha(t_0)$ rotates the curve, and changing the value of $\gamma(t_0)$ translates the curve, thus the solution is unique up to a rigid motion. □

One can relax the condition on f in the previous theorem by requiring it to be continuous (in which case γ will be C^2) or Riemann-integrable.

Example 1.38. What curve should one use as a transition between two straight parts of a train track? The first idea coming to mind is a circular arc. But then, traveling along this track with a constant speed you will experience a jolt to one side when entering the turn and a jolt to the opposite side when leaving it. The graph of the curvature as a function of length demonstrates this clearly, see Figure 7.

A better solution is to prescribe a linear increase in the curvature followed by a linear decrease. The curve whose curvature is a linear function of the arc length is called the *clothoid* or the Euler spiral, see Figure 8. Clothoid arcs are actually used for this purpose, see the Wikipedia article.

The clothoid is also an *elastic curve*, that is it minimizes the bending energy $\int \kappa^2\,ds$.

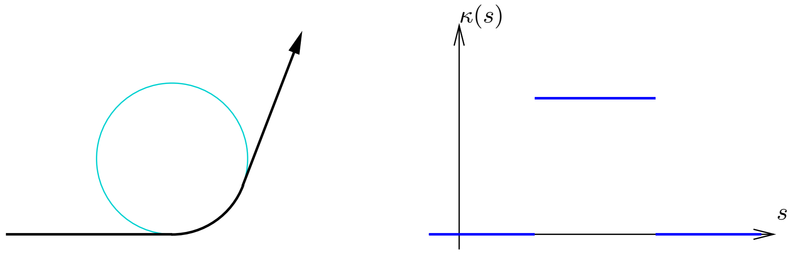

Figure 7: Traveling along this curve is uncomfortable.

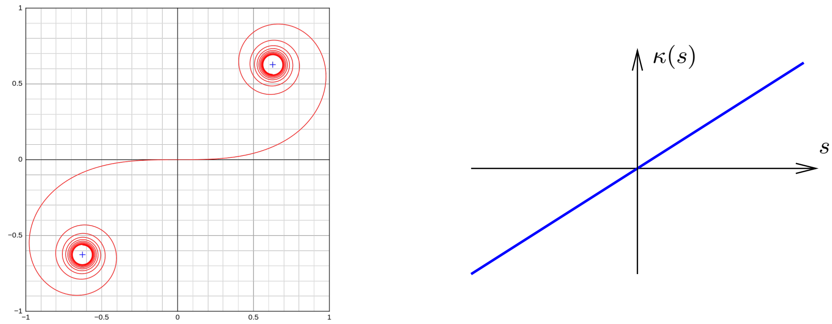

Figure 8: An image from Wikipedia: “A double-end Euler spiral”, copyright AdiJapan, Licence CC BY-SA 3.0. The curvature changes linearly with the length.

1.8 Space curves: curvature and torsion

While a plane curve is determined by its curvature uniquely up to a rigid motion, a space curve is not. For example, there are different helices with the same constant curvature, see Example 1.28. In addition to the curvature there is another quantity which describes the bending of a space curve.

Let γ be a unit-speed curve in $\mathbb{R}^3$. Its unit tangent $\dot\gamma(t)$ will now be denoted as

$$T(t) := \dot\gamma(t).$$

As we have seen, $\|\dot\gamma(t)\| = 1$ for all t implies that the vector $\ddot\gamma(t)$ is orthogonal to γ at the point $\gamma(t)$. If $\ddot\gamma(t) \neq 0$, then the unit vector

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

is called the *(principal) normal* of γ at t. By definition of curvature one has

$$\dot T = \kappa N.$$

The plane through $\gamma(t)$ spanned by the vectors $T(t)$ and $N(t)$ is called the *osculating plane* of γ at $\gamma(t)$. This plane has contact of second order with γ. Indeed, one has

$$\gamma(t) = \gamma(t_0) + \dot{\gamma}(t_0)(t - t_0) + \ddot{\gamma}(t_0)\frac{(t - t_0)^2}{2} + o((t - t_0)^2),$$

and the sum of the first three terms lies in the osculating plane.

The curvature κ of γ measures how fast the tangent to γ turns. Let us now measure how fast does the osculating plane turn. It turns as fast as its normal vector does.

Definition 1.39. *The* binormal *to γ at the point $\gamma(t)$ is defined as*

$$B(t) = T(t) \times N(t).$$

The family of orthonormal bases $(T(t), N(t), B(t))$ is called the Frenet-Serret frame *associated with the curve γ.*

Definition-Lemma 1.40. *The derivative of the binormal is collinear with the principal normal:*

$$\dot{B}(t) = -\tau(t)N(t).$$

The number $\tau(t)$ is called the torsion *of γ at $\gamma(t)$.*

Proof. Since $\|B(t)\| = 1$ for all t, one has $\dot{B} \perp B$. By differentiating the identity $B = T \times N$ one obtains

$$\dot{B} = \dot{T} \times N + T \times \dot{N} = \kappa N \times N + T \times \dot{N} = T \times \dot{N}.$$

Thus $\dot{B} \perp T$.

As $\dot{B}$ is orthogonal to both B and T, it is collinear with N. □

By definition, the torsion can be positive or negative. Positive torsion corresponds to the "right screw".

Theorem 1.41 (Frenet-Serret formulas)**.** *The Frenet-Serret frame satisfies the following system of equations:*

$$\begin{aligned} \dot{T} &= \kappa N \\ \dot{N} &= -\kappa T + \tau B \\ \dot{B} &= -\tau N \end{aligned}$$

Proof. The first and the third equations are the definitions of curvature and torsion. In order to prove the second, we need to show

$$\langle \dot{N}, T\rangle = -\kappa, \quad \langle \dot{N}, N\rangle = 0, \quad \langle \dot{N}, B\rangle = -\tau. \tag{1}$$

Use the standard trick, differentiate the inner product:

$$\langle N, T\rangle = 0 \Rightarrow \frac{d}{dt}\langle N, T\rangle = 0 \Rightarrow \langle \dot{N}, T\rangle + \langle N, \dot{T}\rangle = 0.$$

Since $\langle N, \dot{T} \rangle = \langle N, \kappa N \rangle = \kappa$, we obtain $\langle \dot{N}, T \rangle = -\kappa$, and the first of equations (1) is proved.

The second of equations (1) follows from $\|N\| = 1$, and the last is proved similarly to the first one. $\square$

In order to compute the torsion of a given parametric curve, one can proceed according to the definitions: reparametrize, compute the normal and binormal, differentiate the binormal. As in the case of the curvature, there is an explicit formula for an arbitrarily parametrized curve.

Lemma 1.42. *The torsion of an arbitrary regular curve in $\mathbb{R}^3$ is given by*

$$\tau = \frac{\det(\dot{\gamma}, \ddot{\gamma}, \dddot{\gamma})}{\|\dot{\gamma} \times \ddot{\gamma}\|^2}.$$

In particular, for a unit-speed curve one has

$$\tau = \frac{\det(\dot{\gamma}, \ddot{\gamma}, \dddot{\gamma})}{\kappa^2}.$$

We omit the proof.

The Frenet-Serret formulas in the matrix form:

$$\begin{pmatrix} \dot{T} \\ \dot{N} \\ \dot{B} \end{pmatrix} = \begin{pmatrix} 0 & \kappa & 0 \\ -\kappa & 0 & \tau \\ 0 & -\tau & 0 \end{pmatrix} \begin{pmatrix} T \\ N \\ B \end{pmatrix}$$

The matrix in this formula is skew-symmetric, and this is a manifestation of the following more general fact.

Theorem 1.43. *Let $(e_1(t), \ldots, e_n(t))$ be a smooth family of orthonormal bases of $\mathbb{R}^n$, and let*

$$\begin{pmatrix} \dot{e}_1(t) \\ \vdots \\ \dot{e}_n(t) \end{pmatrix} = Q(t) \begin{pmatrix} e_1(t) \\ \vdots \\ e_n(t) \end{pmatrix}$$

Then the matrix $Q(t)$ is skew-symmetric: $Q^\top(t) = -Q(t)$.

For the proof we will need the following lemma.

Lemma 1.44. *If $M(t) \in SO(n)$ is a smooth family of orthogonal matrices with $M(t_0) = I$, then the matrix $\dot{M}(t_0)$ is skew-symmetric.*

Proof. Differentiate the orthogonality condition:

$$M^\top M = I \Rightarrow \dot{M}^\top M + M^\top \dot{M} = 0 \Rightarrow \dot{M}^\top(t_0) + \dot{M}(t_0) = 0.$$

$\square$

Proof of Theorem 1.43. Denote $\mathcal{E}(t) = (e_1(t), \dots e_n(t))^\top$. Let $\mathcal{E}(t) = M(t)\mathcal{E}(t_0)$. Then $M(t) \in SO(n)$ and $M(t_0) = I$. It follows that $\dot{\mathcal{E}}(t) = \dot{M}(t)\mathcal{E}(t_0)$, and in particular $\dot{\mathcal{E}}(t_0) = \dot{M}(t_0)\mathcal{E}(t_0)$. By Lemma 1.44 the matrix $\dot{M}(t_0)$ is skew-symmetric, which proves Theorem 1.43 for $t = t_0$. Since the choice of t_0 is arbitrary, the statement holds for all t. $\square$

As in the plane the signed curvature determines the curve uniquely, so do the curvature and the torsion for space curves.

Theorem 1.45. *Let $f\colon I \to \mathbb{R}$ and $g\colon I \to \mathbb{R}$ be any smooth functions with $f(t) > 0$ for all t. Then there is a smooth curve $\gamma\colon I \to \mathbb{R}^3$ with the curvature f and the torsion g. This curve is unique up to a rigid motion.*

Sketch of proof. From the theory of ordinary differential equations it follows that the matrix differential equation

$$\begin{pmatrix} \dot{T} \\ \dot{N} \\ \dot{B} \end{pmatrix} = \begin{pmatrix} 0 & f & 0 \\ -f & 0 & g \\ 0 & -g & 0 \end{pmatrix} \begin{pmatrix} T \\ N \\ B \end{pmatrix}$$

has a unique solution with the initial data $T(t_0) = (1,0,0)^\top$, $N(t_0) = (0,1,0)^\top$, $B(t_0) = (0,0,1)^\top$. Similarly to the proof of Lemma 1.44 one can show that the basis $(T(t), N(t), B(t))$ is orthonormal for all t. Define

$$\gamma(t) = \int_{t_0}^{t} T(t)\, dt.$$

Then $T(t) = \dot{\gamma}(t)$, and the above differential equation implies that N and B are the normal and the binormal, and κ and τ are the curvature and the torsion of γ. $\square$

1.9* The total curvature of a space curve

Theorem 1.46 (Fenchel)**.** *The total curvature of a closed space curve is at least 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 1.47. *Let γ be a unit-speed space curve. Then the curve on the unit sphere traced by the vector $T(t) = \dot{\gamma}(t)$ 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 1.48. *The total curvature of a space curve γ is the length of the tangent indicatrix T.*

Proof. Indeed,

$$\int_a^b \kappa(t)\,dt = \int_a^b \|\dot{T}\|\,dt = L(T).$$

□

Lemma 1.49. *The tangent indicatrix intersects each great circle at least twice.*

Proof. Every great circle can be represented as

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

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

$$T(t) \in C_v \Leftrightarrow \langle T(t), v\rangle = 0 \Leftrightarrow \frac{d}{dt}\langle \gamma(t), v\rangle = 0$$

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 T intersects C_v at least twice. □

The rule $v \mapsto v^\perp \cap \mathbb{S}^2$ establishes a bijection between pairs of antipodes on $\mathbb{S}^2$ and the great circles on $\mathbb{S}^2$. The Lebesgue measure on $\mathbb{S}^2$ thus induces a measure on the space of great circles, with the total measure 2π.

Theorem 1.50 (Crofton)**.** *The length of a smooth curve T on the unit sphere is equal half the integral over the space of great circles of the number of intersection points of a great circle with the curve:*

$$L(T) = \frac{1}{2}\int |T \cap C|\,dC.$$

Proof. For a formal proof, see [Hsi97]. But there is a nice straightforward proof in the case when the curve T is an arc of a great circle. □

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. □

1.10 Tubular neighborhoods and normal coordinates

Lecture 3

We start with the planar analog of the Frenet-Serret formulas. Let $\gamma\colon I \to \mathbb{R}^2$ be a unit-speed planar curve, and let $\nu_s\colon I \to \mathbb{R}^2$ be its signed unit normal. Then $(T(t), N(t)) := (\dot{\gamma}(t), \nu_s(t))$ is a positively oriented orthonormal basis, for every $t \in I$. This family of bases is the Frenet-Serret frame of the planar curve γ.

Theorem 1.51. *The Frenet-Serret frame of a planar curve satisfies the following differential equations.*

$$\dot{T} = \kappa_s N$$
$$\dot{N} = -\kappa_s T$$

Proof. The first equation is the definition of the signed curvature, see Lemma 1.32 and Definition 1.33. The second equation follows from the first by differentiating the equations $\langle N, N\rangle = 1$ and $\langle T, N\rangle = 0$. □

Definition 1.52. *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 $\varepsilon \in \mathbb{R}$. The curve*

$$\gamma_\varepsilon = \gamma + \varepsilon \nu_s$$

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

The following lemma explains the term "parallel".

Lemma 1.53. *The velocity vectors of γ and γ_ε at the corresponding points are parallel.*

Proof. By Theorem 1.51 one has $\dot{\nu}_s = -\kappa_s T$. It follows that

$$\dot{\gamma}_\varepsilon = \dot{\gamma} + \varepsilon\dot{\nu}_s = (1 - \varepsilon\kappa_s)\dot{\gamma},$$

and the lemma is proved. □

Lemma 1.54. *Let $K = \max\{\kappa(t) \mid t \in I\}$ be the maximum curvature of γ. Then for all $|\varepsilon| < \frac{1}{K}$ the parallel curve γ_ε is regular.*

Proof. By the previous lemma one has $\|\dot{\gamma}_s\| = |1 - \varepsilon\kappa_s|$, and the definition of K implies $|\varepsilon\kappa_s| < 1$. Therefore $\|\dot{\gamma}_s\| \neq 0$ everywhere on the interval I. □

Example 1.55. The maximum curvature of the standard parabola $\gamma(t) = (t, t^2)$ is equal to 2, attained at $t = 0$. Figure 9 shows the parallel curves for $\varepsilon \in \{0.1, 0.2, 0.3, 0.4, 0.5\}$. One can notice a kink on the curve for $\varepsilon = 0.5$.

Consider the union of parallel curves γ_ε for all $\varepsilon \in (-r, r)$. For r small enough this union is "well-behaved", as explained in the following theorem.

Theorem 1.56. *Let $I \subset \mathbb{R}$ be a closed interval, and let $\gamma \colon I \to \mathbb{R}^2$ be a simple curve parametrized with unit speed. Then for $r > 0$ small enough the map*

$$\Phi \colon I \times (-r, r) \to \mathbb{R}^2, \quad \Phi(t, \varepsilon) = \gamma_\varepsilon(t)$$

is a diffeomorphism onto its image.

The same holds for simple closed curves.

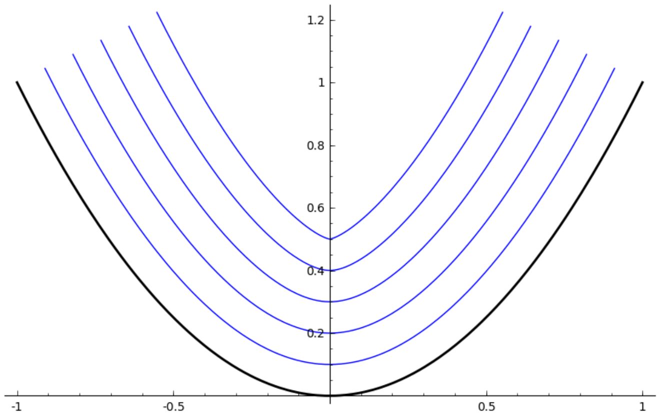

Figure 9: The parabola and its parallel curves.

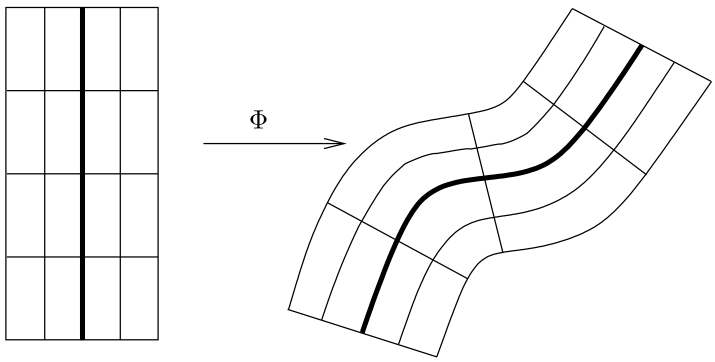

Figure 10: Parametrization of a tubular neighborhood.

The map Φ sends the coordinate net on $I \times (-r, r)$ to the net of parallel curves and straight line segments which intersect these curves orthogonally (they are orthogonal to γ because of $\nu_s \perp \dot{\gamma}$, and orthogonal to γ_ε because of Lemma 1.53). See Figure 10.

Sketch of proof. First, one shows that Φ is a local diffeomorphism. For this, it suffices to prove that the Jacobian determinant of Φ does not vanish. By the Lemmas above one has

$$\det\left(\frac{\partial \Phi}{\partial t}, \frac{\partial \Phi}{\partial \varepsilon}\right) = \det(\dot{\gamma}_\varepsilon, \nu_s) = (1 - \varepsilon\kappa_s)\det(\dot{\gamma}, \nu_s) = 1 - \varepsilon\kappa_s, \tag{2}$$

which does not vanish for all $\varepsilon \in \left(-\frac{1}{K}, \frac{1}{K}\right)$.

A local diffeomorphism may fail to be a bijection globally, see Figure 11. However, from the compactness of I one can derive that for r sufficiently small the map Φ is a bijection. A detailed argument can be found in [Lee13]. □

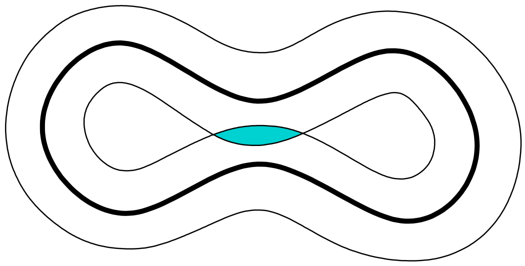

Figure 11: Failure of injectivity for global reasons.

Definition 1.57. *The union of parallel curves for which the map Φ from Theorem 1.56 is a diffeomorphism is called the* r-tubular neighborhood *of γ.*

Definition 1.58. *Let p be a point in a tubular neighborhood of γ. Then (t, ε) such that $p = \gamma_\varepsilon(t)$ are called the* normal coordinates *of p.*

Theorem 1.59 (Weyl tube formula in the plane)**.** *For every non-self-intersecting regular planar curve γ the area of its tubular r-neighborhood is equal to $2rL(\gamma)$.*

For the proof we will use the following fact.

Fact 1.60. *Let $\Phi\colon U \to V$ be a diffeomorphism between two domains in $\mathbb{R}^n$. Then*

$$\mathrm{vol}(V) = \int_U \det(d\Phi(x))\,dx,$$

where $d\Phi(x)$ is the Jacobian matrix of Φ at $x \in U$.

This fact should be known from a multivariable calculus course and is based on the observation that the determinant of the Jacobian of Φ is the volume scaling factor under the map Φ.

Proof of Theorem 1.59. Use the normal coordinates on the tubular neighborhood $N_r(\gamma)$. We have computed the Jacobian in (2). Now integrate:

$$\begin{aligned}\mathrm{area}(N_r(\gamma)) = \int\limits_{I\times(-r,r)} (1 - \varepsilon\kappa_s(t))\,d\varepsilon\,dt &= \int\limits_a^b \left(\int\limits_{-r}^r (1 - \varepsilon\kappa_s(t))\,d\varepsilon\right) dt \\ &= \int\limits_a^b 2r\,dt = 2r(b-a) = 2rL(\gamma).\end{aligned}$$

□

The tubular neighborhood can be defined for space curves. If the curvature of a curve does not vanish, then one can introduce the normal coordinates $(t, \varepsilon_1, \varepsilon_2)$ by setting $\Phi(t, \varepsilon_1, \varepsilon_2) = \gamma(t) + \varepsilon_1 N(t) + \varepsilon_2 B(t)$. If the curvature vanishes at some point, then the Frenet-Serret frame is not defined. This is not really a problem because one can choose any smooth family of orthonormal bases $(E_1(t), E_2(t))$ of $T(t)^\perp$ and put

$$\Phi(t, \varepsilon_1, \varepsilon_2) = \gamma(t) + \varepsilon_1 E_1(t) + \varepsilon_2 E_2(t).$$

The map Φ is a again a diffeomorphism when restricted to a small neighborhood of $I \times \{(0,0)\}$. An r-tubular neighborhood is the image of $I \times B_r$, where $B_r = \{(\varepsilon_1, \varepsilon_2) \mid \varepsilon_1^2 + \varepsilon_2^2 \leq r^2\}$.

Weyl's tube formula also holds:

Theorem 1.61 (Weyl tube formula for curves)**.** *For every non-self-intersecting regular curve γ in $\mathbb{R}^3$ the volume of its tubular r-neighborhood is equal to $\pi r^2 L(\gamma)$.*

We leave the proof to the reader.

1.11 Evolutes and involutes

Definition 1.62. *The* osculating circle *of a curve at a given point is the circle which is tangent to the curve at this point and has the same signed curvature (under the assumption that the circle and the curve are oriented accordingly).*

In other words, if γ is a regular curve, and $\nu_s(t)$ is the signed normal to γ at the point $\gamma(t)$, then the osculating circle at $\gamma(t)$ has center $\gamma(t) + \frac{1}{\kappa_s(t)}\nu_s(t)$ and radius $\frac{1}{\kappa(t)}$. These are also called the *curvature center* and the *curvature radius* at $\gamma(t)$. Intuitively, an osculating circle is a circle through three infinitely close points on the curve (similarly to a tangent being a line through two infinitely close points). It can be shown that, if $\gamma(t)$ is not a point of local maximum or minimum of curvature, then the osculating circle at $\gamma(t)$ separates the curve locally. See Figure 12.

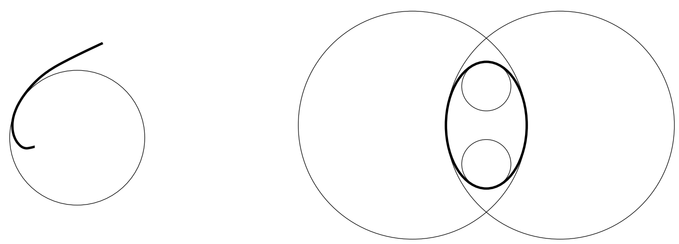

Figure 12: Osculating circles locally separate the curve except at the local extrema of curvature.

The family of osculating circles has also the following property.

Theorem 1.63 (Tait–Kneser)**.** *The osculating circles of a curve with a monotonic curvature are pairwise disjoint and nested.*

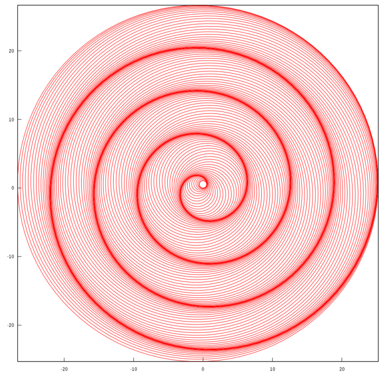

Figure 13: The osculating circles of an Archimedean spiral. By Adam majewski - Own work, CC BY-SA 4.0, https://commons.wikimedia.org/w/index.php?curid=79262500

See Figure 13 for an illustration.

Given a family C_t of curves (lines, circles,...), an *envelope* of C_t is a curve γ which is tangent to C_t at time t. A curve is an envelope of the family of its osculating circles (and also of the family of its tangents). If only the curves C_t are drawn but not γ, then one can still "see" the curve: it lies where the curves C_t become very dense. The spiral in Figure 13 is actually not drawn. Another word for an envelope is *caustic* ("burning"). If C_t is a family of rays of light reflected or refracted by an object, then these rays concentrate along the caustic.

Note however that not every family of curves has an envelope, and some have several envelopes.

Definition 1.64. *The curve formed by the centers of the osculating circles of γ is called the* evolute *of γ.*

We denote the evolute of γ by $\mathcal{E}_\gamma$. By definition one has

$$\mathcal{E}_\gamma(t) = \gamma(t) + \frac{1}{\kappa_s(t)}\nu_s(t). \tag{3}$$

Lemma 1.65. *The evolute of γ consists of the singular points of curves parallel to γ.*

Proof. Indeed, as we have seen in the proofs of Lemma 1.53 and 1.54, the parallel curve γ_ε is singular at t if and only if $\varepsilon = \frac{1}{\kappa_s(t)}$. The corresponding point of γ_ε is exactly the center of the osculating circle of γ at t. $\square$

See Figure 14 for an illustration.

Theorem 1.66. *The tangents to the evolute $\mathcal{E}_\gamma$ go through the corresponding points of γ and are orthogonal to γ. (In ohter words, the evolute is an envelope of the lines orthogonal to the curve.) Singular points of $\mathcal{E}_\gamma$ correspond to critical points of the curvature of γ.*

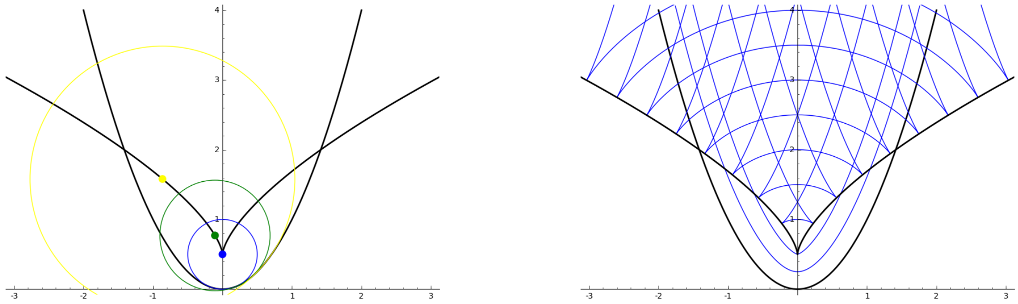

Figure 14: Left: the evolute of the parabola. Right: the evolute as the set of singular points of parallel curves.

Proof. Differentiate the parametrization (3) under the assumption that γ has unit speed. Using $\dot{\nu}_s = -\kappa_s \dot{\gamma}$ one obtains

$$\dot{\mathcal{E}}_\gamma = \dot{\gamma} + \frac{d}{dt}\left(\frac{1}{\kappa_s}\right)\nu_s + \frac{1}{\kappa_s}\dot{\nu}_s = \frac{d}{dt}\left(\frac{1}{\kappa_s}\right)\nu_s.$$

Thus the tangent of $\mathcal{E}_\gamma$ at $\mathcal{E}_\gamma(t)$ is parallel to the normal of γ at $\gamma(t)$. Since $\mathcal{E}_\gamma(t) = \gamma(t) + \lambda\nu(t)$, it follows that the tangent of the evolute passes through the corresponding point of the curve and is orthogonal to the curve there.

The above formula also shows that singular points of $\mathcal{E}_\gamma$ correspond to the points of γ where the derivative of the curvature vanishes. □

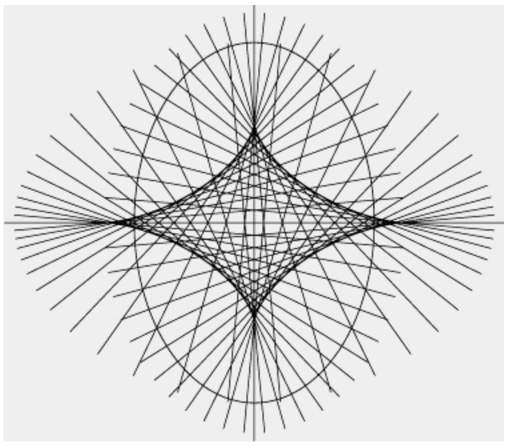

Figure 15: Lines orthogonal to an ellipse.

Figure 15 shows a family of lines orthogonal to an ellipse. It allows you to see the evolute of the ellipse as well as the curvature centers at the extrema of curvatures.

Remark 1.67. At the points where the curvature of a curve vanishes (in particular, at the inflection points which are the points where the signed curvature changes its sign) the osculating circle degenerates into the tangent line. The normal at such a point is an asymptote of the evolute.

Let us describe another way of producing a new curve from an old one. Imagine a string wrapped along a curve. When the string is being unwrapped, its endpoint describes a curve, called an *involute* of the initial curve. In order to derive the equation of an involute, assume the following:

- The curve is parametrized as $\gamma\colon I \to \mathbb{R}^2$.
- The string is wrapped onto $\gamma\left(I \cap [t_0, +\infty)\right)$ so that its free end is at $\gamma(t_0)$.

Then at the moment when the string touches the curve at the point $\gamma(t)$ the unwrapped piece of the string has the length $L_{t_0}^t(\gamma)$ (the length of $\gamma\mid_{[t_0,t]}$). Thus we arrive at the following formal description of involutes.

Definition 1.68. *The* involute *with the starting point $\gamma(t_0)$ of a curve γ is the parametric curve given by*

$$\mathcal{I}_{\gamma,t_0}(t) = \gamma(t) - L_{t_0}^t(\gamma)\frac{\dot\gamma(t)}{\|\dot\gamma(t)\|}, \quad t \in [t_0, +\infty).$$

In particular, if γ has unit speed, then

$$\mathcal{I}_{\gamma,t_0}(t) = \gamma(t) - (t - t_0)\dot\gamma(t).$$

See Figure 16 for an illustration.

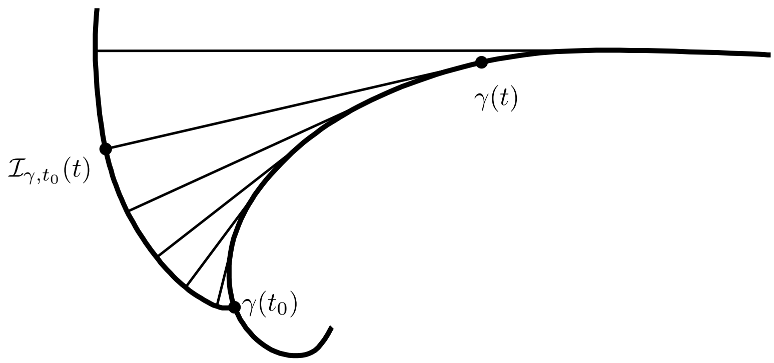

Figure 16: Construction of an involute $\mathcal{I}_{\gamma,t_0}$ of a curve γ.

Example 1.69. The involutes of logarithmic spirals are logarithmic spirals again.

Lemma 1.70. *The tangents to γ intersect every involute of γ orthogonally.*

Proof. Without loss of generality assume that γ has unit speed. The tangent to γ at $\gamma(t_0)$ passes through the point $\mathcal{I}_{\gamma,t_0}(t)$ of the involute. Compute the velocity vector of the involute at this point:

$$\dot{\mathcal{I}}_{\gamma,t_0} = \dot\gamma - (\dot\gamma + (t - t_0)\ddot\gamma) = -(t - t_0)\ddot\gamma. \tag{4}$$

One has $\ddot\gamma \perp \dot\gamma$, and the lemma is proved. □

The previous lemma implies that the lines orthogonal to an involute are tangent to the curve. So it looks like the curve is the evolute of any of its involutes. We cannot yet claim this for sure because we only know that the evolute is an envelope of normals, but cannot say that an envelope of normals is the evolute... So let us give a formal proof of this theorem.

Theorem 1.71. *Let γ be a curve with nowhere vanishing curvature. Then the evolute of an involute of γ is γ itself.*

We need a lemma.

Lemma 1.72. *Let γ be a unit-speed curve. Denote its involute starting at $\gamma(t_0)$ by Γ. Then the curvature of Γ at the point $\delta(t)$ is equal to*

$$\kappa^\Gamma(t) = \frac{1}{t - t_0}.$$

Besides, the signed curvatures of γ and Γ have the same sign.

Proof. Let us use Theorem 1.30: the curvature of Γ is equal to the rotation speed of its tangent relative to the motion speed of the point:

$$\kappa^\Gamma = \frac{|\dot\alpha^\Gamma|}{\|\dot\Gamma\|}.$$

The turning angle of Γ differs from the turning angle of γ by 90°, because tangents to γ are orthogonal to Γ. This implies $|\dot\alpha^\Gamma| = |\dot\alpha^\gamma| = \kappa^\gamma$, as γ is of unit speed. Also this implies that the signed curvatures of γ and Γ are of the same sign. To compute the norm of $\dot\Gamma$, use (4):

$$\|\dot\Gamma\| = (t - t_0)\|\ddot\gamma\| = (t - t_0)\kappa^\gamma.$$

Thus we get

$$\kappa^\Gamma = \frac{\kappa^\gamma}{(t - t_0)\kappa^\gamma} = \frac{1}{t - t_0}.$$

□

Proof of Theorem 1.71. Let $\Gamma = \mathcal{I}_{\gamma,t_0}$ be an involute of a curve γ. Without loss of generality, let γ have unit speed. The evolute of Γ is

$$\mathcal{E}_\Gamma(t) = \Gamma(t) + \frac{1}{\kappa_s^\Gamma(t)}\nu_s^\Gamma(t).$$

Let us assume that the curve γ has a positive turn: $\kappa_s^\gamma > 0$. By Lemma 1.72 $\kappa_s^\Gamma > 0$ as well. By our assumption, the signed normal ν_s^γ is a positive multiple of $\ddot{\gamma}$. In the proof of Lemma 1.70 we have seen that the vector $\dot{\Gamma}$ is a negative multiple of $\ddot{\gamma}$. It follows that the signed normal of Γ equals $\dot{\gamma}$. Substituting this into the above equation and using Lemma 1.72 we get

$$\mathcal{E}_\Gamma(t) = \Gamma(t) + (t - t_0)\dot{\gamma}.$$

But $\Gamma(t) = \gamma(t) - (t - t_0)\dot{\gamma}$ as the involute of γ. Thus $\mathcal{E}_\Gamma(t) = \gamma(t)$, and the theorem is proved. $\square$

1.12* The four-vertex theorem

First we need to learn how to integrate functions on curves.

Definition 1.73. *Let $C \subset \mathbb{R}^n$ be a simple curve, and let $f\colon C \to \mathbb{R}$ be a function on C. One defines the integral of f along C as follows:*

$$\int_C f\, dL := \int_I f(\gamma(t))\, \|\dot{\gamma}(t)\|\, dt,$$

where $\gamma\colon I \to C$ is any regular parametrization of C.

Lemma 1.74. *The value of the integral of a function along a curve does not change under reparametrization.*

Proof. if $\varphi\colon J \to I$ is a diffeomorphism between two intervals, and one puts $\delta = \gamma \circ \varphi$, then one has

$$\int_J f(\delta(s))\, \|\dot{\delta}(s)\|\, ds = \int_J f(\gamma(\varphi(s)))\, \|\dot{\gamma}(\varphi(s))\| \varphi'(s)\, ds = \int_I f(\gamma(t))\, \|\dot{\gamma}(t)\|\, dt$$

$\square$

The expression $dL = \|\dot{\gamma}(t)\|$ is called the *line element.*

Remark 1.75. In contrast to the integral of a 1-form along a curve, the integral of a function along a curve does not change its sign under orientation-reversing reparametrizations.

Definition 1.76. *A simple closed curve is called* convex *if it lies on one side from each of its tangents.*

It is possible to show (but we do not do it) that the signed curvature of a convex curve has a constant sign (it can however vanish at some points or on some intervals). In this section we will always parametrize a convex curve in such a way that its signed curvature is non-negative. This implies that the signed normal is pointing inside the figure bounded by the curve.

Theorem 1.77 (Minkowski formulas in dimension one)**.** *For every smooth closed convex curve $C \subset \mathbb{R}^2$ the following identities hold:*

$$\int_C \kappa\nu \, dL = 0, \qquad \int_C \nu \, dL = 0.$$

Here ν is the field of outward pointing unit normals along the curve.

(Note that we are integrating a vector-valued function here.)

Proof. Let $\gamma\colon [a,b] \to \mathbb{R}^2$ be a unit-speed parametrization of C with non-negative signed curvature. Due to $\ddot{\gamma} = \kappa_s \nu_s = -\kappa\nu$ one has

$$\int_C \kappa\nu \, dL = -\int_a^b \ddot{\gamma} \, dt = -\dot{\gamma}(b) + \dot{\gamma}(a) = 0,$$

which proves the first Minkowski formula.

To prove the second formula, recall the divergence theorem:

$$\int_D \operatorname{div} X \, dx = \int_{\partial D} \langle X, \nu \rangle \, dL$$

(the integral of the divergence over a domain is equal to the flow through the boundary of the domain). For D we take the figure bounded by C, and for X take any constant vector field. As the divergence of a constant vector field vanishes, one has

$$0 = \int_{\partial D} \langle X, \nu \rangle \, dL = \left\langle X, \int_{\partial D} \nu \, dL \right\rangle.$$

Since X is an arbitrary vector, it follows that the integral of the normal vector field along the curve vanishes. $\square$

Remark 1.78. The second Minkowski formula has the following physical interpretation. Imagine a (two-dimensional) balloon. The air inside the balloon exerts a constant pressure at each point of the membrane in the direction normal to the membrane. Thus the integral of the normal vector field is the total force exerted by the air. This force must be zero, otherwise the balloon will move just by itself.

The first Minkowski formula will be used in our proof of the four-vertex theorem.

Definition 1.79. *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 1.80 (Four-vertex theorem)**.** *Every simple closed smooth curve has at least four vertices.*

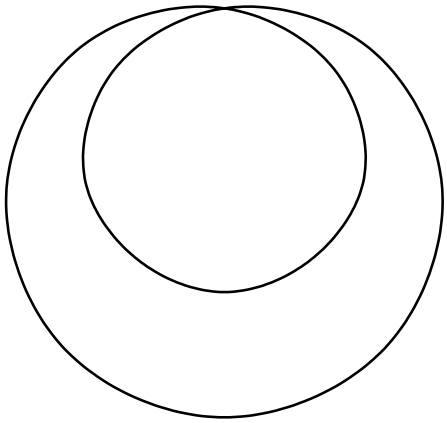

Figure 17: A smooth closed curve with two extrema of curvature.

The simplicity assumption is essential. Figure 17 shows a self-intersecting curve with two vertices.

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

Lemma 1.81. *For every smooth convex closed curve parametrized with the unit speed the following identity holds:*

$$\int_C \dot{\kappa}\iota \, dL = 0,$$

where $\iota\colon C \to \mathbb{R}^2$ *is the inclusion map* $\iota(p) = p$.

Proof. If $\gamma\colon [a,b] \to \mathbb{R}^2$ is a unit-speed parametrization of C, then the integral in the lemma is equal to

$$\int\limits_a^b \dot{\kappa}\gamma(t)\, dt.$$

Apply integration by parts:

$$\int\limits_a^b \dot{\kappa}\gamma \, dt = -\int\limits_a^b \kappa\dot{\gamma}\, dt.$$

Let $J\colon \mathbb{R}^2 \to \mathbb{R}^2$ be the rotation by $90°$. Then $\dot{\gamma} = -J(\nu)$, and one has

$$-\int\limits_a^b \kappa\dot{\gamma}\, dt = \int\limits_a^b \kappa J(\nu)\, dt = J\left(\int\limits_a^b \kappa\nu\, dt\right) = J(0) = 0,$$

due to the first Minkowski formula. □

Proof of the four-vertex theorem. Assume the contrary. Let C be a curve that has only two vertices: the global minimum of the curvature and the global maximum of the curvature. The line ℓ through the vertices cuts C into two arcs C_+ and C_-, where C_+ runs from the minimum point of the curvature to the maximum point, and C_- runs from the maximum to the minimum. (We assume that the curve is oriented counterclockwise.)

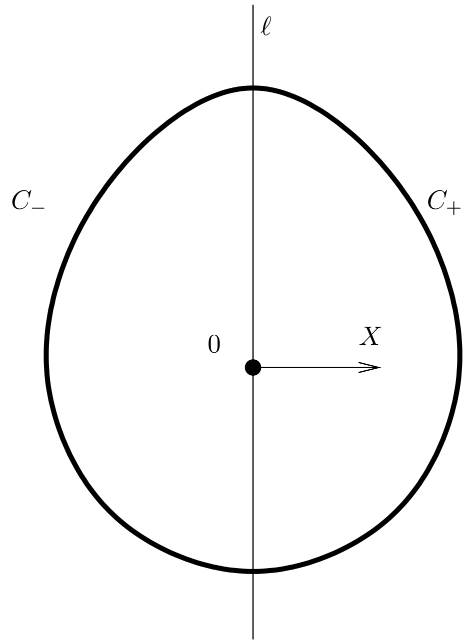

Figure 18: To the proof of the four-vertex theorem.

Translate the curve so that $0 \in \mathbb{R}^2$ lies on the line ℓ. Let $X \in \mathbb{R}^2$ be a vector orthogonal to ℓ and pointing towards the arc C_+, see Figure 18. One has

$$0 = \left\langle X, \int_C \dot{\kappa} \iota \, dL \right\rangle = \int_C \dot{\kappa} \langle X, \iota \rangle \, dL = \int_{C_+} \dot{\kappa} \langle X, \iota \rangle \, dL + \int_{C_-} \dot{\kappa} \langle X, \iota \rangle \, dL$$

The first integrand on the right hand side is non-negative because κ increases monotonically on C_+ and C_+ lies in the positive half-plane with respect to the vector X. The second integrand is also non-negative because here κ decreases while C_- lies in the negative half-plane. Also one can show that the integrands cannot be identically zero. Thus we get $0 > 0$. This contradiction shows that the curve cannot have exactly two vertices. $\square$

2 Surfaces: metric

Lecture 4

In this section we will first clarify the notion of a smooth surface and then we will learn how to measure lengths and areas on surfaces. This represents the *metric* aspect of surface theory; the *curvature* will be the subject of the next section.

2.1 What is a surface?

When we want to answer the question what is a surface we encounter similar problems as we had with curves. On one hand one can try to see surfaces as the images of nice maps, on the other hand one can see them as subsets which locally look like a plane in the space. Let us start with this "look like" approac.

Definition 2.1. *A* smooth surface *in* $\mathbb{R}^3$ *(or a two-dimensional smooth submanifold of* $\mathbb{R}^3$*) is a connected subset* $M \subset \mathbb{R}^3$ *such that for every* $p \in M$ *there is an open subset* $V \subset \mathbb{R}^3$ *such that* $p \in V$ *and a diffeomorphism* $\Phi\colon V \to W$ *of* V *to another open subset* $W \subset \mathbb{R}^3$ *such that* $\Phi(V \cap M) = W \cap \{x_3 = 0\}$.

This definition almost word by word repeats Definition 1.21, see also Figure 1.5 (and the reader can easily guess how a smooth n-dimensional submanifold of $\mathbb{R}^m$ is defined).

There are equivalent characterizations of a smooth surface, again similar to those we had for curves.

Theorem 2.2. *Let* $M \subset \mathbb{R}^3$*. Then the following properties are equivalent.*

1. *M is a smooth surface.*
2. *For every point* $p \in M$ *there is a neighborhood* $V \subset \mathbb{R}^3$ *of* p *and a smooth function* $F\colon V \to \mathbb{R}$ *with nowhere vanishing gradient such that* $V \cap M = F^{-1}(0)$.
3. *For every point* $p \in M$ *there is a neighborhood* $V \subset \mathbb{R}^3$ *of* p*, an open subset* $U \subset \mathbb{R}^2$*, and a smooth map* $\sigma\colon U \to \mathbb{R}^3$ *such that* $\sigma(U) = V \cap M$*, the differential of* σ *(represented by the Jacobian matrix) has rank two everywhere, and* σ *is a homeomorphism onto its image.*

The implications 1. ⇒ 2. and 1. ⇒ 3. are quite straightforward: the corresponding properties of the plane $\{x_3 = 0\}$ are obvious, and the diffeomorphism Φ allows to transfer them to V. The inverse implications can be proved with the help of the inverse function theorem, see e. g. [Laf15, Section 1.5.1].

The third characterization of a smooth surface is an analog of a parametric curve and is the one we will work with. The requirement that the differential of σ is non-degenerate is similar to the requirement $\dot\gamma \neq 0$, and we know already why σ is required to be a homeomoprphism onto its image and not just injective.

Figure 19 shows examples of smooth surfaces in $\mathbb{R}^3$.

We have defined smooth surfaces *without boundary*. If we want to allow boundary, then Definition 2.1 should be modified by adding the possibility that $\Phi(V \cap M) = W \cap \{x_3 = 0, x_2 \geq 0\}$. That is, while surfaces without boundary are modelled on planes, surfaces with boundary are modelled on planes and half-planes. For example, the closed disk $\{x^2 + y^2 \leq 1,\, z = 0\}$ is a surface with boundary while the open disk is a surface without boundary.

A *closed surface* is a compact surface without boundary. In Figure 19 only the torus and the surface of genus 2 are closed surfaces. Every closed surface in $\mathbb{R}^3$ is homeomorphic to a surface of genus $g \geq 0$, which means a sphere with g handles. Torus with a hole

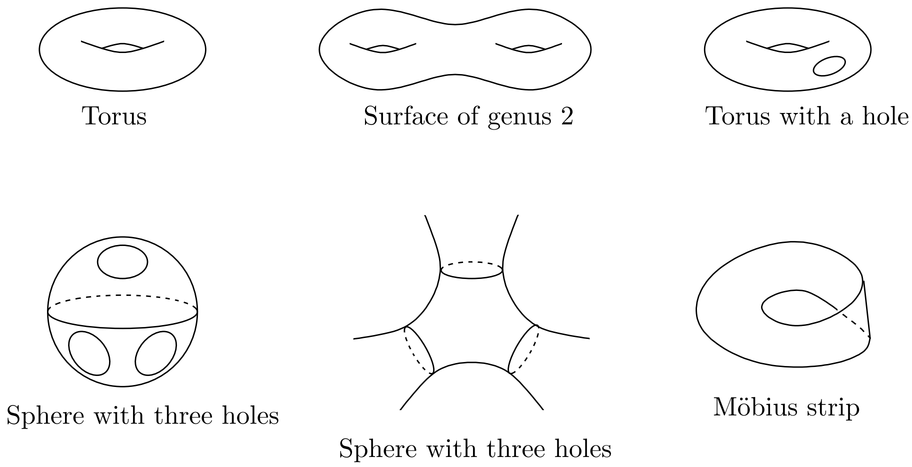

Figure 19: Examples of surfaces

and sphere with three holes are understood as surfaces without boundary (so that the boundary of the hole is a part of the hole and not of the surface), but one can make surfaces with boundary out of them. The same holds for the Möbius strip: there is a version with and a version without boundary.

Example 2.3. The sphere

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

is a smooth surface. Indeed, the sphere is the zero set of the function

$$F(x, y, z) = x^2 + y^2 + z^2 - 1$$

whose gradient does not vanish in the neighborhood $V = \mathbb{R}^3 \setminus \{0\}$ of $\mathbb{S}^2$.

Not every surface can be represented as $M = \{F(x, y, z) = 0\}$ with the gradient of F not vanishing on M. For example, the Möbius strip cannot (think why).

2.2 Surface patches, reparametrization, transition maps

Our main tool in studying surfaces will be, as in the case of curves, the parametrization (the third characterization in Theorem 2.2).

A smooth map $\sigma\colon U \to \mathbb{R}^3$ defined on an open subset $U \subset \mathbb{R}^2$ is called an *immersion* or a *regular map* if its differential has rank 2 everywhere.

Definition 2.4. *Let $M \subset \mathbb{R}^3$ be a smooth surface. A* surface patch *or a* local parametrization *of M is an injective immersion whose image is contained in M.*

Theorem 2.2 says that every smooth surface can be represented as a patchwork of immersions. Note that there are immersions whose images are not surfaces (self-intersecting or others), but for every $x \in U$ there is a smaller neighborhood V such that the restriction of σ to V is a homeomorphism onto the image. So, the local theory of surfaces is in some sense the same as the local theory of immersions.

The global aspect includes not only geometry but also topology: while every smooth curve is homeomorphic to an interval or to a circle, there are infinitely many topological types of smooth surfaces. It is still not too hard to classify the topological types of surfaces, really interesting things start in dimension 3...

In coordinates a surface patch is represented as follows:

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

The differential of σ is a linear map $\mathbb{R}^2 \to \mathbb{R}^3$ represented by the matrix

$$\begin{pmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \\ \frac{\partial z}{\partial u} & \frac{\partial z}{\partial v} \end{pmatrix}.$$

The columns of this matrix are denoted by

$$\sigma_u := \frac{\partial \sigma}{\partial u}, \quad \sigma_v := \frac{\partial \sigma}{\partial v}.$$

Thus the rank two condition on the differential of σ means that the vectors $\sigma_u(u_0, v_0)$ and $\sigma_v(u_0, v_0)$ are linearly independent for every $(u_0, v_0) \in U$.

Example 2.5. Let $N = (0, 0, 1)$ and $S = (0, 0, -1)$ be the north and the south pole of the sphere $\mathbb{S}^2$. The stereographic projection maps every point $p \in \mathbb{S}^2$ different from N to the intersection point of the line Np with the xy-plane. It is given by the formula

$$(x, y, z) \mapsto \left(\frac{x}{1 - z}, \frac{y}{1 - z}, 0 \right).$$

The inverse map is

$$\sigma^N(u, v) = \left(\frac{2u}{u^2 + v^2 + 1}, \frac{2v}{u^2 + v^2 + 1}, \frac{u^2 + v^2 - 1}{u^2 + v^2 + 1} \right).$$

In order to check that this is a surface patch for $\mathbb{S}^2$, compute the partial derivatives:

$$\sigma_u^N = \begin{pmatrix} \frac{2(v^2 - u^2 + 1)}{(u^2 + v^2 + 1)^2} \\ \frac{-4uv}{(u^2 + v^2 + 1)^2} \\ \frac{4u}{(u^2 + v^2 + 1)^2} \end{pmatrix}, \quad \sigma_v^N = \begin{pmatrix} \frac{-4uv}{(u^2 + v^2 + 1)^2} \\ \frac{2(u^2 - v^2 + 1)}{(u^2 + v^2 + 1)^2} \\ \frac{4v}{(u^2 + v^2 + 1)^2} \end{pmatrix}.$$

We have to check that σ_u^N and σ_v^N are linearly independent. Two vectors in $\mathbb{R}^3$ are linearly independent if and only if their cross-product does not vanish. Compute the cross-product:

$$\sigma_u^N \times \sigma_v^N = -4(u^2+v^2+1)\begin{pmatrix} 2u \\ 2v \\ u^2+v^2-1 \end{pmatrix} \neq 0.$$

The surface patch σ^N covers all of $\mathbb{S}^2$ with exception of the north pole. One defines similarly a surface patch σ^S covering the complement of the south pole.

Example 2.6. The well-known latitude-longitude parametrization of the sphere can be written as

$$\sigma\colon (0,\pi)\times(0,2\pi) \to \mathbb{R}^3,$$
$$\sigma(u,v) = (\sin u\cos v, \sin u \sin v, \cos u).$$

(u is the distance from the south pole, that is the latitude in radiants plus $\pi/2$, v is the longitude). Here one has

$$\sigma_u = \begin{pmatrix} \cos u \cos v \\ \cos u \sin v \\ -\sin u \end{pmatrix}, \quad \sigma_v = \begin{pmatrix} -\sin u \sin v \\ \sin u \cos v \\ 0 \end{pmatrix}$$

and

$$\sigma_u \times \sigma_v = \sin u \begin{pmatrix} \sin u \cos v \\ \sin u \sin v \\ \cos u \end{pmatrix}.$$

Again, the vector $\sigma_u \times \sigma_v$ never vanishes, thus the vectors σ_u and σ_v are linearly independent.

This surface patch covers the sphere with the exception of the "zero meridian".

If $\sigma\colon U \to \mathbb{R}^3$ is a surface patch for a surface M, and $\varphi\colon V \to U$ is a diffeomorphism, then the map $\sigma\circ\varphi\colon V\to\mathbb{R}^3$ is also a surface patch for M, called a *reparametrization* of σ.

Reparametrizations appear naturally at the intersections of the images of two patches.

Definition 2.7. *Let $\sigma\colon U\to\mathbb{R}^3$ and $\tau\colon V\to\mathbb{R}^3$ be two patches for the same surface M with non-empty intersection $\Omega = \sigma(U)\cap\tau(V)$. Then the map*

$$\sigma^{-1}\circ\tau\colon \tau^{-1}(\Omega)\to\sigma^{-1}(\Omega)$$

is called the transition map *from σ to τ.*

See Figure 20 for an illustration.

Lemma 2.8. *Transition maps are diffeomorphisms. Thus two intersecting surface patches are reparametrizations of each other on the preimages of their intersection.*

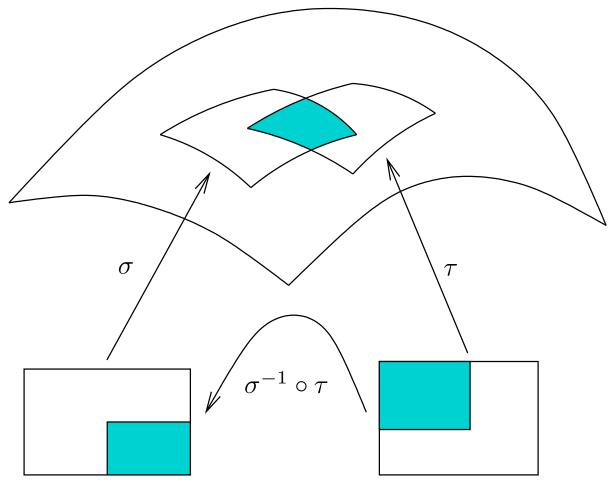

Figure 20: A transition map.

This follows from the definition of a surface (the possibility of straightening).

Example 2.9. The transition map $(\sigma^S)^{-1} \circ \sigma^N$ in Example 2.5 is the inversion in the unit circle

$$\begin{pmatrix} x \\ y \end{pmatrix} \mapsto \frac{1}{x^2+y^2} \begin{pmatrix} x \\ y \end{pmatrix}.$$

This can be proved either by direct computation or with the help of Figure 2.9. Indeed, $OA = \cot \alpha, OB = \tan \alpha \Rightarrow OA \cdot OB = 1$.

2.3 Examples of surfaces

Example 2.10 (Surfaces of revolution)**.** A surface of revolution is obtained by rotating a plane curve (called the *profile curve*) around a line in the plane.

Let the plane be the xz-plane and the axis of rotation the z-axis. Write the profile curve in a parametric form:

$$\gamma(u) = (f(u), 0, g(u)).$$

Then one obtains a parametrization of the surface of revolution by taking v for the rotation angle:

$$\sigma(u, v) = (f(u) \cos v, f(u) \sin v, g(u)).$$

Let us check whether σ is an immersion:

$$\sigma_u = \begin{pmatrix} \dot{f} \cos v \\ \dot{f} \sin v \\ \dot{g} \end{pmatrix}, \quad \sigma_v = \begin{pmatrix} -f \sin v \\ f \cos v \\ 0 \end{pmatrix}$$

$$\sigma_u \times \sigma_v = \begin{pmatrix} -f\dot{g} \cos v \\ -f\dot{g} \sin v \\ f\dot{f} \end{pmatrix}$$

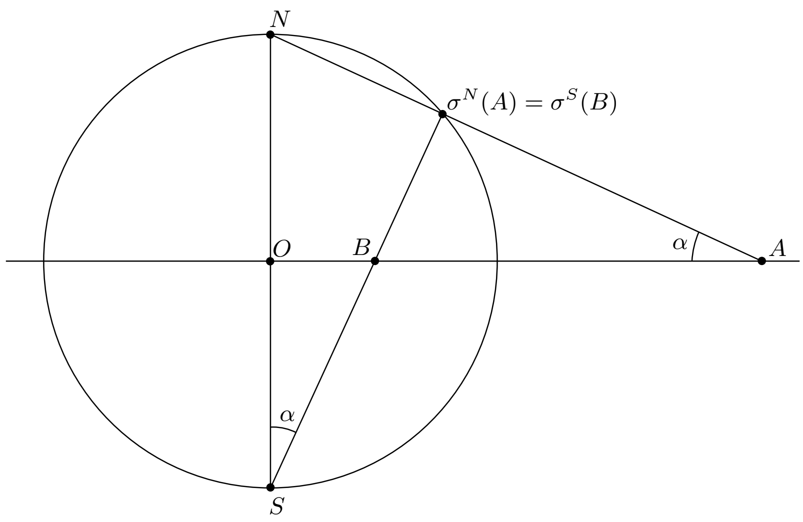

Figure 21: Computing the transition map between stereographic projections from the north and the south pole.

One has $\|\sigma_u \times \sigma_v\| = |f| \cdot \|\dot{\gamma}\|$, thus σ_u and σ_v are linearly independent provided that γ is regular and f does not vanish. If these conditions are fulfilled, then the image of σ is a smooth surface. The image can also be a smooth surface around the points where f does vanish, the geographic parametrization of the sphere being an example. That is, a good surface can also have a bad parametrization.

Definition 2.11. *A* ruled surface *is a union of straight lines called the* rulings *of the surface.*

Let γ be a curve that intersects all of the generators, and let $\delta(u)$ be a direction vector of the line through the point $\gamma(u)$. Then the corresponding ruled surface has a parametrization

$$\sigma(u, v) = \gamma(u) + v\delta(u).$$

Compute:

$$\sigma_u = \dot{\gamma} + v\dot{\delta}, \quad \sigma_v = \delta.$$

There is no simple criterion when σ_u and σ_v are linearly independent. But if $\dot{\gamma}$ and δ are linearly independent, then so are σ_u and σ_v for all sufficiently small v. Thus, if the curve γ is never tangent to the rulings, then in its neighborhood the map σ is an immersion. The map $\sigma|_{I\times(-\varepsilon,\varepsilon)}$ is injective and a homeomorphism onto its image if I is compact and $\gamma|_I$ is injective.

In a weaker sense, a ruled surface is one that contains a line segment through every point.

Example 2.12. An example of a ruled surface is the *helicoid*, see Figure 22. Here the curve γ is a straight line, and the direction of the ruling is orthogonal to this line and

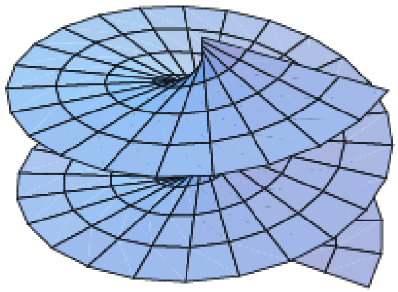

Figure 22: A helicoid.

rotates with a constant speed as one moves along the line. Take

$$\gamma(u) = (0, 0, u), \quad \delta(u) = (\cos u, \sin u, 0)$$

(the ruling rotates with the unit angular speed). This results in the following parametrization of the helicoid:

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

One has

$$\sigma_u = \begin{pmatrix} -v \sin u \\ v \cos u \\ 1 \end{pmatrix}, \quad \sigma_v = \begin{pmatrix} \cos u \\ \sin u \\ 0 \end{pmatrix}$$

$$\sigma_u \times \sigma_v = \begin{pmatrix} -\sin u \\ \cos u \\ -v \end{pmatrix} \neq 0,$$

so that $\sigma\colon \mathbb{R}^2 \to \mathbb{R}^3$ is an immersion. It is easy to see that σ is injective. It is also proper (preimages of compact sets are compact), and therefore a homeomorphism onto its image.

Example 2.13. The one-sheeted hyperboloid $x^2 + y^2 - z^2 = 1$ is a ruled surface. Moreover it is doubly ruled: there are two lines through every point.

Example 2.14. One can construct a ruled Möbius strip by moving and simultaneously rotating a line along a circle.

2.4 Curves on surfaces and the first fundamental form

Let $M \subset \mathbb{R}^3$ be a smooth surface. By a smooth regular curve in M we mean a smooth regular curve $\gamma\colon I \to \mathbb{R}^3$ such that $\gamma(I) \subset M$.

If $\sigma\colon U \to \mathbb{R}^3$ is a surface patch for M, and $\delta\colon J \to U$ is a smooth regular curve, then clearly $\gamma := \sigma \circ \delta$ is a smooth regular curve on M. The converse is also true: locally, every smooth curve in M is the image (under any surface patch) of a smooth curve in the parameter domain. This follows from Definition 2.1: the statement is straightforward for curves in $W \cap (\mathbb{R}^2 \times \{0\})$, and the straightening diffeomorphism Φ respects the smoothness and the regularity of curves.

This allows us to write any smooth regular curve in M locally as

$$\gamma(t) = \sigma(u(t), v(t)), \tag{5}$$

where $(u(t), v(t))$ is a smooth regular curve in U.

Definition 2.15. *A* tangent vector *to a surface M at a point $p \in M$ is the tangent vector at p to a curve on M passing through p. The set of all tangent vectors to M at p is denoted by T_pM.*

Lemma 2.16. *Let $p \in M$ be a point on a smooth surface M. Let $\sigma\colon U \to \mathbb{R}^3$ be a surface patch for M such that $p = \sigma(u_0, v_0)$. Then the set T_pM is the 2-dimensional linear subspace of $\mathbb{R}^3$ spanned by the vectors $\sigma_u(u_0, v_0)$ and $\sigma_v(u_0, v_0)$.*

This lemma allows us to call T_pM the *tangent plane* to M at p. Note however that in reality T_pM does not need to touch M; it is the affine plane $p + T_pM$ that touches M at p.

Proof. Let γ be a smooth curve in M such that $\gamma(t_0) = p$. In a neighborhood of p the curve γ can be represented in terms of the local parametrization σ as (5). One computes the tangent vectors to γ by the chain rule:

$$\dot\gamma = \sigma_u \dot u + \sigma_v \dot v. \tag{6}$$

In particular, the tangent vector to γ at p is a linear combination of the vectors $\sigma_u(u_0, v_0)$ and $\sigma_v(u_0, v_0)$. This proves $T_pM \subset \operatorname{span}\{\sigma_u(u_0, v_0), \sigma_v(u_0, v_0)\}$.

In the opposite direction, take any $X = \lambda\sigma_u(u_0, v_0) + \mu\sigma_v(u_0, v_0)$. Consider the curve

$$\gamma(t) = \sigma(u_0 + \lambda t, v_0 + \mu t).$$

For this curve one has $\dot u = \lambda, \dot v = \mu$, so that its tangent vector at $p = \sigma(u_0, v_0)$ is equal to X. Thus $\operatorname{span}\{\sigma_u(u_0, v_0), \sigma_v(u_0, v_0)\} \subset T_pM$ and the lemma is proved. □

The length of a curve in a surface is computed by the usual formula

$$L(\gamma) = \int_I \|\dot\gamma\|\, dt.$$

However, if a local parametrization σ is fixed and we have to compute the lengths of many curves given in terms of (5), then it is good to have a formula in terms of $u(t)$ and $v(t)$. This is easily obtained by computing the norm of the vector (6):

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

That is, the information from σ needed to compute the lengths of curves is contained in the matrix

$$\begin{pmatrix} E & F \\ F & G \end{pmatrix} := \begin{pmatrix} \langle \sigma_u, \sigma_u \rangle & \langle \sigma_u, \sigma_v \rangle \\ \langle \sigma_u, \sigma_v \rangle & \langle \sigma_v, \sigma_v \rangle \end{pmatrix}. \tag{7}$$

The length of γ is then given by

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

The notations E, F, G are a bit awkward but have a very long tradition.

Definition 2.17. *Let p be a point on a smooth surface M. The* first fundamental form *of M at p is a map I which associates to any two vectors $X, Y \in T_pM$ their inner product:*

$$I(X, Y) := \langle X, Y \rangle.$$

The matrix (7) is the matrix of the first fundamental form in the basis (σ_u, σ_v). Indeed, representing vectors X and Y with respect to this basis as

$$X = X_1\sigma_u + X_2\sigma_v, \quad Y = Y_1\sigma_u + Y_2\sigma_v$$

one obtains

$$\langle X, Y \rangle = \begin{pmatrix} X_1 & X_2 \end{pmatrix} \begin{pmatrix} E & F \\ F & G \end{pmatrix} \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix}.$$

In many books one finds the following notation for the first fundamental form:

$$I = Edu^2 + 2Fdudv + Gdv^2.$$

This can be interpreted by considering u and v as functions of t and setting $du = \dot{u}\, dt$, $dv = \dot{v}\, dt$, so that the expression

$$\sqrt{Edu^2 + 2Fdudv + Gdv^2} = \sqrt{E\dot{u}^2 + 2F\dot{u}\dot{v} + G\dot{v}^2}\, dt$$

(called the *element of length* on a surface) is the integrand in the formula for the length of a curve.

A modern rewriting of the first fundamental form is

$$I = Edu \otimes du + F(du \otimes dv + dv \otimes du) + Gdv \otimes dv,$$

which is an element of the tensor product $(T_pM)^* \otimes (T_pM)^*$, the space of all bilinear functions on the vector space T_pM. We will speak about tensors and tensor fields later.

Here is one more metric concept related to curves.

Definition 2.18. *Let γ and δ be two curves in $\mathbb{R}^n$ meeting at a point. The* angle *between γ and δ at the intersection point is the angle between their tangent vectors at these points.*

In other words, the angle α between γ and δ satisfies

$$\cos\alpha = \frac{\langle\dot\gamma, \dot\delta\rangle}{\|\dot\gamma\|\|\dot\delta\|}.$$

If the curves γ and δ lie in a surface $M \subset \mathbb{R}^3$, then the definition of the angle between them does not change. However, we get an additional tool for computing the angle: if the curves are given in terms of the coordinates on a surface patch, then the inner products and the norms can be computed with the help of the element of length in the patch coordinates:

$$\cos\angle(X, Y) = \frac{I(X, Y)}{\sqrt{I(X, X)I(Y, Y)}}.$$

Example 2.19. The angle α between the coordinate curves $u = u_0$ and $v = v_0$ at the point $\sigma(u_0, v_0)$ satisfies

$$\cos\alpha = \frac{F}{\sqrt{EG}}.$$

In particular, the coordinate curves are orthogonal if and only if $F = 0$.

The standard parametrizations of surfaces of revolution and of the helicoid possess this property.

2.5 Area of a surface

Definition 2.20. *Let $M \subset \mathbb{R}^3$ be a smooth surface, $\sigma \colon U \to \mathbb{R}^3$ be a surface patch for M, and $Q = \sigma(U)$ the corresponding piece of M. The* area *of Q is defined as*

$$\operatorname{area}(Q) := \int_U \|\sigma_u \times \sigma_v\|\, dudv.$$

Here is an intuitive explanation of this formula. The parameter curves $u = \text{const}$ and $v = \text{const}$ cut the surface Q into curvilinear quadrilaterals. For small Δu and Δv the quadrilateral between $u = u_0, u = u_0 + \Delta u, v = v_0$ and $v = v_0 + \Delta v$ is approximately the parallelogram spanned by the vectors $\sigma_u(u_0, v_0)\Delta u$ and $\sigma_v(u_0, v_0)\Delta v$. The area of this parallelogram is $\|\sigma_u \times \sigma_v\|\Delta u\Delta v$, and the sum of the areas of these small parallelograms converges to the above integral as Δu and Δv converge to zero.

Remark 2.21. Recall that the length of a curve is the least upper bound of the lengths of inscribed polygons. For smooth surfaces one is tempted to consider inscribed polyhedra, but this does not work in general. One can inscribe into a (finite) cylinder a sequence of polyhedra whose areas tend to infinity, see Schwarz lantern. The definition works if the surface and the inscribed polyhedra are convex.

Now it makes a lot of sense to prove that the value of the area does not change under reparametrization of Q.

Lemma 2.22. *The area of a surface is independent of the choice of a parametrization.*

Proof. Let $\varphi\colon V \to U$ be a diffeomorphism, and

$$\tau = \sigma \circ \varphi\colon V \to \mathbb{R}^3$$

the reparametrized surface patch. Let (s,t) be Cartesian coordinates in V. By the chain rule one has

$$\begin{aligned}\tau_s &= \sigma_u \frac{\partial u}{\partial s} + \sigma_v \frac{\partial v}{\partial s},\\ \tau_t &= \sigma_u \frac{\partial u}{\partial t} + \sigma_v \frac{\partial v}{\partial t}.\end{aligned}$$

It follows that

$$\tau_s \times \tau_t = \sigma_u \times \sigma_v \left(\frac{\partial u}{\partial s}\frac{\partial v}{\partial t} - \frac{\partial u}{\partial t}\frac{\partial v}{\partial s} \right) = \sigma_u \times \sigma_v \det J_\varphi.$$

Thus the area computed in terms of the parametrization τ equals

$$\int_V \|\tau_s \times \tau_t\|\, dsdt = \int_V \|\sigma_u \times \sigma_v\| \cdot |\det J_\varphi|\, dsdt = \int_U \|\sigma_u \times \sigma_v\|\, dudv,$$

and the lemma is proved. □

Lemma 2.23. *One has* $\|\sigma_u \times \sigma_v\| = \sqrt{EG - F^2}$*, where* E*,* F *and* G *are given by* (7)*.*

Proof. This is a special case of the following general fact: the volume of the parallelepiped spanned by linearly independent vectors $v_1, \ldots, v_m \in \mathbb{R}^n$ is equal to the square root of the determinant of the Gram matrix:

$$\operatorname{vol}(P(v_1,\ldots,v_m))^2 = \det \begin{pmatrix} \langle v_1, v_1\rangle & \ldots & \langle v_1, v_m\rangle \\ \vdots & \ddots & \vdots \\ \langle v_m, v_1\rangle & \ldots & \langle v_m, v_m\rangle \end{pmatrix}$$

□

This leads to a formula for the area of a smooth surface in terms of its first fundamental form:

$$\operatorname{area}(Q) = \int_U \sqrt{EG - F^2}\, dudv.$$

Naturally, the expression $\sqrt{EG-F^2}\, dudv$ (as well as $\|\sigma_u \times \sigma_v\|\, dudv$) is called the *area element* and sometimes denoted dA.

2.6 Conformal parametrizations

A unit-speed parametrization of a smooth curve is a length-preserving map between a line segment and the curve. A length-preserving parametrization of a surface is not always possible. A surface which allows such a parametrization is called *developable.* It can be proved (but it is not easy) that a developable surface contains a line segment through every point, thus it is a ruled surface in a weak sense. Note however that not every ruled surface is developable.

Since a length-preserving parametrization is not always possible, one should try to ask for less.

Definition 2.24. *A surface patch* $\sigma\colon U \to \mathbb{R}^3$ *is called* conformal *if it preserves angles.*

Theorem 2.25. *A surface patch is conformal if and only if its first fundamental form has the form*

$$I = \lambda^2(u,v)\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

Proof. If I is like in the theorem, then

$$\cos\angle(X,Y) = \frac{I(X,Y)}{\sqrt{I(X,X)I(Y,Y)}} = \frac{\langle X,Y\rangle}{\|X\|\|Y\|}.$$

For the opposite direction we have to prove the following: if two positive definite symmetric bilinear forms in a two-dimensional vector space define the same angle measurement, then they are proportional. Denote the two symmetric bilinear forms for short α and β. By assumption one has

$$\frac{\alpha(X,Y)}{\sqrt{\alpha(X,X)\alpha(Y,Y)}} = \frac{\beta(X,Y)}{\sqrt{\beta(X,X)\beta(Y,Y)}} \tag{8}$$

for all X, Y. Let (e_1, e_2) be an orthonormal basis for α (it can be constructed by the Gram-Schmidt method). Substituting $X = e_1$ and $Y = e_2$ in (8) one gets $\beta(e_1, e_2) = 0$. Let

$$\beta(e_1,e_1) = \lambda, \quad \beta(e_2,e_2) = \mu.$$

It suffices to show that $\lambda = \mu$. For this, put $X = e_1$ and $Y = e_1 + e_2$. The left hand side of (8) yields $\frac{1}{\sqrt{2}}$, and the right hand side yields

$$\frac{\beta(e_1, e_1+e_2)}{\sqrt{\beta(e_1,e_2)\beta(e_1+e_2,e_1+e_2)}} = \frac{\lambda}{\sqrt{\lambda(\lambda+\mu)}}.$$

Solving $\frac{1}{\sqrt{2}} = \frac{\lambda}{\sqrt{\lambda(\lambda+\mu)}}$ one gets $\lambda = \mu$, and the theorem is proved. □

Example 2.26. The inverse of the stereographic projection, the patch $\sigma\colon \mathbb{R}^2 \to \mathbb{S}^2 \setminus \{(0,0,1)\}$ from Example 2.5, is conformal. This can be checked by a direct computation from the formulas for σ_u and σ_v computed in that example:

$$\begin{aligned}
\langle \sigma_u, \sigma_u \rangle &= \cdots = \frac{4}{(u^2+v^2+1)^2} \\
\langle \sigma_u, \sigma_v \rangle &= 0 \\
\langle \sigma_v, \sigma_v \rangle &= \cdots = \frac{4}{(u^2+v^2+1)^2}
\end{aligned}$$

Thus the element of length of this surface patch is

$$\frac{4}{(u^2+v^2+1)^2}(du^2+dv^2).$$

A more direct geometric approach to the conformality of the stereographic projection is through a geometric transformation called inversion.

Every smooth surface has a conformal patch in the neighborhood of every point. This was first proved by Gauss for analytic surfaces, and later generalized to surfaces with a very low degree of smoothness.

Conformal surface patches are not unique, but the transition functions between them are controlled by the following theorem.

Theorem 2.27. *A diffeomorphism $U \to V$ between two domains in $\mathbb{R}^2$ is conformal if and only if it is either holomorphic or anti-holomorphic, with respect to the canonical identification $\mathbb{R}^2 \to \mathbb{C}$.*

The coordinates (u, v) produced on a surface by a conformal surface patch are called *isothermal coordinates.*

2.7 Smooth functions and smooth maps

Lecture 5

Definition-Lemma 2.28. *Let $M \subset \mathbb{R}^3$ be a smooth surface and let $p \in M$ be a point on M. A function $f: M \to \mathbb{R}$ is called* smooth at p *if for some (and then for any) surface patch $\sigma: U \to M$ covering p the composition $f \circ \sigma$ is smooth at $\sigma^{-1}(p)$.*

Proof. It has to be proved that if for some patch σ the composition $f \circ \sigma$ is smooth at $\sigma^{-1}(p)$, then for any patch $\tau: V \to M$ the composition $f \circ \tau$ is smooth at $\tau^{-1}(p)$. One has

$$f \circ \tau = (f \circ \sigma) \circ (\sigma^{-1} \circ \tau)$$

(on the definition domain $\tau^{-1}(\sigma(U) \cap \tau(V)) \ni \tau^{-1}(p)$). Since the transition map $\sigma^{-1} \circ \tau$ is smooth, the smoothness of $f \circ \sigma$ at $\sigma^{-1}(p)$ implies the smoothness of $f \circ \tau$ at $\tau^{-1}(p)$. $\square$

Definition 2.29. *A function $f: M \to \mathbb{R}$ is called* smooth *if it is smooth at every $p \in M$, i. e. if the composition $f \circ \sigma$ is smooth for all surface patches σ.*

Definition-Lemma 2.30. *Let $M \subset \mathbb{R}^3$ and $N \subset \mathbb{R}^3$ be smooth surfaces, and let $p \in M$ be a point on M. A map $F\colon M \to N$ is called* smooth at p *if for some (and then for any) pair of surface patches $\sigma\colon U \to M$ and $\tau\colon V \to N$ such that $p \in \sigma(U)$ and $F(p) \in \tau(V)$ the composition $\tau^{-1} \circ F \circ \sigma$ is smooth at $\sigma^{-1}(p)$.*

Proof. Again, the smoothness of the composition is independent of the choice of patches due to the identity

$$(\tau')^{-1} \circ F \circ \sigma' = ((\tau')^{-1} \circ \tau) \circ (\tau^{-1} \circ F \circ \sigma) \circ (\sigma^{-1} \circ \sigma')$$

(with the appropriately restricted domains of definition which, however, contain the point $(\sigma')^{-1}(p)$) and the smoothness of the transition maps. □

Definition 2.31. *A map $F\colon M \to N$ is called* smooth *if it is smooth at every $p \in M$, i. e. if the composition $\tau^{-1} \circ F \circ \sigma$ is smooth for all pairs of surface patches $\sigma\colon U \to M$ and $\tau\colon V \to N$.*

Let us reformulate in the same spirit the smoothness of a curve in a surface as discussed at the beginning of Section 2.4.

Definition-Lemma 2.32. *A curve $\gamma\colon I \to M$ is called* smooth at $t \in I$ *if for some (and then for any) surface patch $\sigma\colon U \to M$ such that $\gamma(t) \in \sigma(U)$ the curve $\sigma^{-1} \circ \gamma$ is smooth at t.*

(Again, $\sigma^{-1} \circ \gamma$ may be defined not on all of I but on a neighborhood of t.)

We now have three kinds of smooth maps: smooth curves $I \to M$, smooth maps between surfaces $M \to N$ and smooth functions $M \to \mathbb{R}$. As one can expect, "composition of smooth is smooth", as detailed in the next lemma.

Lemma 2.33. *Let $\gamma\colon I \to M$ be a smooth curve, $F\colon M \to N$ and $G\colon N \to P$ be smooth maps between smooth surfaces, and $f\colon M \to \mathbb{R}$ and $g\colon N \to \mathbb{R}$ be smooth functions. Then the following holds:*

1. *$F \circ \gamma$ is a smooth curve in N;*
2. *$f \circ \gamma$ is a smooth function on I;*
3. *$G \circ F$ is a smooth map from M to P;*
4. *$g \circ F$ is a smooth function on M.*

Proof. All this is reduced to compositions of smooth maps between domains in $\mathbb{R}^2$ and intervals. For example, if γ is a smooth curve, then by definition $\sigma^{-1} \circ \gamma$ is a smooth curve, and then for any appropriate surface patch τ for N one has

$$\tau^{-1} \circ (F \circ \gamma) = (\tau^{-1} \circ F \circ \sigma) \circ (\sigma^{-1} \circ \gamma)$$

which is smooth because F and γ are smooth. □

Remark 2.34. As an illustration of the above definitions and arguments it is convenient to draw diagrams such as this one:

$$
\begin{tikzcd}
 & M \arrow[r, "F"] & N \arrow[dr, "f"] & \\
I \arrow[ur, "\gamma"] \arrow[dr, dashed] & & & \mathbb{R} \\
 & U \arrow[u, "\sigma"] \arrow[r, dashed] & V \arrow[u, "\tau"'] \arrow[ur, dashed] &
\end{tikzcd}
$$

The curve/map/function $\gamma/F/f$ is smooth if and only if the dashed arrow underneath corresponds to a map which is smooth in the usual sense.

2.8 The differential of a smooth map

Definition 2.35. *Let $F\colon M \to N$ be a smooth map between smooth surfaces. The* differential *of F at a point $p \in M$ is a linear map*

$$dF_p\colon T_pM \to T_{F(p)}N$$

defined as follows. For any $X \in T_pM$ let $\gamma\colon I \to M$ be a smooth curve such that $\gamma(t_0) = p$ and $\dot\gamma(t_0) = X$. Then

$$dF_p(X) := \frac{d(F \circ \gamma)}{dt}(t_0).$$

The above definition can be summarized as $dF(\dot\gamma) = (F \circ \gamma)^{\cdot}$.

Since there are many curves γ with the same tangent vector at $t = t_0$, one has to check that the vector on the right hand side is independent of the choice among these curves. We will do this with the help of surface patches and will derive the matrix of the differential with respect to the patch coordinates.

Lemma 2.36. *The differential of a smooth map is well-defined. Besides, it is a linear map between tangent spaces.*

Proof. Let $\sigma\colon U \to \mathbb{R}^3$ be a surface patch for M containing p, and $\tau\colon V \to \mathbb{R}^3$ be a surface patch for N containing $F(p)$. Denote the coordinates in U by (u_1, u_2) and in V by (v_1, v_2). Since the map F is smooth, one has

$$F(\sigma(u, v)) = \tau(v_1(u_1, u_2), v_2(u_1, u_2))$$

for some smooth functions $v_1(u_1, u_2)$ and $v_2(u_1, u_2)$. Since the curve γ is smooth, one has

$$\gamma(t) = \sigma(u_1(t), u_2(t))$$

and therefore

$$(F \circ \gamma)(t) = \tau(v_1(t), v_2(t)), \tag{9}$$

where $v_i(t) = v_i(u_1(t), u_2(t))$. The tangent vector to γ is

$$\dot{\gamma} = \dot{u}_1\sigma_1 + \dot{u}_2\sigma_2,$$

where we denote $\sigma_i = \sigma_{u_i}$. We are interested in $\dot{\gamma}(t_0)$.

Differentiating (9) with respect to t one obtains

$$\frac{d(F \circ \gamma)}{dt} = \dot{v}_1\tau_1 + \dot{v}_2\tau_2 = \left(\frac{\partial v_1}{\partial u_1}\dot{u}_1 + \frac{\partial v_1}{\partial u_2}\dot{u}_2\right)\tau_1 + \left(\frac{\partial v_2}{\partial u_1}\dot{u}_1 + \frac{\partial v_2}{\partial u_2}\dot{u}_2\right)\tau_2.$$

It shows that the map $\dot{\gamma}(t_0) \mapsto \frac{d(F\circ\gamma)}{dt}(t_0)$ is a linear map from T_pM to $T_{F(p)}N$ given in the bases (σ_1, σ_2) and (τ_1, τ_2) by the matrix

$$\begin{pmatrix} \dfrac{\partial v_1}{\partial u_1} & \dfrac{\partial v_1}{\partial u_2} \\ \dfrac{\partial v_2}{\partial u_1} & \dfrac{\partial v_2}{\partial u_2} \end{pmatrix} \tag{10}$$

evaluated at $(u_1(t_0), u_2(t_0))$. This also shows the independence of the choice of γ, as the image of X depends on its components in the basis (σ_1, σ_2) only. $\square$

Theorem 2.37 (Chain rule for differentials)**.** *Let $F \colon M \to N$ and $G \colon N \to P$ be smooth maps between smooth surfaces. Then for every $p \in M$ one has*

$$d(G \circ F)_p = dG_{F(p)} \circ dF_p.$$

Proof. Let $X \in T_pM$ and let $\gamma \colon I \to M$ be a curve such that $\gamma(t_0) = p$ and $\dot{\gamma}(t_0) = X$. Then one has

$$d(G \circ F)_p(X) = d(G \circ F)_p(\dot{\gamma}(t_0)) = (G \circ F \circ \gamma)^{\cdot}(t_0),$$
$$(dG_{F(p)} \circ dF_p)(X) = dG_{F(p)}((F \circ \gamma)^{\cdot}(t_0)) = (G \circ F \circ \gamma)^{\cdot}(t_0).$$

This proves the theorem. $\square$

2.9 Isometries of surfaces

Definition 2.38. *A smooth map $F \colon M \to N$ between surfaces is called a* path isometry *if it takes every smooth curve in M to a curve of the same length in N:*

$$L(\gamma) = L(F \circ \gamma) \quad \text{for all } \gamma \colon I \to M.$$

Example 2.39. The map $F(u, v) = (\cos v, \sin v, u)$ wraps the plane around the cylinder $\{(x, y, z) \mid x^2 + y^2 = 1\}$. It is intuitively clear (and can be easily proved) that this map is a path isometry.

Let us look for equivalent and more practicable definitions of a path isometry.

Recall that an *inner product space* is a vector space equipped with a positive definite symmetric bilinear form. A tangent plane to a smooth surface is an inner product space: it inherits the inner product from the Euclidean space it lies in.

Theorem 2.40. *A smooth map $F\colon M \to N$ is a path isometry if and only if the differential $dF_p\colon T_pM \to T_{F(p)}N$ is an isometry of inner product spaces for all $p \in M$:*

$$\langle X, Y\rangle = \langle dF_p(X), dF_p(Y)\rangle \quad \text{for all } X, Y \in T_pM.$$

Proof. Let $\gamma\colon [a,b] \to M$ be a curve in M and $F \circ \gamma\colon [a,b] \to N$ the corresponding curve in N. Their lengths are

$$L(\gamma) = \int_a^b \sqrt{\langle \dot\gamma, \dot\gamma\rangle}\, dt, \quad L(F \circ \gamma) = \int_a^b \sqrt{\langle (F\circ\gamma)^{\cdot}, (F\circ\gamma)^{\cdot}\rangle}\, dt$$

By definition of the differential, $(F \circ \gamma)^{\cdot} = dF(\dot\gamma)$, so that

$$L(F \circ \gamma) = \int_a^b \sqrt{\langle dF(\dot\gamma), dF(\dot\gamma)\rangle}\, dt.$$

Thus, if dF_p is an isometry for all p, then $L(\gamma) = L(F \circ \gamma)$.

For the opposite direction look at the restriction of γ to $[a,t] \subset [a,b]$. We have $L_a^t(\gamma) = L_a^t(F \circ \gamma)$. Differentiating the corresponding integrals with respect to the upper limit of integration one obtains $\langle \dot\gamma, \dot\gamma\rangle = \langle dF(\dot\gamma), dF(\dot\gamma)\rangle$ for all t. Since for all $p \in M$ and for all $X \in T_pM$ there is a curve through p with the tangent vector X, this implies $\langle X, X\rangle = \langle dF(X), dF(X)\rangle$ for all $X \in TM$. By a well-known trick

$$\langle X, Y\rangle = \frac{1}{2}(\langle X+Y, X+Y\rangle - \langle X, X\rangle - \langle Y, Y\rangle),$$

a quadratic form determines the associated symmetric bilinear form. It follows that dF is an isometry. □

Tools which allow to check if the differential of a smooth map is an isometry is provided by the following lemma.

Lemma 2.41. *Let $F\colon M \to N$ be a smooth map between smooth surfaces, and let $p \in M$. Let $\sigma\colon U \to \mathbb{R}^3$ and $\tau\colon V \to \mathbb{R}^3$ be surface patches for M and N such that $p \in \sigma(U)$ and $F(p) \in \tau(V)$. Then the following are equivalent.*

1. *The map $dF_p\colon T_pM \to T_{F(p)}N$ is an isometry.*
2. *One has*
$$I_\sigma(u_1, u_2) = J_\Phi^\top(u_1, u_2) I_\tau(v_1, v_2) J_\Phi(u_1, u_2),$$
where $\sigma(u_1, u_2) = p$, $\tau(v_1, v_2) = F(p)$, I_σ and I_τ are the matrices of the first fundamental forms of M and N with respect to the patches σ and τ, and J_Φ is the Jacobi matrix of $\Phi = \tau^{-1} \circ F \circ \sigma$.
3. *One has $I_\sigma = I_{F\circ\sigma}$.*

Proof. Let us prove that 1) and 2) are equivalent. For a vector $X \in T_pM$ denote by $X_\sigma \in \mathbb{R}^2$ its components in the basis (σ_1, σ_2). (Recall that $\sigma_i = \frac{\partial \sigma}{\partial u_i}$.) Then for every $X, Y \in T_pM$ one has

$$\langle X, Y \rangle = X_\sigma^\top I_\sigma Y_\sigma.$$

On the other hand, from the proof of Lemma 2.36 we know that the components of the vector $dF(X)$ in the basis τ_1, τ_2 are given by $dF(X)_\tau = J_\Phi X_\sigma$. Thus one has

$$\langle dF(X), dF(Y) \rangle = (J_\Phi X_\sigma)^\top I_\tau (J_\Phi Y_\sigma) = X_\sigma^\top (J_\Phi^\top I_\tau J_\Phi) Y_\sigma.$$

It follows that $\langle X, Y \rangle = \langle dF(X), dF(Y) \rangle$ holds for all $X, Y \in T_pM$ if and only if $I_\sigma = J_\Phi^\top I_\tau J_\Phi$, and the equivalence of 1) and 2) is proved.

Now we prove that 1) and 3) are equivalent. For this, compute the entries of the matrices I_σ and $I_{F \circ \sigma}$.

$$E_\sigma = \langle \sigma_1, \sigma_1 \rangle,$$
$$E_{F \circ \sigma} = \langle (F \circ \sigma)_1, (F \circ \sigma)_1 \rangle = \langle dF(\sigma_1), dF(\sigma_1) \rangle,$$

and similarly for the F and G entries. One sees that $I_\sigma = I_{F \circ \sigma}$ if and only if $\langle X, Y \rangle = \langle dF(X), dF(Y) \rangle$ holds for $X, Y \in \{\sigma_1, \sigma_2\}$. But since (σ_1, σ_2) is a basis of T_pM, if the latter holds for $X, Y \in \{\sigma_1, \sigma_2\}$, then it holds for any X, Y. □

Example 2.42. Consider the catenoid C parametrized by

$$\sigma(u_1, u_2) = (\cosh u_1 \cos u_2, \cosh u_1 \sin u_2, u_1)$$

and the helicoid H parametrized by

$$\tau(v_1, v_2) = (v_1 \cos v_2, v_1 \sin v_2, v_2).$$

I claim that the map $F : C \to H$ given in the patch coordinates by

$$\Phi(u_1, u_2) = (\sinh u_1, u_2)$$

is a path isometry.

Before doing the computation try to visualize this map. The curves $u_2 = \text{const}$ are the meridians of the catenoid, and for the map to be well-defined, one has to cut the catenoid along one of the meridians, that is to restrict σ to e. g. $\mathbb{R} \times (0, 2\pi)$. Then the image of F is the part of the helicoid between the planes $z = 0$ and $z = 2\pi$. The "equator" $z = 0$ of the catenoid goes to the "axis" of the helicoid $x = y = 0$, and the meridians go to the rulings. The inverse map F^{-1} can be extended to the whole helicoid. The helicoid wraps around the catenoid as the plane around the cylinder.

Let us show that F is an isometry in by checking both condition 2) and the condition 3) from Lemma 2.41. Compute

$$\sigma_1 = (\sinh u_1 \cos u_2, \sinh u_1 \sin u_2, 1), \quad \sigma_2 = (-\cosh u_1 \sin u_2, \cosh u_1 \cos u_2, 0).$$

The first fundamental form of C wrt σ is

$$I_\sigma = \begin{pmatrix} \cosh^2 u_1 & 0 \\ 0 & \cosh^2 u_1 \end{pmatrix}.$$

The first fundamental form of the helicoid wrt τ is

$$I_\tau = \begin{pmatrix} 1 & 0 \\ 0 & v_1^2 + 1 \end{pmatrix}.$$

The Jacobian of Φ is

$$J_\Phi = \begin{pmatrix} \cosh u_1 & 0 \\ 0 & 1 \end{pmatrix}$$

At $(v_1, v_2) = \Phi(u_1, u_2) = (\sinh u_1, u_2)$ one has

$$J_\Phi^\top I_\tau J_\Phi = \begin{pmatrix} \cosh u_1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & \cosh^2 u_1 \end{pmatrix} \begin{pmatrix} \cosh u_1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} \cosh^2 u_1 & 0 \\ 0 & \cosh^2 u_1 \end{pmatrix} = I_\sigma.$$

Thus condition 2) is satisfied.

Now let us compute the first fundamental form of the helicoid wrt $F \circ \sigma$. From

$$F \circ \sigma(u_1, u_2) = (\sinh u_1 \cos u_2, \sinh u_1 \sin u_2, u_2)$$

one computes

$$(F \circ \sigma)_1 = (\cosh u_1 \cos u_2, \cosh u_1 \sin u_2, 0), \quad (F \circ \sigma)_2 = (-\sinh u_1 \sin u_2, \sinh u_1 \cos u_2, 1)$$

and

$$I_{F\circ\sigma} = \begin{pmatrix} \cosh^2 u_1 & 0 \\ 0 & \cosh^2 u_1 \end{pmatrix} = I_\sigma.$$

Thus condition 3) is satisfied as expected.

Isometric surfaces are extremely rare. The above example is very special: the helicoid and the catenoid are so-called conjugate minimal surfaces. (Some information on minimal surfaces will follow.) There is even a one-parameter family of pairwise isometric minimal surfaces containing the helicoid and the catenoid, and you can find on the internet animations showing a continuous isometric deformation of one surface into another.

The following theorem illustrates how difficult it is to isometrically deform surfaces.

Theorem 2.43 (Rigidity of convex surfaces)**.** *If M and N are smooth convex closed surfaces, path isometric to each other, then M and N are congruent.*

Any open subset of the sphere can be isometrically deformed (even if this set is the sphere with a small hole), but already this is not easy to describe explicitly.

2.10 Developable surfaces

As we have indicated at the beginning of Section 2.6, a surface which is path isometric to a piece of the plane is called developable, and such surfaces are rare.

Definition 2.44. *A smooth surface M is called* developable *if there is an open subset $U \subset \mathbb{R}^2$ and a surjective path isometry $U \to M$.*

A piece of the plane can be seen as a smooth surface in $\mathbb{R}^3$, so that the definition of path isometry specializes to this situation.

Remark 2.45. The relation of being path isometric is not an equivalence relation because we do not require a path isometry to be bijective. In formulating the above definition we had a choice between requring a path isometry $U \to M$ or a path isometry $M \to U$. In the first sense a cylinder is developable, in the second it is not. However, locally the difference disappears: every path isometry is a local diffeomorphism, thus being locally developable has two equivalent definitions.

The cylinder over any curve and the cone over any curve are developable (if one removes the singular points, for the cone this is at least the apex). The next theorem describes a set of non-obvious examples.

Theorem 2.46. *Let $\gamma\colon I \to \mathbb{R}^3$ be a space curve with non-vanishing curvature. Put*

$$\sigma\colon I \times \mathbb{R} \to \mathbb{R}^3, \quad \sigma(u, v) = \gamma(u) + v\dot{\gamma}(u),$$

that is, σ is the ruled surface made of the tangents to γ. Then the restriction of σ to $I \times (\mathbb{R} \setminus \{0\})$ is an immersion, and its image is a developable surface.

The surface described in the above theorem is called the *tangent developable* of γ. Figure 23 shows the tangent developable of a helix.

Proof. To check that σ is an immersion away from $v = 0$, compute

$$\sigma_u \times \sigma_v = (\dot{\gamma} + v\ddot{\gamma}) \times \dot{\gamma} = v\ddot{\gamma} \times \dot{\gamma}.$$

Since $v \neq 0$ and γ has non-zero curvature, the right hand side does not vanish. Along γ the tangent developable is singular, it has a so-called cuspidal edge.

Although σ is an immersion, it need not be injective. Thus when we say that its image is developable, we mean this only for pieces which are smooth surfaces. And this holds at least locally: every immersion is locally injective.

We will prove the developability of σ by showing that its first fundamental form coincides with the first fundamental form of a specially parametrized plane.

Without loss of generality, let γ have unit speed (parameter change for γ results in a reparametrization of σ, so that we are still speaking about the same surface). Then we have

$$\begin{aligned}
\langle \sigma_u, \sigma_u \rangle &= \langle \dot{\gamma} + v\ddot{\gamma}, \dot{\gamma} + v\ddot{\gamma} \rangle = \|\dot{\gamma}\|^2 + 2v\langle \dot{\gamma}, \ddot{\gamma} \rangle + v^2\|\ddot{\gamma}\|^2 = 1 + v^2\kappa^2, \\
\langle \sigma_u, \sigma_v \rangle &= \langle \dot{\gamma} + v\ddot{\gamma}, \dot{\gamma} \rangle = \|\dot{\gamma}\|^2 + v\langle \dot{\gamma}, \ddot{\gamma} \rangle = 1, \\
\langle \sigma_v, \sigma_v \rangle &= \|\dot{\gamma}\|^2 = 1.
\end{aligned}$$

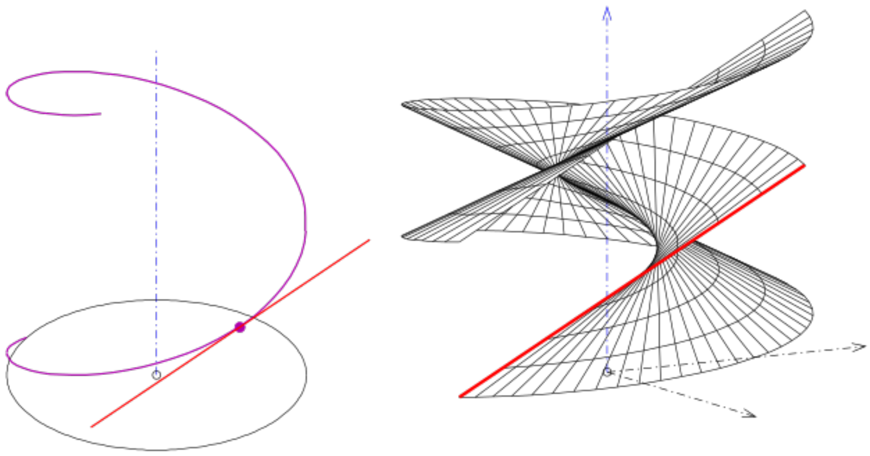

Figure 23: The tangent developable of a helix. Source: `https://commons.wikimedia.org/wiki/File:Schraub-torse-def.svg`, License CC-BY-SA-4.0.

Thus the first fundamental form is

$$I_\sigma = \begin{pmatrix} 1 + v^2\kappa^2 & 1 \\ 1 & 1 \end{pmatrix}.$$

Remarkably, the torsion of γ does not appear in this formula. That is, if we take any other unit-speed curve with the same curvature: $\kappa_\delta(u) = \kappa_\gamma(u)$, and let $\tau(u, v) = \delta(u) + v\dot{\delta}(u)$ be its tangent developable, then we get $I_\sigma = I_\tau$. This means that the map $\Phi(u, v) = (u, v)$ defines a path isometry between the two tangent developables.

By Theorem 1.37 there is a unit-speed curve δ in the plane with the signed curvature κ. Its tangent developable, restricted to small v, is an open subset of the plane: both sheets cover a neighborhood of the convex side of the curve. It follows that the tangent developable of γ is locally developable. $\square$

A path isometry is a map between two surfaces. A related concept is that of an *isometric deformation*. This is a family of surfaces M_t which are diffeomorphically path isometric to each other and smoothly depend on the parameter t. The smooth dependence means that the trajectory of every point $p \in M_0$ under the isometry $F_t \colon M_0 \to M_t$ is a smooth curve. A helix can be smoothly deformed to a circle while preserving its curvature. This defines an isometric deformation of its tangent developable to an annulus.

2.11 Conformal maps

Definition 2.47. *A smooth map $F \colon M \to N$ between two smooth surfaces is called* conformal *if it preserves the angles between curves.*

Theorem 2.48. *The map $F\colon M \to N$ is conformal if and only if $I_{F\circ\sigma} = \lambda^2(u,v)I_\sigma$.*

Proof. Use the arguments from the proof of Theorem 2.25. $\square$

Thus, a conformal map stretches the surface at every point p by the same factor $\lambda(p)$ in all directions.

2.12* Discrete differential geometry

Discrete differential geometry is a relatively new discipline. It studies, among other things, discretization of smooth surfaces. This finds applications in computer graphics and architecture.

A discrete analog of a smooth curve is a polygonal chain. Since the signed curvature of a smooth curve is related to the turning angle, its discrete analog is the oriented angle between two adjacent sides of a polygon, see Figure 24, left. One can easily prove that the total signed curvature of a closed polygon is an integer multiple of 2π.

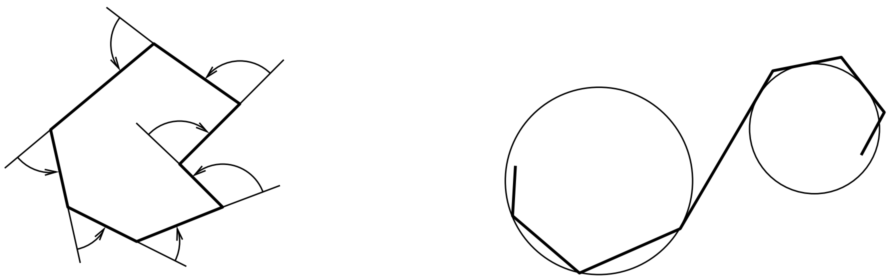

Figure 24: Left: discrete signed curvature. Right: discrete analogs of osculating circles.

An osculating circle of a smooth curve is a circle through "three infinitely close points". For a polygonal chain this can be translated as the circle through three consecutive vertices, but also as the circle tangent to three consecutive edges. Both definitions lead to notions of discrete evolutes. These were studied in detail in [AFI$^+$17].

A discrete analog of a smooth surface is a polyhedral surface. Similarly to isometric deformations of smooth surfaces, one can study isometric bendings of polyhedral surfaces: the faces are viewed as rigid plates connected by rotational joints along the edges. Two classical theorems say that smooth closed convex surfaces and closed convex polyhedra are rigid. There are examples of non-rigid non-convex polyhedra. However, the existence of non-rigid non-convex closed surfaces is an open problem.

A conformal map stretches the surface near a given point p by the same factor $\lambda(p)$ in all directions. Thus small circles on a surface (whatever one means by a circle on a surface) are mapped to "almost circles" on the other surface. This leads to circle packings as discrete analogs of conformal maps. See Figure 25 for a map between two plane regions.

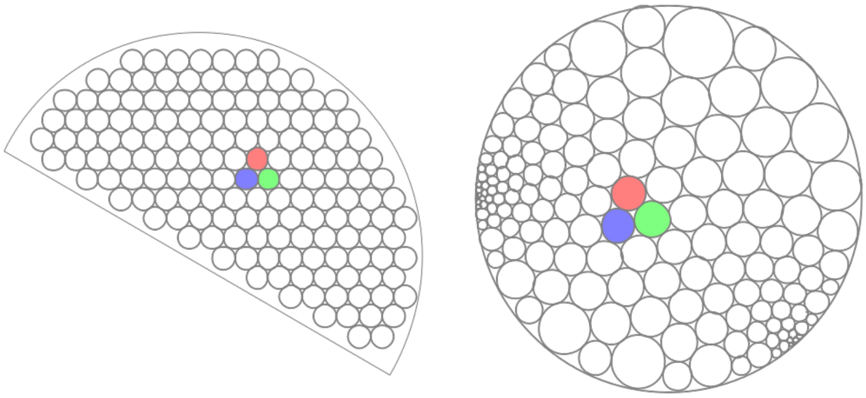

Figure 25: Circle packings approximating a conformal map from a half-disk to a disk. There is a tangency preserving bijection between circles on the left and on the right.

3 Surfaces: curvatures

Lecture 6

3.1 The differential and the gradient of a function

In Section 2.8 we have defined the differential of a smooth map $F\colon M \to N$ between smooth surfaces. For a smooth function $f\colon M \to \mathbb{R}$ the differential can be defined in a similar manner.

Definition 3.1. *Let $f\colon M \to \mathbb{R}$ be a smooth function on a smooth surface. The* differential *of f at a point $p \in M$ is a linear functional*

$$df_p\colon T_pM \to \mathbb{R}$$

defined as follows. For any $X \in T_pM$ let $\gamma\colon I \to M$ be a smooth curve such that $\gamma(t_0) = p$ and $\dot\gamma(t_0) = X$. Then

$$df_p(X) := (f \circ \gamma)^{\cdot}(t_0) = \frac{d(f \circ \gamma)}{dt}(t_0).$$

In other words, $df(\dot\gamma) = (f \circ \gamma)^{\cdot}$.

The well-definedness and the linearity of df can be proved in the same way, by using a surface patch σ.

Closely related to the differential is the *gradient*.

Definition 3.2. *Let $f\colon M \to \mathbb{R}$ be a smooth function. The* gradient *of f at $p \in M$ is a vector* $\operatorname{grad} f_p \in T_pM$ *such that*

$$\langle \operatorname{grad} f_p, X\rangle := df_p(X)$$

for all $X \in T_pM$.

Lemma 3.3. *The components of the gradient of f in the basis (σ_u, σ_v) of T_pM are*

$$I_\sigma^{-1} \begin{pmatrix} (f \circ \sigma)_u \\ (f \circ \sigma)_v \end{pmatrix}.$$

Proof. Let $(\operatorname{grad} f)_\sigma, X_\sigma \in \mathbb{R}^2$ be the components of the gradient and of a tangent vector X in the basis (σ_u, σ_v). One has

$$\langle \operatorname{grad} f, X \rangle = (\operatorname{grad} f)_\sigma^\top I_\sigma X_\sigma.$$

On the other hand, in the coordinates of σ the differential acts on $X = X_u \sigma_u + X_v \sigma_v$ as

$$df(X) = (f \circ \sigma)_u X_u + (f \circ \sigma)_v X_v.$$

This implies that

$$\langle \operatorname{grad} f, X \rangle = df(X) = \begin{pmatrix} (f \circ \sigma)_u \\ (f \circ \sigma)_v \end{pmatrix}^\top X_\sigma.$$

The statement follows. □

Lemma 3.4. *The differential of a function satisfies the product rule:*

$$d(fg) = (df)g + f(dg).$$

That is, for every $p \in M$ and every $X \in T_pM$ one has

$$d(fg)_p(X) = df_p(X)g(p) + f(p)dg_p(X).$$

Proof. Let γ be a curve such that $\gamma(t_0) = p$ and $\dot\gamma(t_0) = X$. Then one has

$$\begin{aligned} d(fg)_p(X) &= d(fg)_p(\dot\gamma(t_0)) = ((f \circ \gamma) \cdot (g \circ \gamma))^{\cdot}(t_0) \\ &= (f \circ \gamma)^{\cdot}(t_0) \cdot g(\gamma(t_0)) + f(\gamma(t_0)) \cdot (g \circ \gamma)^{\cdot}(t_0) = df_p(X)g(p) + f(p)dg_p(X) \end{aligned}$$

□

3.2 Directional derivatives

The image of a vector X under the differential of f is also called the *derivative of f in the direction X*, and denoted

$$D_X f := df(X).$$

One can take directional derivatives not only of functions, but also of vector fields.

Definition 3.5. *Let $M \subset \mathbb{R}^3$ be a smooth surface. A* vector field along M *is a map $Y \colon M \to \mathbb{R}^3$, which is smooth at every point $p \in M$ with respect to some (and then any) surface patch σ. The vector $Y(p)$ is denoted by Y_p.*

If $Y_p \in T_pM$, then the vector field Y is called a tangent vector field *on M.*

Definition 3.6. *Let $X \in T_pM$ be a tangent vector to M at p, and let Y be a vector field along M. The* derivative of Y *in the direction of X is a vector in $\mathbb{R}^3$ given by*

$$D_XY = (Y \circ \gamma)^{\cdot}(t_0),$$

where γ is a curve on M such that $\gamma(t_0) = p$, $\dot{\gamma}(t_0) = X$, and the derivative on the right hand side is taken at t_0.

We will use the notations X, Y sometimes for individual tangent vectors in some T_pM and sometimes for vector fields along M or along some neighborhood of $p \in M$. The notation X_p for an individual vector (value of X at p) will be used sometimes but not always, in order not to overcrowd the formulas.

If X and Y are vector fields, X tangent to M, and Y along M, then D_XY is a vector field along M. Observe that the value of D_XY at a point p depends on the value of X at p only but requires the knowledge of Y in a neighborhood of p. (This is clear from the formula $(D_XY)_p = D_{X_p}Y$ which uses consequent/pedantic notations.) Further properties of the directional derivatives of vector fields are given in the next lemma.

Lemma 3.7. *The map $(X, Y) \mapsto D_XY$ has the following properties.*

1. *It is $C^\infty(M)$-linear in X:*

$$D_{X_1+X_2}Y = D_{X_1}Y + D_{X_2}Y, \quad D_{fX}Y = fD_XY.$$

2. *It is additive in Y and satisfies the product rule:*

$$D_X(Y_1 + Y_2) = D_XY_1 + D_XY_2, \quad D_X(fY) = (D_Xf)Y + fD_XY.$$

Proof. Since the value of D_XY at p depends on the value of X at p only, the $C^\infty(M)$-linearity in X is simply the pointwise $\mathbb{R}$-linearity. Namely, let p be any point in M, and let γ be a curve such that $\gamma(0) = p$ and $\dot{\gamma}(0) = X_p$. Let further $f(p) = c$. Then one has $(D_{fX}Y)_p = D_{cX_p}Y$. We need a curve through p with the tangent vector cX_p there. For this purpose one can take $\delta(t) = \gamma(ct)$. Then one computes

$$D_{f(p)X_p}Y = D_{cX_p}Y = (Y \circ \delta)^{\cdot}(0) = c(Y \circ \gamma)^{\cdot}(0) = cD_{X_p}Y = f(p)D_{X_p}Y.$$

The pointwise $\mathbb{R}$-linearity in X can also be stated as the linearity of the "differential" $DY : T_pM \to \mathbb{R}^3$.

The additivity with respect to Y is obvious, and the product rule is proved by computations similar to those in the proof of Lemma 3.4. □

Lemma 3.8. *The following two product rules hold:*

$$D_X\langle Y, Z\rangle = \langle D_XY, Z\rangle + \langle Y, D_XZ\rangle,$$
$$D_X(Y \times Z) = (D_XY) \times Z + Y \times (D_XZ).$$

Proof. Here again, one uses the arguments from the proof of Lemma 3.4. □

Note that if both vector fields X and Y are tangent to M, then the vector field D_XY is not necessarily tangent to M.

3.3 The shape operator

Let $\nu \colon M \to \mathbb{R}^3$ be a field of unit normals to M. In fact, it can happen that such a field does not exist (take for example, the Möbius band). Our considerations will be local, and locally a field of unit normals does exist for any surface. Besides, there is a choice between two oppositely directed fields of unit normals.

If $\sigma \colon U \to \mathbb{R}^3$ is a surface patch for M, then one can put

$$\nu(\sigma(u,v)) = \frac{\sigma_u \times \sigma_v}{\|\sigma_u \times \sigma_v\|}.$$

Two surface patches $\sigma \colon U \to \mathbb{R}^3$ and $\tau \colon V \to \mathbb{R}^3$ define the same unit normals on $\sigma(U) \cap \tau(V)$ if and only if the determinant of the transition map $\sigma^{-1} \circ \tau$ is positive everywhere in $\sigma(U) \cap \tau(V)$. Indeed, as we have seen in the proof of Lemma 2.22, one has

$$\tau_s \times \tau_t = \sigma_u \times \sigma_v \det J_{\sigma^{-1} \circ \tau},$$

so that normalization of $\sigma_u \times \sigma_v$ and of $\tau_s \times \tau_t$ produces the same unit vector if and only if $\det J_{\sigma^{-1} \circ \tau} > 0$.

Observe that choosing one of the two unit normals to T_pM is equivalent to choosing an orientation of T_pM. A field of unit normals thus corresponds to a continuous choice of orientations of tangent planes.

Definition 3.9. *A surface in $\mathbb{R}^3$ is* orientable *if it can be covered by surface patches whose pairwise transition maps have positive Jacobian determinants everywhere. Equivalently, a surface is orientable if it has a field of unit normals.*

But, again, even for a non-orientable surface a field of unit normals exists locally.

Definition 3.10. *Let $p \in M$ and let $\nu \colon \Omega \to \mathbb{R}^3$ be a field of unit normals to M in a neighborhood $\Omega \subset M$ of p. Then the* shape operator *of $M \subset \mathbb{R}^3$ at a point p (also called the* Weingarten map*) is the linear operator*

$$S_p \colon T_pM \to T_pM, \quad S_p(X) = -D_X\nu.$$

With S we denote the collection of maps S_p (a field of linear operators). Note that replacing ν by $-\nu$ replaces S by $-S$. If M is orientable, then the field S can be defined continuously over all of M.

The definition claims that $D_X\nu \in T_pM$ for all $X \in T_pM$. This holds due to

$$\langle D_X\nu, \nu\rangle = \frac{1}{2} D_X\langle \nu, \nu\rangle = 0.$$

Example 3.11. The shape operator of the unit sphere is $S_p = -\mathrm{Id}$ for the choice of the outward pointing unit normal, and $S_p = \mathrm{Id}$ for the choice of the inward pointing unit normal. Here Id denotes the identity map.

There is an analogy between the shape operator and the curvature of a curve. The curvature of a curve was defined as the turning speed of its tangent. The vector $S_p(X) = -D_X\nu$ shows how the tangent plane to M rotates when the point of tangency moves from p in the direction of X.

3.4 The second fundamental form

Theorem 3.12. *The shape operator is self-adjoint with respect to the first fundamental form. That is, for all $p \in M$ and all $X, Y \in T_pM$ one has*

$$\langle S(X), Y\rangle = \langle X, S(Y)\rangle. \tag{11}$$

Self-adjoint operators naturally correspond to symmetric bilinear forms. We use this correspondence in the following definition.

Definition 3.13. *The* second fundamental form *of a smooth surface M is a family of symmetric bilinear forms in the tangent planes to M defined by*

$$II(X, Y) := \langle S(X), Y\rangle = -\langle D_X\nu, Y\rangle$$

for every $X, Y \in T_pM$ for some $p \in M$.

Again, inverting the direction of the unit normal field changes the sign of the second fundamental form.

We now prove Theorem 3.12 and at the same time compute the matrix of the second fundamental form.

Proof of Theorem 3.12. Let $p = \sigma(u_0, v_0)$. The vectors $\sigma_u(u_0, v_0) = \frac{\partial\sigma}{\partial u}\big|_{(u_0,v_0)}$ and $\sigma_v = \frac{\partial\sigma}{\partial v}\big|_{(u_0,v_0)}$ form a basis of the tangent plane T_pM. In order to prove that the operator S is self-adjoint it suffices to establish the equality (11) for $X = \sigma_u$ and $Y = \sigma_v$. It will then follow for arbitrary X, Y by linearity. One has

$$S(\sigma_u) = -D_{\sigma_u}\nu.$$

To compute this directional derivative, we must choose a curve in M with the velocity vector $\sigma_u(u_0, v_0)$. An obvious choice is

$$\gamma(t) = \sigma(u_0 + t, v_0).$$

Assume that we have chosen $\nu = \frac{\sigma_u\times\sigma_v}{\|\sigma_u\times\sigma_v\|}$. Then we have

$$D_{\sigma_u}\nu = \frac{\partial}{\partial u}\frac{\sigma_u\times\sigma_v}{\|\sigma_u\times\sigma_v\|} = \frac{\sigma_{uu}\times\sigma_v}{\|\sigma_u\times\sigma_v\|} + \frac{\sigma_u\times\sigma_{vu}}{\|\sigma_u\times\sigma_v\|} - \frac{\|\sigma_u\times\sigma_v\|_u}{\|\sigma_u\times\sigma_v\|^2}\sigma_u\times\sigma_v.$$

It follows that

$$\langle S(\sigma_u), \sigma_v\rangle = -\frac{\langle\sigma_u\times\sigma_{vu}, \sigma_v\rangle}{\|\sigma_u\times\sigma_v\|} = \frac{\det(\sigma_u, \sigma_v, \sigma_{uv})}{\|\sigma_u\times\sigma_v\|} = \langle\sigma_{uv}, \nu\rangle.$$

Similarly,

$$D_{\sigma_v}N = \frac{\sigma_{uv}\times\sigma_v}{\|\sigma_u\times\sigma_v\|} + \text{ terms orthogonal to } \sigma_u,$$

$$\langle\sigma_u, S(\sigma_v)\rangle = -\frac{\langle\sigma_u, \sigma_{uv}\times\sigma_v\rangle}{\|\sigma_u\times\sigma_v\|} = \frac{\det(\sigma_u, \sigma_v, \sigma_{uv})}{\|\sigma_u\times\sigma_v\|} = \langle\sigma_{uv}, \nu\rangle.$$

Thus the shape operator is self-adjoint. Let us finish the calculation of the components of the second fundamental form in the basis (σ_u, σ_v).

$$II(\sigma_u, \sigma_u) = -\langle D_{\sigma_u}\nu, \sigma_u\rangle = \frac{\det(\sigma_u, \sigma_v, \sigma_{uu})}{\|\sigma_u \times \sigma_v\|} = \langle \sigma_{uu}, \nu\rangle,$$
$$II(\sigma_v, \sigma_v) = -\langle D_{\sigma_v}\nu, \sigma_v\rangle = \frac{\det(\sigma_u, \sigma_v, \sigma_{vv})}{\|\sigma_u \times \sigma_v\|} = \langle \sigma_{vv}, \nu\rangle.$$

□

Recall the traditional notation for the entries of the matrix of the first fundamental form:

$$I_\sigma = \begin{pmatrix} E & F \\ F & G \end{pmatrix} = \begin{pmatrix} \langle \sigma_u, \sigma_u\rangle & \langle \sigma_u, \sigma_v\rangle \\ \langle \sigma_u, \sigma_v\rangle & \langle \sigma_v, \sigma_v\rangle \end{pmatrix}.$$

The traditional notation for the second fundamental form is

$$II_\sigma = \begin{pmatrix} L & M \\ M & N \end{pmatrix} = \begin{pmatrix} \langle \sigma_{uu}, \nu\rangle & \langle \sigma_{uv}, \nu\rangle \\ \langle \sigma_{uv}, \nu\rangle & \langle \sigma_{vv}, \nu\rangle \end{pmatrix}.$$

(Keep in mind that this is the matrix we obtain if $(\sigma_u, \sigma_v, \nu)$ form a right-hand basis.)

Lemma 3.14. *Let I and II be the matrices of the first and the second fundamental forms in some basis of T_pM. Then the matrix of the shape operator in the same basis is $I^{-1}II$.*

Proof. For any basis (e_1, e_2) of T_pM there is a surface patch σ around p such that at this point $(\sigma_u, \sigma_v) = (e_1, e_2)$. For any two vectors $X, Y \in T_pM$ one has

$$I(X, Y) = X_\sigma^\top I_\sigma\, Y_\sigma, \quad II(X, Y) = X_\sigma^\top II_\sigma\, Y_\sigma$$

We are looking for the matrix S_σ such that $S(X) = S_\sigma X_\sigma$ for all X. By definition of II one has

$$II(X, Y) = I(X, S(Y)) = X_\sigma^\top I_\sigma\, (S_\sigma Y_\sigma) = X_\sigma^\top (I_\sigma S_\sigma)\, Y_\sigma.$$

Comparing this with the second equation above, we get $II_\sigma = I_\sigma S_\sigma$, which implies $S_\sigma = I_\sigma^{-1} II_\sigma$. □

3.5 Principal curvatures, Gaussian and mean curvature

The following fact should be known from the linear algebra course.

Fact 3.15. *For every self-adjoint operator $A\colon V \to V$ on a finite-dimensional vector space V equipped with a positive definite inner product there is an orthonormal basis of V consisting of eigenvectors of A.*

In our case the positive definite inner product is the first fundamental form in the tangent plane, and the self-adjoint operator is the shape operator.

Definition 3.16. *The eigenvalues of the shape operator are called the* principal curvatures*, and the eigenvectors the* principal curvature directions*.*

Viewed geometrically, when a point on the surface moves in the principal direction, the tangent plane to the surface tilts forward or backward, but not to a side. To get a feel of it, imagine a cylinder and move a point over it. You will easily find the principal directions.

Lemma 3.17. *The principal curvatures are the roots of the equation*

$$\det(II - \kappa I) = 0.$$

If κ_i is a root of the above equation, then the principal curvature directions corresponding to κ_i are found by solving the system of linear equations

$$(II - \kappa_i I)X = 0.$$

Proof. By Lemma 3.14, the matrix of the shape operator is $I^{-1}II$. Thus the eigenvectors and the eigenvalues of S satisfy the equation

$$I^{-1}IIX = \kappa X$$

or, equivalently,

$$(II - \kappa I)X = 0.$$

The lemma follows. □

As we have already noted, changing the unit normal to the opposite vector changes the sign of the shape operator and changes therefore the signs of the principal curvatures. For example, the surface in Figure 26, left, has both principal curvatures negative, and the surface in Figure 26, right, has both principal curvatures positive.

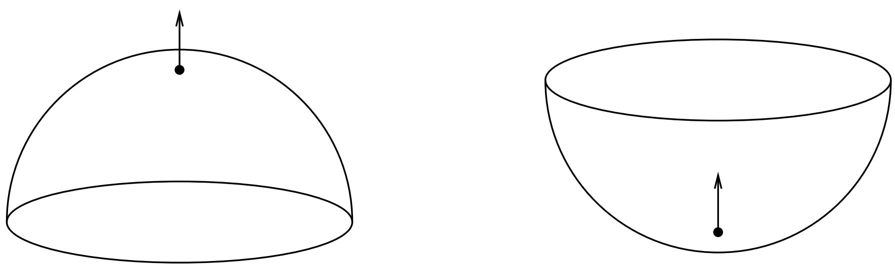

Figure 26: Left: $\kappa_1, \kappa_2 < 0$. Right: $\kappa_1, \kappa_2 > 0$.

Points on a smooth surface can be classified according to the signs of their principal curvatures.

- If the principal curvatures have the same sign, then the point is called *elliptic*.

- If the principal curvatures have opposite signs, then the point is called *hyperbolic.*
- If exactly one of the principal curvatures vanishes, then the point is called *parabolic.*
- If both principal curvatures vanish, then the point is called *planar.*

A point where the principal curvatures coincide is called *umbilic.* On the sphere all points are umbilic. At an umbilic point, all directions are principal curvature direction.

Example 3.18. The torus of revolution contains two circles of parabolic points which separate it into an elliptic and a hyperbolic region.

Definition 3.19. *The determinant of the shape operator at a point $p \in M$ is called the* Gaussian curvature *of M at p and is denoted by K. The half of the trace of the shape operator at a point $p \in M$ is called the* mean curvature *of M at p and is denoted by H:*

$$K := \det S, \quad H := \frac{1}{2}\operatorname{tr} S.$$

In terms of the principal curvatures one has

$$K = \kappa_1\kappa_2, \quad H = \frac{\kappa_1 + \kappa_2}{2}.$$

In practice, one can compute the Gaussian and the mean curvature with the help of the following Lemma.

Lemma 3.20. *The Gaussian and the mean curvatures can be computed from the matrices of the first and the second fundamental forms by the following formulas:*

$$K = \frac{\det II}{\det I}, \quad H = \frac{\det(I, II)}{\det I}.$$

Here $\det(A, B)$ *denotes the* mixed determinant *of two* (2×2)*-matrices, defined as*

$$\det\left(\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}, \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix}\right) = \frac{1}{2}(a_{11}b_{22} + a_{22}b_{11} - a_{12}b_{21} - a_{21}b_{12}).$$

Proof. By Lemma 3.14 the matrix of the shape operator is $I^{-1}II$. It follows that

$$K = \det S = \det(I^{-1}II) = \frac{\det II}{\det I}.$$

For the mean curvature let us compute the matrix of S explicitly:

$$\begin{pmatrix} E & F \\ F & G \end{pmatrix}^{-1} \begin{pmatrix} L & M \\ M & N \end{pmatrix} = \frac{1}{\det I} \begin{pmatrix} G & -F \\ -F & E \end{pmatrix} \begin{pmatrix} L & M \\ M & N \end{pmatrix}$$
$$= \frac{1}{\det I} \begin{pmatrix} GL - FM & GM - FN \\ EM - FL & EN - FM \end{pmatrix}$$

The trace of this matrix is twice $\det(I, II)$ divided by the determinant of I, and the lemma is proved. $\square$

3.6 Curvature of surfaces of revolution

Consider a surface of revolution

$$\sigma(u,v) = (f(u)\cos v, f(u)\sin v, g(u)).$$

$$\sigma_u = \begin{pmatrix} \dot f\cos v \\ \dot f\sin v \\ \dot g \end{pmatrix}, \quad \sigma_v = f\begin{pmatrix} -\sin v \\ \cos v \\ 0 \end{pmatrix}, \quad \sigma_u \times \sigma_v = f\begin{pmatrix} -\dot g\cos v \\ -\dot g\sin v \\ \dot f \end{pmatrix}$$

$$\nu = \frac{\sigma_u \times \sigma_v}{\|\sigma_u \times \sigma_v\|} = \frac{1}{\sqrt{\dot f^2 + \dot g^2}}\begin{pmatrix} -\dot g\cos v \\ -\dot g\sin v \\ \dot f \end{pmatrix}$$

$$\sigma_{uu} = \begin{pmatrix} \ddot f\cos v \\ \ddot f\sin v \\ \ddot g \end{pmatrix}, \quad \sigma_{uv} = \dot f\begin{pmatrix} -\sin v \\ \cos v \\ 0 \end{pmatrix}, \quad \sigma_{vv} = f\begin{pmatrix} -\cos v \\ -\sin v \\ 0 \end{pmatrix}$$

$$I = \begin{pmatrix} \dot f^2 + \dot g^2 & 0 \\ 0 & f^2 \end{pmatrix}, \qquad II = \frac{1}{\sqrt{\dot f^2 + \dot g^2}}\begin{pmatrix} \dot f\ddot g - \ddot f\dot g & 0 \\ 0 & f\dot g \end{pmatrix}$$

It follows that the principal curvatures and directions are

$$\kappa_1 = \frac{\dot f\ddot g - \ddot f\dot g}{(\dot f^2 + \dot g^2)^{\frac{3}{2}}} \qquad \text{for direction } \sigma_u,\ \text{tangent to a profile curve,}$$

$$\kappa_2 = \frac{\dot g}{f\sqrt{\dot f^2 + \dot g^2}} \qquad \text{for direction } \sigma_v,\ \text{tangent to a circular section.}$$

Let us apply these formulas in some concrete situations.

Example 3.21. For the catenoid one has $f(u) = \cosh u$, $g(u) = u$. This yields

$$\kappa_1 = -\frac{1}{\cosh^2 u}, \quad \kappa_2 = \frac{1}{\cosh^2 u}.$$

In particular, the catenoid has zero mean curvature.

Example 3.22. Rotate the tractrix about the z-axis. Choosing the parametrization

$$f(u) = \sin u, \quad g(u) = \log\cot\frac{u}{2} - \cos u$$

one computes

$$\dot g = \sin u - \frac{1}{\sin u}, \quad \ddot g = \cos u + \frac{\cos u}{\sin^2 u}, \quad \dot f^2 + \dot g^2 = \frac{\cos^2 u}{\sin^2 u}.$$

This leads to

$$\kappa_1 = \tan u, \quad \kappa_2 = -\cot u.$$

In particular, the Gaussian curvature of this surface of revolution is everywhere equal to -1. The surface is called a *pseudosphere* and it played a prominent role in the history of the non-Euclidean geometry.

3.7 Surfaces of revolution of constant Gaussian curvature

Without loss of generality assume that the profile curve is parametrized with the unit speed: $\dot{f}^2 + \dot{g}^2 = 1$. Then the formulas for the principal curvatures become simpler:

$$\kappa_1 = \dot{f}\ddot{g} - \ddot{f}\dot{g} \qquad \text{for direction } \sigma_u \text{ tangent to a profile curve}$$

$$\kappa_2 = \frac{\dot{g}}{f} \qquad \text{for direction } \sigma_v \text{ tangent to a circular section}$$

It can be shown that $\dot{f}^2 + \dot{g}^2 = 1$ implies

$$\kappa_1 = \dot{f}\ddot{g} - \ddot{g}\dot{g} = -\frac{\ddot{f}}{\dot{g}}.$$

Thus the Gaussian curvature is equal to

$$K = -\frac{\ddot{f}}{f}.$$

Surfaces of revolution with $K = 1$ correspond to $f(u) = a\cos u + b\sin u$. The reparametrization $u \to u + u_0$ allows to restrict our attention to $f(u) = a\cos u$. For $a = 1$ we obtain $g(u) = \sin u$ and the surface is the unit sphere. For other values of a one has

$$g(u) = \int \sqrt{1 - a^2 \sin^2 u}\, du,$$

which is an elliptic integral. The surfaces look differently for $a > 1$ and for $a < 1$.

Surfaces of revolution with $K = -1$ correspond to $f(u) = ae^u + be^{-u}$ which by parameter shift or its sign change can be reduced to three cases:

$$f(u) = e^u \text{ or } f(u) = a\cosh u \text{ or } f(u) = a\sinh u.$$

In the first case one obtains

$$g(u) = \int \sqrt{1 - e^{2u}}\, du = \sqrt{1 - e^{2u}} - \operatorname{arcosh}(e^{-u}),$$

which is just another parametrization of the tractrix. The other two cases yield two different families of surfaces.

3.8 Gauss map

Lecture 7

Let $\nu\colon M \to \mathbb{R}^3$ be a field of unit normals to M. (Again, our considerations are local, so if M is not orientable, we consider the unit normals over some subset $\Omega \subset M$.) The map ν can be viewed as the map from M to the unit sphere centered at the origin

$$\mathbb{S}^2 = \{x \in \mathbb{R}^3 \mid \|x\| = 1\}.$$

In this interpretation it becomes not a vector field along M but a smooth map between smooth surfaces. In order to stress this distinction, we use different notations for each case.

Definition 3.23. *Let $M \subset \mathbb{R}^3$ be an orientable smooth surface, and $\nu\colon M \to \mathbb{R}^3$ be the corresponding field of unit normals. The smooth map*

$$\Gamma\colon M \to \mathbb{S}^2, \quad \Gamma(p) = \nu_p$$

is called the Gauss map *associated with the chosen orientation of M.*

Lemma 3.24. *For every $p \in M$ one has $T_pM = T_{\Gamma(p)}\mathbb{S}^2$.*

Proof. On one hand, the tangent plane T_pM is the orthogonal complement of the unit normal ν_p. On the other hand, the tangent plane to the sphere is orthogonal to the corresponding radius vector. Thus one has

$$T_pM = \nu_p^\perp = T_{\Gamma(p)}S^2.$$

□

Lemma 3.25. *The shape operator is the negative of the differential of the Gauss map:*

$$S = -d\Gamma.$$

Proof. By definition of the shape operator, for every $X \in T_pM$ one has

$$S(X) = -D_X\nu = -\left.\frac{d}{dt}\right|_{t=0}\Gamma(\gamma(t))$$

for an appropriate curve γ. At the same time, on the right hand side one recognizes the vector $-d\Gamma(X)$. □

Theorem 3.26. *The Gauss map is a local diffeomorphism at $p \in M$ if and only if the Gauss curvature of M at p is different from zero.*

Proof. This follows from a general statement: a smooth map $F\colon M \to N$ is a local diffeomorphism at $p \in M$ if and only if $DF_p\colon T_pM \to T_pN$ is a linear isomorphism. The statement itself is proved by looking at the matrix of the differential with respect to any choice of patches around p and $F(p)$: this matrix is the Jacobian of $\tau \circ F \circ \sigma^{-1}$ and is non-degenerate iff $\tau \circ F \circ \sigma^{-1}$ is a local diffeomorphism, which in turn is equivalent to F being a local diffeomorphism.

Now, $K_p = \det S_p = \det d\Gamma_p$, thus Γ is a local diffeomorphism at p iff $K_p \neq 0$. □

Example 3.27. Let us show that the Gaussian curvature of a tangent developable (see Theorem 2.46) vanishes. The surface is parametrized as $\sigma(u, v) = \gamma(u) + v\dot\gamma(u)$, and its normal can be computed from

$$\sigma_u \times \sigma_v = v\ddot\gamma \times \dot\gamma.$$

One sees that $\Gamma(u, v)$ depends on u only: the unit normals along each ruling $u = \text{const}$ coincide. It follows that the Gauss map is not a local homeomorphism, thus the Gaussian curvature vanishes.

The same observation about the normals shows of course that σ_u is a principal curvature direction for the principal curvature 0.

Theorem 3.28. *If the Gauss map is a diffeomorphism onto its image, then*

$$\int_M K\, dA = \pm\, \mathrm{area}(\Gamma(M)).$$

The integral of a function over a surface is defined first for a domain covered by a single patch:

$$\int\limits_Q f\, dA = \int\limits_{\sigma^{-1}(Q)} (f \circ \sigma)\, \|\sigma_u \times \sigma_v\|\, dudv = \int\limits_{\sigma^{-1}(Q)} (f \circ \sigma) \sqrt{\det I}\, dudv.$$

To integrate a function over all of M, the surface must be cut into domains which can be covered by patches.

From this definition one can derive the following "global change of variables" formula. For every smooth function $f \colon M \to \mathbb{R}$ and every diffeomorphism $F \colon L \to M$ one has

$$\int_M f\, dA^M = \int_L f \circ F |\det(dF)|\, dA^L$$

(where we put the superscripts in order to distinguish between the area elements on M and on L).

With this being said, the proof of Theorem 3.28 is done in two lines.

Proof. Replace integration over M by integration over $\mathbb{S}^2$:

$$\int_M K\, dA^M = \int_{\Gamma(M)} K \circ \Gamma^{-1} |\det(dN^{-1})|\, dA^{\mathbb{S}^2} = \int_{\Gamma(M)} \det S |\det S^{-1}|\, dA^{\mathbb{S}^2} = \pm\, \mathrm{area}(\Gamma(M))$$

(note that if the Gaussian curvature does not vanish, then it has a constant sign over M, and thus we are integrating a constant function ± 1). □

Remark 3.29. It can be shown that for a smooth convex closed surface with everywhere positive Gaussian curvature the Gauss map is a diffeomorphism to $\mathbb{S}^2$. This implies

$$\int_M K\, dA = 4\pi.$$

We will come back to the integral of the Gaussian curvature in a few weeks from a more general perspective.

3.9 Minimal surfaces

A *minimal surface* is a surface whose area cannot be decreased by a continuous deformation. If the surface has boundary, then one assumes that the boundary remains fixed during the deformation (otherwise the surface may shrink). If the surface has no boundary, then one assumes that the deformation is supported on a compact subset (for a similar reason to above if the surface is open and bounded and in order to avoid improper integrals). We consider only smooth deformations of surfaces. Let us give formal definitions.

Let $M \subset \mathbb{R}^3$ be a smooth surface (without boundary).

Definition 3.30. *A* smooth deformation *of a surface M is a family of surfaces $\{M_t\}$, $t \in (-\varepsilon, \varepsilon)$ together with a family of diffeomorphisms $\Phi_t \colon M \to M_t$ such that the map*

$$\Phi \colon M \times (-\varepsilon, \varepsilon) \to \mathbb{R}^3, \quad \Phi(p,t) = \Phi_t(p)$$

is smooth in the sense that for every surface patch σ the map $U \times (-\varepsilon, \varepsilon) \to \mathbb{R}^3$, $(u,v,t) \mapsto \Phi(\sigma(u,v),t)$ is smooth.

A deformation is called compactly supported *if there is a compact subset $Q \subset M$ such that $\Phi_t = \mathrm{id}$ outside of Q.*

The definition implies that every point $p \in M$ moves along a smooth curve $\gamma_p(t) = \Phi_t(p)$. Note also that for every surface patch σ for M the map $\sigma^t = \Phi_t \circ \sigma$ is a surface patch for M_t.

A surface is called *minimal* if its area cannot be decreased through a smooth compactly supported deformation. Since the area of M may be infinite, one has to correct this by saying: the area of any compact piece cannot be decreased by a smooth deformation supported on this piece. Let us formulate a weaker property which relates to the above in the same way as a critical point relates to a local minimum.

A surface is called a *critical point of the area functional* if for every compact subset $Q \subset M$ one has $\frac{d}{dt}\big|_{t=0} \operatorname{area}(\Phi_t(Q)) = 0$ for every smooth deformation of M supported on Q. Clearly, every minimal surface is a critical point of the area functional.

Definition 3.31. *The* velocity vector field *of a smooth deformation (M, Φ) is the vector field*

$$\dot\Phi \colon M \to \mathbb{R}^3, \quad \dot\Phi_p = \dot\Phi(p,0) = \dot\gamma_p(0).$$

Lemma 3.32. *For every smooth deformation of M supported on a compact subset Q covered by a surface patch σ one has*

$$\frac{d}{dt}\bigg|_{t=0} \operatorname{area}(\Phi_t(Q)) = \int_Q \frac{\det(I_\sigma, D^2\langle \sigma, \xi\rangle)}{\det I_\sigma}\, dA,$$

where $\xi(u,v) = \dot\Phi_{\sigma(u,v)}$, and

$$D^2\langle\sigma,\xi\rangle = \begin{pmatrix} 2\langle\sigma_u, \xi_u\rangle & \langle\sigma_u,\xi_v\rangle + \langle\sigma_v,\xi_u\rangle \\ \langle\sigma_u,\xi_v\rangle + \langle\sigma_v,\xi_u\rangle & 2\langle\sigma_v,\xi_v\rangle \end{pmatrix}.$$

Proof. Let us differentiate the formula

$$\operatorname{area}(\Phi_t(Q)) = \int_Q \sqrt{\det I_{\sigma^t}}\, dudv.$$

For this, observe that

$$\sigma^t_u = \sigma_u + t\xi_u + o(t), \quad \sigma^t_v = \sigma_v + t\xi_v + o(t)$$

and compute

$$I_{\sigma^t} = \begin{pmatrix} \langle\sigma^t_u, \sigma^t_u\rangle & \langle\sigma^t_u,\sigma^t_v\rangle \\ \langle\sigma^t_u,\sigma^t_v\rangle & \langle\sigma^t_v,\sigma^t_v\rangle \end{pmatrix} = I_\sigma + tD^2\langle\sigma,\xi\rangle + o(t).$$

Thus one has

$$\det I_{\sigma^t} = \det I_\sigma + 2t \det(I_\sigma, D^2\langle\sigma, \xi\rangle) + o(t) = \det I_\sigma \left(1 + 2t\frac{\det(I_\sigma, D^2\langle\sigma, \xi\rangle)}{\det I_\sigma}\right) + o(t)$$

which implies

$$\sqrt{\det I_{\sigma^t}} = \sqrt{\det I_\sigma}\left(1 + t\frac{\det(I_\sigma, D^2\langle\sigma, \xi\rangle)}{\det I_\sigma}\right) + o(t).$$

It follows that

$$\operatorname{area}(\Phi_t(Q)) = \operatorname{area}(Q) + t\int_Q \frac{\det(I_\sigma, D^2\langle\sigma, \xi\rangle)}{\det I_\sigma}\, dA + o(t),$$

and the lemma is proved. □

Lemma 3.33. *If the velocity vector field of the deformation Φ is orthogonal to M, that is $\dot{\Phi} = f\nu$, then for every deformation supported on Q one has*

$$\left.\frac{d}{dt}\right|_{t=0} \operatorname{area}(\Phi_t(Q)) = -2\int_Q fH\, dA,$$

where H is the mean curvature of M with respect to the normal ν.

Proof. By abuse of notation let $\xi = f\nu$. Then one has

$$\langle\sigma_u, \xi_u\rangle = \langle\sigma_u, f_u\nu + f\nu_u\rangle = f\langle\sigma_u, \nu_u\rangle,$$

and similarly

$$\langle\sigma_u, \xi_v\rangle = f\langle\sigma_u, \nu_v\rangle, \quad \langle\sigma_v, \xi_u\rangle = f\langle\sigma_v, \nu_u\rangle, \quad \langle\sigma_v, \xi_v\rangle = f\langle\sigma_v, \nu_v\rangle.$$

Since $\nu_u = D_{\sigma_u}\nu = -S(\sigma_u)$, one has

$$D^2\langle\sigma, f\nu\rangle = -2II_\sigma.$$

Thus Lemma 3.32 implies that

$$\left.\frac{d}{dt}\right|_{t=0} \operatorname{area}(\Phi_t(Q)) = -2\int_Q f\frac{\det(I_\sigma, II_\sigma)}{\det I_\sigma}\, dA = -2\int_Q fH\, dA.$$

□

Lemma 3.34. *For every smooth deformation of M supported on a compact subset Q covered by a surface patch one has*

$$\left.\frac{d}{dt}\right|_{t=0} \operatorname{area}(\Phi_t(Q)) = \int_Q -2fH\, dA,$$

where $f = \langle\dot{\Phi}, \nu\rangle$ is the normal component of the velocity field of the deformation.

The statement can be globalized, that is the assumption that Q is covered by a surface patch can be omitted, by cutting M into topologically trivial pieces. Note that both f and H change their signs if the unit normal field is replaced by its opposite. Therefore the integrand is well-defined even on non-orientable surfaces.

Proof. We will use the following two properties of the formula from Lemma 3.32. First, the derivative of the area depends only on the velocity vector field of the deformation: if $\dot{\Phi} = \dot{\Psi}$, then the corresponding derivatives coincide. Second, it depends on the velocity vector field linearly and in particular additively. Indeed, the matrix $D^2(\sigma, \xi)$ depends linearly on ξ, and the mixed determinant is bilinear.

Take any smooth deformation Φ_t, and let $X = \dot{\Phi}$ be its velocity vector field. Decompose X into the tangent and the normal components:

$$X = \top X + \perp X, \quad (\top X)_p \in T_pM, \, (\perp X)_p = f\nu.$$

Now construct smooth deformations $\top\Phi_t$ and $\perp\Phi_t$ with the velocity fields $\top X$ and $\perp X$ respectively. The deformation $\perp\Phi_t$ will be very simple:

$$\perp\Phi_t(p) = p + t(\perp X)_p,$$

while for $\top\Phi_t$ we take the so-called flow of the tangent vector field $\top X$. Then all $\top\Phi_t$ are diffeomorphisms from M to M, which implies that the derivative of the area under $\top\Phi_t$ is equal to zero. Thus one has

$$\left.\frac{d}{dt}\right|_{t=0} \operatorname{area}(\Phi_t(Q)) = \left.\frac{d}{dt}\right|_{t=0} \operatorname{area}(\perp\Phi_t(Q)) = \int_Q fH \, dA,$$

and the lemma is proved. $\square$

Theorem 3.35. *A smooth surface is a critical point of the area functional if and only if $H = 0$ at every point.*

Proof. The "if" part follows from the (globalized) Lemma 3.34. Assume that $H(p) \neq 0$ for some $p \in M$. Then H has a constant sign in some neighborhood of p. Take a normal deformation supported in this neighborhood and with $\langle \dot{\Phi}, \nu \rangle$ of a constant sign. Then according to the formula from Lemma 3.34 the derivative of the area with respect to the deformation Φ_t is non-zero, and the surface is not a critical point of the area functional. $\square$

It is of course interesting to ask if a critical point of the area functional is a point of local minimum for all smooth deformations Φ. The answer is yes, but the proof requires more work. Thus minimal surfaces are exactly the surfaces of zero mean curvature.

Example 3.36. The catenoid is a minimal surface, as we have established in Example 3.21. All the surfaces from the helicoid-catenoid family (and the helicoid in particular) are minimal.

Theorem 3.37. *The Gauss map of a minimal surface is conformal.*

Proof. Since $d\Gamma = -S$, the Gauss map is conformal if and only if there is a function $\lambda \colon M \to \mathbb{R}$ such that

$$\langle X, Y \rangle = \lambda^2(p) \langle S(X), S(Y) \rangle \quad \text{for all } X, Y \in T_pM.$$

Let (e_1, e_2) be an orthonormal basis of T_pM which diagonalizes the shape operator. Since the mean curvature of M vanishes, one has

$$S(e_1) = \kappa(p) e_1, \quad S(e_2) = -\kappa(p) e_2.$$

Thus the above equation is satisfied for $X, Y \in \{e_1, e_2\}$ with $\lambda^2(p) = \kappa^2(p) = -K(p)$. Then it is also satisfied for all $X, Y \in T_pM$. $\square$

By reverting the above argument one shows that if the Gauss map is conformal at p, then $|\kappa_1(p)| = |\kappa_2(p)|$. That is, either $H(p) = 0$ or the point p is umbilic.

Conformal maps are closely related to holomorphic functions. There is a formula, due to Enneper and Weierstrass, which constructs a minimal surface out of a pair of holomorphic functions.

Plateau's problem asked if there is a minimal surface with a prescribed boundary (a collection of simple disjoint curves in $\mathbb{R}^3$). Raised in the 18th century, it was completely solved only in the 20th.

A local minimizer is not necessarily a global minimizer. Examples can be constructed by looking at the Plateau problem for two parallel circles of equal radii. Sometimes there is no catenoid spanning these disks, and sometimes there are two different catenoids of different areas. For some values of parameters a pair of disks has a smaller area than any of the two catenoids.

3.10 Surfaces of constant mean curvature

Constant mean curvature (CMC) surfaces are equilibrium surfaces in the presence of different pressure on two sides of the surface. In other words, they are critical points of the area under variations preserving the volumes on either side of the surface.

Theorem 3.38. *If M is a surface of a constant non-zero mean curvature H, then the parallel surface to M at the distance $\frac{1}{2H}$ has a constant Gaussian curvature $4H^2$.*

Conversely, if M is a surface of a constant positive Gaussian curvature K, then the parallel surfaces to M at distances $\pm\frac{1}{\sqrt{K}}$ have constant mean curvatures $\mp\frac{\sqrt{K}}{2}$.

Proof. In the exercises we have derived the following formulas for the principal curvatures of the parallel surface at the distance δ:

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

This implies that

$$K^\delta = \frac{K}{1 - 2\delta H + \delta^2 K}, \quad H^\delta = \frac{H - \delta K}{1 - 2\delta H + \delta^2 K}.$$

If we assume that H is constant and substitute into the first formula $\delta = \frac{1}{2H}$, then the variable K will cancel out and we get $K^\delta = \frac{1}{\delta^2} = 4H^2$. Similarly, if we assume K to be constant and substitute $\delta = \pm\frac{1}{\sqrt{K}}$ into the second formula, then again a cancellation happens which makes the variable H to disappear and we arrive at the formula in the theorem. □

We do not investigate the regularity of the surfaces obtained in the above theorem.

We have studied surfaces of revolution of constant Gaussian curvature in Section 3.7. By applying to them the above construction one obtains a family of CMC surfaces.

CMC surfaces of revolution were classified by Delaunay and have a very nice geometric description. The profile curve of a Delaunay surface is traced by a focus of conics when the conic is rolled on a line. Figure 27 shows one of these surfaces, the unduloid. Rolling a hyperbola produces a self-intersecting surface called the nodoid, and rolling a parabola produces the catenary.

Figure 27: Unduloid, a surface of constant mean curvature. From the collection of models at TU Wien.

4 Extrinsic vs. intrinsic geometry of surfaces

Lecture 8

Intrinsic geometry of surfaces deals with properties and quantities which can be established or measured "without leaving the surface", that is only by measuring lengths of curves on the surface. In particular, isometric surfaces have the same intrinsic geometry. *Extrinsic geometry* takes into account how a surface is bent in the space.

The first fundamental form is an intrinsic quantity: although it comes from the inner product of the ambient space, it is intrinsic because it expresses the lengths of "infinitesimally short curves" on the surface. The first fundamental form also allows to compute the length of any curve on the surface. Thus all intrinsic geometry is contained in the first fundamental form.

By contrast, the shape operator and the second fundamental form are extrinsic by their nature, and so are the principal curvatures and the principal curvature directions. This can be illustrated by taking two isometric surfaces (for example, a piece of the plane and a cylinder) and studying their second fundamental forms.

Being a minimal surface is also an extrinsic property. Although there are isometric families of minimal surfaces, such as the catenoid-helicoid family, a small piece of a minimal surfaces can be isometrically deformed so that its mean curvature becomes non-zero. An example is an isometric deformation of a piece of the catenoid in the class of surfaces of revolution. (We did not study isometric deformations of surfaces of revolution this year.)

Sometimes an extrinsically defined property or quantity turns out to be intrinsic. In this section we will be focusing on the interplay between the intrinsic and extrinsic geometry.

4.1 Curves on surfaces and their curvature

Let $\gamma\colon I \to M$ be a unit-speed curve. Its velocity vector $\dot\gamma(t)$ belongs to the tangent plane $T_{\gamma(t)}M$. Due to $\|\dot\gamma\| = 1$, the acceleration vector $\ddot\gamma$ is orthogonal to $\dot\gamma$. It follows that $\ddot\gamma$ is a linear combination of the vectors ν and $\nu \times \dot\gamma$. Here ν is a unit normal to the surface, and $\nu \times \dot\gamma$ by basic properties of the cross-product is a unit vector in the tangent plane orthogonal to $\dot\gamma$.

Definition 4.1. *Let γ be a unit-speed curve, and let its acceleration vector decompose into a normal and a tangential component as*

$$\ddot\gamma = \kappa_n \nu + \kappa_g (\nu \times \dot\gamma). \tag{12}$$

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

See Figure 28.

As usual, for a non-unit-speed curve its normal and geodesic curvatures are defined as curvatures of its unit-speed reparametrization.

Changing the unit normal ν to the opposite vector changes signs of the normal and the geodesic curvatures.

Since ν and $\nu \times \dot\gamma$ are orthogonal unit vectors, the curvature $\kappa = \|\ddot\gamma\|$ of the curve γ as a space curve satisfies

$$\kappa^2 = \kappa_n^2 + \kappa_g^2.$$

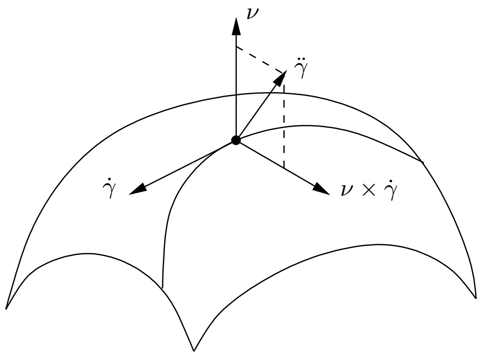

Figure 28: Decomposition of the acceleration vector of a unit-speed curve on a surface.

Lemma 4.2. *For every unit-speed curve γ on a surface one has*

$$\kappa_n = II(\dot{\gamma}, \dot{\gamma}).$$

In particular, all curves through the same point with the same tangent at this point have the same normal curvature.

Proof. Equation (12) implies

$$\kappa_n = \langle \ddot{\gamma}, \nu \rangle.$$

By differentiating the identity $\langle \dot{\gamma}, \nu \rangle = 0$ one gets

$$0 = \frac{d}{dt} \langle \dot{\gamma}, \nu \rangle = \langle \ddot{\gamma}, \nu \rangle + \langle \dot{\gamma}, \dot{\nu} \rangle.$$

On the other hand, one has

$$\dot{\nu} = \frac{d}{dt} \nu(\gamma(t)) = D_{\dot{\gamma}} \nu.$$

It follows that

$$\kappa_n = -\langle \dot{\gamma}, D_{\dot{\gamma}} \nu \rangle = \langle \dot{\gamma}, S(\dot{\gamma}) \rangle = II(\dot{\gamma}, \dot{\gamma}).$$

□

Definition 4.3. *A smooth curve on a smooth surface is called a* geodesic *if it has zero geodesic curvature.*

4.2 Length and energy of curves

Any two points on a surface can be connected by a smooth curve. How do we find a shortest curve between two given points?

Example 4.4. The shortest curve does not always exist. Take a punctured plane (a plane without a point) and consider two points collinear with the puncture and separated by it. There are curves going around the puncture whose length is arbitrarily close to the distance between the two chosen points in the plane, but no curve has exactly this length.

Example 4.5. The shortest curve is not always unique. On the sphere, the meridians are the shortest curves between the poles, and there are infinitely many of them. On a surface of revolution for any two points whose connecting segment intersects the rotation axis the shortest cannot be unique (if it exists) due to symmetry reasons: for any curve between these points there is another curve of the same length.

To formalize the problem, consider the space of all curves connecting two given points. Since a reparametrization does not change the length of a curve, without loss of generality we can assume that the parameter interval is $[0, 1]$. On this space, we are looking for the points of minimum of the length functional

$$L(\gamma) = \int_0^1 \|\dot\gamma\|\, dt.$$

As it turns out, minimizing the length is almost equivalent to minimizing something different.

Definition 4.6. *The* energy *of a smooth curve $\gamma\colon [0, 1] \to M$ is defined as*

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

Lemma 4.7. *For every curve $\gamma\colon [0, 1] \to \mathbb{R}^n$ one has*

$$E(\gamma) \geq \frac{1}{2} L(\gamma)^2.$$

The equality is attained if and only if γ has a constant speed.

Proof. Apply the Cauchy-Schwarz inequality

$$\int_0^1 f^2\, dt \int_0^1 g^2\, dt \geq \left(\int_0^1 fg\, dt \right)^2$$

to the functions $f = \|\dot\gamma\|$ and $g = 1$. The equality is attained only if f and g are proportional, which means $\|\dot\gamma\|$ is constant. $\square$

Theorem 4.8. *A smooth curve minimizes the energy among all curves connecting two given points if and only if it minimizes the length and has constant speed.*

Proof. If a curve minimizes the energy, then it has a constant speed: otherwise its constant-speed reparametrization has a smaller energy due to Lemma 4.7. By the same lemma, for constant-speed curves one has $E(\gamma) = \frac{1}{2} L(\gamma)^2$. Thus minimizing the energy among all curves is equivalent to minimizing the energy among constant speed curves, which is equivalent to minimizing the length among constant speed curves. $\square$

A disadvantage of the length functional is that its minimum points are never isolated: a reparametrization preserves the length. We do not know yet if the energy is free of flaws, but at least it has a chance. Another advantage of the energy compared to the length is that the integrand $\|\dot\gamma\| = \sqrt{\langle\dot\gamma, \dot\gamma\rangle}$ contains a square root while the integrand for the energy does not. This simplifies the calculus that we are going to do in the next section.

4.3 Critical points of the energy and the geodesic curvature

Every minimum point is a critical point. In order to use this idea in our situation, we need to do some setup.

Let $\gamma\colon [0,1] \to M$ be a smooth curve. A *smooth deformation* of γ is a family of curves

$$\gamma_s\colon [0,1] \to M, \quad s \in (-\varepsilon, \varepsilon)$$

such that $\gamma_0 = \gamma$ and the map

$$\Gamma\colon (-\varepsilon, \varepsilon) \times [0,1] \to M, \quad \Gamma(s,t) = \gamma_s(t)$$

is smooth. In particular, this implies that each curve γ_s is smooth. A deformation γ_s is said to have *fixed endpoints* if

$$\gamma_s(0) = \gamma(0), \quad \gamma_s(1) = \gamma(1) \quad \text{for all } s \in (-\varepsilon, \varepsilon).$$

If a smooth curve $\gamma\colon [0,1] \to M$ minimizes the energy among all curves on M connecting two given points, then it is a critical point of the energy functional under deformations with fixed endpoints:

$$\left.\frac{d}{ds}\right|_{s=0} E(\gamma_s) = 0 \text{ if } \gamma_s(0) = \gamma(0), \gamma_s(1) = \gamma(1). \tag{13}$$

Every smooth deformation of a curve has a *derivative* given by

$$\left.\frac{\partial\gamma_s(t)}{\partial s}\right|_{s=0} = \left.\frac{\partial\Gamma(s,t)}{\partial s}\right|_{s=0}$$

The derivative of a deformation of γ is a tangent vector field along γ, that is a map $I \to TM$ which sends each $t \in I$ to a vector in $T_{\gamma(t)}M$. Conversely, every tangent vector field along γ is the derivative of some deformation of γ. This is easy to prove if the image of γ is covered by a surface patch: take the image of a linear deformation in U:

$$\gamma_s(t) = \sigma(u(t) + sx(t), v(t) + sy(t)),$$

where $X(t) = x(t)\sigma_u + y(t)\sigma_v$.

Theorem 4.9. *Let γ_s be a smooth deformation of a curve $\gamma\colon [0,1] \to M$, and let X be the derivative of γ_s. Then one has*

$$\left.\frac{d}{ds}\right|_{s=0} E(\gamma_s) = \langle\dot\gamma(1), X(1)\rangle - \langle\dot\gamma(0), X(0)\rangle - \int_0^1 \langle\ddot\gamma, X\rangle\, dt$$

Proof. Differentiating under the integral sign one obtains

$$\frac{d}{ds}E(\gamma_s) = \int_0^1 \left\langle \dot\gamma_s, \frac{\partial}{\partial s}\dot\gamma_s \right\rangle dt = \int_0^1 \left\langle \dot\gamma_s, \frac{\partial^2 \Gamma(s,t)}{\partial s \partial t} \right\rangle dt.$$

Substituting $s = 0$ and using the commutation of partial derivatives, one gets

$$\left.\frac{d}{ds}\right|_{s=0} E(\gamma_s) = \int_0^1 \langle \dot\gamma, \dot X \rangle.$$

Now use integration by parts to get

$$\int_0^1 \langle \dot\gamma, \dot X\rangle\, dt = \langle \dot\gamma, X\rangle\Big|_0^1 - \int_0^1 \langle \ddot\gamma, X\rangle\, dt$$

and the theorem is proved. □

Theorem 4.10. *A curve is a geodesic if and only if its constant-speed parametrization is a critical point of the energy functional. Every shortest curve is a geodesic.*

Proof. If $\gamma_s(0) = \gamma(0)$ and $\gamma_s(1) = \gamma(1)$, then $X(0) = X(1) = 0$, and the formula from Theorem 4.9 reads

$$\left.\frac{d}{ds}\right|_{s=0} E(\gamma_s) = -\int_0^1 \langle \ddot\gamma, X\rangle\, dt.$$

If γ is a unit-speed geodesic, then equation (5) implies $\ddot\gamma = \kappa_n \nu$. If γ is a geodesic of constant speed c, then $\ddot\gamma = c^2\kappa_n\nu$. In any case one has $\langle \ddot\gamma, X\rangle = 0$ for every tangent vector X. Thus constant-speed parametrizations of geodesics are critical points of the energy functional.

For the opposite direction, let γ be a critical point of E. If $\kappa_g(t_0) \neq 0$ for some t_0, then one can find a tangent vector field X along γ supported in a neighborhood of t_0 and such that $\langle \ddot\gamma, X\rangle > 0$ at t_0 and non-negative elsewhere. Taking a deformation with the derivative X one obtains a contradiction to the criticality of γ. Thus $\ddot\gamma$ is everywhere perpendicular to M and $\kappa_g = 0$.

Finally, a constant-speed parametrization of a shortest curve minimizes the energy, thus is a critical point of the energy, and thus is a geodesic. □

This is the first instance of the interaction between intrinsic and extrinsic geometry: to be a critical point of the energy is an intrinsic property, while to be geodesic looks extrinsic because the definition of the geodesic curvature uses the acceleration vector $\ddot\gamma$ which is extrinsic. As we will see later, not only the vanishing of the geodesic curvature, but also its exact value is an intrinsic quantity.

4.4* Exponential map

The existence and uniqueness of a geodesic with initial conditions (see Corollary 4.24) is of local character. If we start from a given point with a given velocity, we do not necessarily can move forever. For example, if M is an open disk, then we will "hit the boundary" in whichever direction we move. This is not a problem for us, because in this and the following sections we will be interested in the local picture only.

Definition 4.11. *Let $X \in T_pM$, $X \neq 0$. Denote by γ^X a geodesic in M which at time 0 passes through p with the velocity X:*

$$\gamma^X : I \to M, \quad \gamma^X(0) = p, \quad \dot{\gamma}^X(0) = X,$$

where I is some interval containing 0 in its interior.

Lemma 4.12. *For every real $c \neq 0$ one has $\gamma^{cX}(t) = \gamma^X(ct)$.*

Proof. Indeed, the curve $t \mapsto \gamma^X(ct)$ passes through p at $t = 0$ and has there the velocity vector

$$\left.\frac{d}{dt}\right|_{t=0} \gamma^X(ct) = c \left.\frac{d}{dt}\right|_{t=0} \gamma^X(t) = cX.$$

□

Definition 4.13. *Assume that the maximum interval of existence of the geodesic γ^X contains $t = 1$. Denote*

$$\exp_p(X) := \gamma^X(1).$$

Besides, put $\exp_p(0) = p$. The resulting map

$$\exp_p : \Omega \to M, \quad \Omega \subset T_pM$$

is called the exponential map *of M at p.*

Example 4.14. The exponential map of the sphere at its north pole maps the disk of radius π to the northern hemisphere. The image of the disk of radius 2π covers the whole sphere; its boundary goes to the south pole. Every vector of length $k\pi$ goes to one of the poles: north if k is even, south if k is odd.

For the sphere the exponential map is defined on all of T_pM. This is true for any compact surface without boundary, but we will not prove this. The main theorem of this section is the following.

Theorem 4.15. *For every smooth surface $M \subset \mathbb{R}^3$ and every point $p \in M$ there is a neighborhood Ω of $0 \in T_pM$ such that the exponential map $\exp_p : \Omega \to M$ is defined and is a diffeomorphism onto its image.*

In the theory of ordinary differential equations one estimates the interval of existence of the solution of an ODE and proves that the solution depends smoothly on the initial data. This implies the following fact.

Fact 4.16. *Let $p \in M$ be a point on a smooth surface. For any $r > 0$ denote by $B_r(0)$ the set of tangent vectors in T_pM of the norm $< r$. Then there are positive real numbers r and ε such that there exists a smooth map*

$$\Gamma : B_r(0) \times (-\varepsilon, \varepsilon) \to M$$

with $\Gamma(X, t) = \gamma^X(t)$ for $X \neq 0$ and $\Gamma(0, t) = p$.

Proof of Theorem 4.15. Fact 4.16 implies that the exponential map $\exp_p$ is defined and is smooth on the ball $B_{c\varepsilon r}(0)$ for any $c < 1$. Indeed, one has

$$\exp_p(X) = \gamma^X(1) = \gamma^{c^{-1}\varepsilon^{-1}X}(c\varepsilon) = \Gamma(c^{-1}\varepsilon^{-1}X, c\varepsilon),$$

and if $\|X\| < c\varepsilon r$, then $(c^{-1}\varepsilon^{-1}X, c\varepsilon) \in B_r(0) \times (-\varepsilon, \varepsilon)$. In order to show that $\exp_p$ is a local diffeomorphism at 0, it suffices to prove that the differential of $\exp_p$ is invertible at 0.

For brevity, denote $c\varepsilon r =: \delta$. We have a smooth map

$$\exp_p \colon B_\delta \to M, \quad \exp_p(0) = p,$$

and want to compute the differential $d\exp_p$ at 0. The tangent space to B_δ at 0 can be canonically identified with T_pM (the velocity vectors of curves in $B_\delta \subset T_pM$ lie in T_pM). To find the image of $X \in T_pM$ under $\left(d\exp_p\right)_0$, we must by definition take a curve

$$\gamma \colon I \to B_\delta, \quad \gamma(0) = 0, \quad \dot\gamma(0) = X$$

and compute the velocity of $\exp_p \circ \gamma$ at 0. An obvious simple choice is $\gamma(t) = tX$. Thus we have

$$\left(d\exp_p\right)_0(X) = \left.\frac{d}{dt}\right|_{t=0} \exp_p(tX) = \left.\frac{d}{dt}\right|_{t=0} \gamma^{tX}(1) = \left.\frac{d}{dt}\right|_{t=0} \gamma^X(t) = X.$$

This means that the differential of $\exp_p$ at 0 is the identity map, in particular invertible, and therefore there is a neighborhood $\Omega \subset B_\delta$ of 0 such that $\exp_p \colon \Omega \to M$ is a diffeomorphism onto the image. □

4.5* Geodesic polar coordinates and the Gauss lemma

Since $\exp_p \colon \Omega \to M$ is a diffeomorphism onto the image, any diffeomorphism $U \to \Omega$ from some $U \subset \mathbb{R}^2$ determines a surface patch for M or, in other words, local coordinates on M. For example, one can identify T_pM with $\mathbb{R}^2$ by choosing any orthonormal basis of T_pM. The corresponding coordinates on M are called *normal coordinates*.

Here we will prefer polar coordinates to the Cartesian ones. Choose a half-line in T_pM starting at 0 and an orientation of T_pM. Then for every vector $X \in T_pM$ put $\rho(X) = \|X\|$ and let $\theta(X) \in \mathbb{R}/2\pi\mathbb{Z}$ be the angle from the chosen line to the vector X. For any $r > 0$ and $I \subset \mathbb{R}$ such that $|I| < 2\pi$ this yields a surface patch $(0, r) \times I \to T_pM$ and, if the image of this patch is contained in Ω, a surface patch for M.

Definition 4.17. *The local coordinates on M produced by the exponential map $\exp_p$ from polar coordinates on T_pM are called* geodesic polar coordinates *around p.*

Theorem 4.18. *The first fundamental form in geodesic polar coordinates has the form*

$$d\rho^2 + G(\rho, \theta)d\theta^2.$$

In particular, the curves $\theta = \mathrm{const}$ are orthogonal to the curves $\rho = \mathrm{const}$.

The latter (orthogonality of parameter curves) is known as the *Gauss lemma.*

Proof. Let $\sigma\colon (0,r) \times I \to M$ be a surface patch corresponding to geodesic polar coordinates. We have to show

$$\|\sigma_\rho\| = 1, \quad \langle \sigma_\rho, \sigma_\theta \rangle = 0$$

at every point (r_0, θ_0). For every θ consider the curve

$$\gamma^\theta\colon [0, r_0] \to M, \quad \gamma^\theta(t) = \sigma(t, \theta).$$

By definition of the exponential map and the geodesic polar coordinates, the curve γ^θ is a unit-speed geodesic (its initial velocity is the vector $X \in T_pM$ with polar coordinates $(1, \theta)$). Therefore

$$\|\sigma_\rho\| = \|\dot\gamma^\theta\| = 1$$

and the first equation is proved.

For the second equation consider a smooth deformation γ^θ of the geodesic γ^{θ_0} and apply the formula of variation of energy, Theorem 4.9. (Clearly, the formula holds for geodesics defined on any interval, not only on $[0,1]$.) The derivative of this deformation is

$$\frac{\partial}{\partial\theta}\gamma^\theta(t) = \sigma_\theta.$$

On the other hand, all curves γ^θ are unit-speed geodesics of length r_0, so that all of them have energy $\frac{r_0}{2}$. Thus we have

$$0 = \frac{d}{d\theta}E(\gamma^\theta) = \langle \dot\gamma^\theta(r_0), \sigma_\theta \rangle = \langle \sigma_\rho, \sigma_\theta \rangle,$$

and the theorem is proved. □

We are now ready to clarify how the vanishing of geodesic curvature is related to being the shortest curve. Theorem 4.10 says that every constant-speed shortest curve is a geodesic. In fact, looking at its proof one can see that it yields a stronger statement: every constant-speed *locally shortest* curve is a geodesic.

Definition 4.19. *A curve $\gamma\colon I \to M$ is called* locally shortest *if for every $t_0 \in I$ there is a neighborhood $J \subset I$ of t_0 such that for every $t_1 \in J$ the restriction of γ to $[t_0, t_1]$ (or to $[t_1, t_0]$ if $t_1 < t_0$) is the shortest (and by this we mean the unique shortest) among all curves connecting $\gamma(t_0)$ with $\gamma(t_1)$.*

Theorem 4.20. *Geodesics are exactly the constant speed locally shortest curves.*

Proof. One direction is already proved. It remains to show that every geodesic is locally the shortest.

Let $\gamma\colon I \to M$ be a geodesic. Take any $t_0 \in I$; by a reparametrization (time shift) one can achieve $t_0 = 0$. Also, by time scaling one can achieve that γ is of unit speed. This will not influence the locally shortest property (whether it holds or not), thus there is no loss of generality. Let $\gamma(0) = p$.

Choose $r > 0$ so that $\exp_p\colon B_r(0) \to M$ is a diffeomorphism onto the image. Then, since a geodesic is uniquely determined by the initial condition, in the geodesic polar coordinates an arc of γ has the equation

$$\gamma\colon [0, r] \to M, \quad \gamma(t) = \sigma(t, \theta_0)$$

for some θ_0. We will show that for any $r_0 \in (0, r)$ the curve $\gamma_{[0,r_0]}$ is the shortest one from p to $\gamma(r_0) =: q$. (The same will hold for any $r_0 \in (-r, 0)$, because that arc corresponds to $\sigma(t, \theta_0 + \pi)$.)

Let

$$\delta\colon [a, b] \to M, \quad \delta(a) = p, \quad \delta(b) = q$$

be any curve connecting p and q. First, let us assume that δ does not leave the set $\exp_p(B_r(0))$. Then δ can be represented in the geodesic polar coordinates as

$$\delta(t) = \sigma(\rho(t), \theta(t)).$$

By Theorem 4.18 one has

$$\|\dot{\delta}\|^2 = \dot{\rho}^2 + G\dot{\theta}^2,$$

and the length of δ can be estimated from below as follows:

$$L(\delta) = \int_a^b \sqrt{\dot{\rho}^2 + G\dot{\theta}^2}\, dt \geq \int_a^b \|\dot{\rho}\|\, dt \geq \int_a^b \dot{\rho}\, dt = \rho(b) - \rho(a) = r_0 = L(\gamma).$$

The equality holds only if $\dot{\theta} = 0$ and ρ is monotone, that is only if δ is a reparametrization of γ. Next, if δ leaves the set $\exp_p(B_r(0))$, then let $c \in [a, b]$ be the first time when δ intersects the geodesic circle $\rho = t_0$. Then by the above estimate the length of $\delta_{[a,c]}$ is already at least r_0, so that δ is longer than γ. $\square$

4.6 Geodesic equation

Here we will learn to compute (at least theoretically) geodesics in coordinates.

Theorem 4.21. *Let $M \subset \mathbb{R}^3$ be a smooth surface, and let $\sigma\colon U \to \mathbb{R}^3$ be a surface patch for M. Then a unit-speed curve $\gamma(t) = \sigma(u(t), v(t))$ is a geodesic if and only if it satisfies the following system of differential equations:*

$$\begin{aligned} \ddot{u} + \Gamma^1_{11}\dot{u}^2 + 2\Gamma^1_{12}\dot{u}\dot{v} + \Gamma^1_{22}\dot{v}^2 = 0, \\ \ddot{v} + \Gamma^2_{11}\dot{u}^2 + 2\Gamma^2_{12}\dot{u}\dot{v} + \Gamma^2_{22}\dot{v}^2 = 0. \end{aligned}$$

Here the coefficients Γ^i_{jk} are certain rational functions of the coefficients of the first fundamental form and of its first order partial derivatives.

Note that this theorem implies the intrinsic nature of geodesics, because the coefficients of the equations depend only on the first fundamental form.

We need a lemma.

Lemma 4.22. *Let $M \subset \mathbb{R}^3$ be a smooth surface, and let $\sigma\colon U \to \mathbb{R}^3$ be a surface patch for M. Then a unit-speed curve $\gamma = \sigma(u(t), v(t))$ has vanishing geodesic curvature 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. The geodesic curvature of γ vanishes at a given point if and only if $\ddot{\gamma}$ is perpendicular to M at this point, that is

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

Viewing $\sigma_u = \sigma_u(u(t), v(t))$ as a function of t, apply the Leibniz rule:

$$\frac{d}{dt}\langle\dot{\gamma}, \sigma_u\rangle = \left\langle\dot{\gamma}, \frac{d}{dt}\sigma_u\right\rangle + \langle\ddot{\gamma}, \sigma_u\rangle = \left\langle\dot{\gamma}, \frac{d}{dt}\sigma_u\right\rangle. \tag{14}$$

Due to $\dot{\gamma} = \dot{u}\sigma_u + \dot{v}\sigma_v$ one has

$$\frac{d}{dt}\langle\dot{\gamma}, \sigma_u\rangle = \frac{d}{dt}\langle\dot{u}\sigma_u + \dot{v}\sigma_v, \sigma_u\rangle = \frac{d}{dt}(E\dot{u} + F\dot{v}).$$

On the other hand,

$$\begin{aligned}\left\langle\dot{\gamma}, \frac{d}{dt}\sigma_u\right\rangle &= \langle\dot{u}\sigma_u + \dot{v}\sigma_v, \dot{u}\sigma_{uu} + \dot{v}\sigma_{uv}\rangle \\ &= \langle\sigma_u, \sigma_{uu}\rangle\dot{u}^2 + (\langle\sigma_u, \sigma_{uv}\rangle + \langle\sigma_v, \sigma_{uu}\rangle)\,\dot{u}\dot{v} + \langle\sigma_v, \sigma_{uv}\rangle\dot{v}^2 \\ &= \frac{1}{2}(E_u\dot{u}^2 + 2F_u\dot{u}\dot{v} + G_u\dot{v}^2),\end{aligned}$$

because of

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

Substituting this into (14) proves the first equation of the lemma. The second equation is proved in a similar way. □

Remark 4.23. If you studied calculus of variations, then you may recognize in the equations of Lemma 4.22 the Euler-Lagrange equations for the Lagrangian

$$L(t, u, v, \dot{u}, \dot{v}) = \frac{1}{2}(E\dot{u}^2 + 2F\dot{u}\dot{v} + G\dot{v}^2).$$

Proof of Theorem 4.21. Rewrite the left hand sides of equations of Lemma 4.22 using the equations

$$\frac{d}{dt}E = E_u \dot{u} + E_v \dot{v}, \quad \text{etc.}$$

We get

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

Bringing all terms of the equations of Lemma 4.22 to the left hand side we obtain

$$\begin{aligned}
E\ddot{u} + F\ddot{v} + \frac{1}{2}(E_u \dot{u}^2 + 2E_v \dot{u}\dot{v} + (2F_v - G_u)\dot{v}^2) &= 0 \\
F\ddot{u} + G\ddot{v} + \frac{1}{2}((2F_u - E_v)\dot{u}^2 + 2G_u \dot{u}\dot{v} + G_v \dot{v}^2) &= 0
\end{aligned}$$

This is an inhomogeneous system of linear equations on $\ddot{u}$ and $\ddot{v}$ whose coefficient matrix is the matrix of the first fundamental form. Since this matrix is non-degenerate, the system can be solved and the solution has the form described in Theorem 4.21. □

Equations in Theorem 4.21 are non-linear and usually cannot be solved explicitly. However, the existence and uniqueness theorem from the theory of ordinary differential equations implies the following.

Corollary 4.24. *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.*

4.7 Geodesics on surfaces of revolution

Theorem 4.25 (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.

See Figure 29.

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

$$\dot{\gamma} = \sigma_u \cos\psi \pm \frac{\sigma_v}{f} \sin\psi.$$

(Indeed, σ_u and $\frac{\sigma_v}{f}$ are unit vectors tangent to the meridian, respectively to the parallel.) Thus we have

$$\dot{u} = \cos\psi, \quad \dot{v} = \pm\frac{\sin\psi}{f}.$$

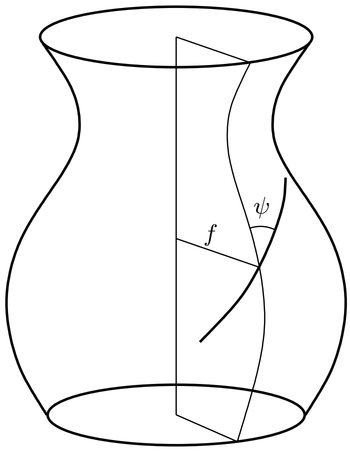

Figure 29: Geodesic on a surface of revolution.

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 above, 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. □

Figure 27 shows some geodesics on an unduloid.

Corollary 4.26. *If a geodesic on a surface of revolution is tangent to two latitudinal circles, then these circles have equal radii.*

For example, if between two latitudinal circles of equal radii f_0 the distance to the axis satisfies $f > f_0$, then there are geodesics which oscillate between these two circles.

Clairaut's theorem can be interpreted mechanically as conservation of angular momentum. Indeed, a geodesic can be viewed as a trajectory of a particle constrained to move on a surface: the orthogonality of $\ddot{\gamma}$ to M means that the only force acting on the particle comes from the surface. On a surface of revolution, all normal vectors pass through the rotation axis, thus the constraining force does not change the angular momentum.

We have taken the so called Lagrangian approach to the geodesics. There is also a very efficient Hamiltonian approach which allows in particular to describe geodesics on an ellipsoid.

4.8 Covariant derivative

Lecture 9

In Section 3.2 we have defined the directional derivative of a vector field along M. The derivative of $Y: M \to \mathbb{R}^3$ at $p \in M$ in the direction $X \in T_pM$ is

$$D_X Y = \left.\frac{d}{dt}\right|_{t=0} Y_{\gamma(t)},$$

where $\gamma(0) = p$ and $\dot{\gamma}(0) = X$. For a tangent vector field ($Y_p \in T_pM$ for all $p \in M$) its directional derivative is not necessarily tangent to M.

For every $p \in M$ there is a direct sum decomposition

$$\mathbb{R}^3 = T_pM \oplus (T_pM)^\perp,$$

where of course $(T_pM)^\perp = \mathbb{R}\nu$ for any unit normal ν to M at p. For every $Z \in \mathbb{R}^3$ the summands in

$$Z = Z_\top + Z_\perp, \quad Z_\top \in T_pM, \quad Z_\perp \in (T_pM)^\perp$$

are called the *tangential* and the *normal component* of Z, respectively. One has

$$Z_\perp = \langle Z, \nu \rangle \nu, \quad Z_\top = Z - \langle Z, \nu \rangle \nu.$$

The normal component of the directional derivative has a simple and useful geometric meaning stated in the next lemma.

Lemma 4.27. *For any tangent vector field Y and any tangent vector X one has*

$$\langle D_X Y, \nu \rangle = II(X, Y).$$

Proof. Indeed, differentiating $\langle Y, \nu \rangle = 0$ by the product rule one obtains

$$0 = D_X \langle Y, \nu \rangle = \langle D_X Y, \nu \rangle + \langle Y, D_X \nu \rangle,$$

which implies

$$\langle D_X Y, \nu \rangle = -\langle D_X \nu, Y \rangle = \langle S(X), Y \rangle = II(X, Y).$$

□

The tangential component of the directional derivative has not shown up yet but will play a prominent role in this lecture.

Definition 4.28. *Let $X \in T_pM$, and let Y be a tangent vector field on M. The tangential component of $D_X Y$ is called the* covariant derivative *of Y in the direction of X and denoted by $\nabla_X Y$.*

Due to the previous lemma one has

$$\nabla_X Y = D_X Y - II(X,Y)\nu.$$

A less explicit but useful description of the covariant derivative:

$$\langle \nabla_X Y, Z\rangle = \langle D_X Y, Z\rangle \quad \text{for all } Z \in T_pM.$$

If X is also a vector field, then by differentiating Y in the direction of X_p at every point p one obtains a vector field $\nabla_X Y$. One can write this as

$$(\nabla_X Y)_p := \nabla_{X_p} Y. \tag{15}$$

For the value of $\nabla_X Y$ at a point p (that is for the vector (15)) one needs from X only its value at p. By contrast, from Y one needs more than that, but how much more? We don't really need the values of Y on all of M, any neighborhood of p will suffice. Even less than that, the values of Y along any curve through p with the velocity vector X at p will do. This gives meaning to the expressions like $\nabla_{\dot\gamma}\nu$ (which has appeared in the proof of Lemma 4.2) and $\nabla_{\dot\gamma}\dot\gamma$.

Lemma 4.29. *Let γ be a unit-speed curve on a surface M. Then the absolute value of the geodesic curvature of γ coincides with the norm of the covariant derivative of the velocity vector field of the curve in the direction of the curve:*

$$|\kappa_g| = \|\nabla_{\dot\gamma}\dot\gamma\|\,.$$

A curve γ on M is a constant-speed geodesic if and only if $\nabla_{\dot\gamma}\dot\gamma = 0$.

Proof. By definition, one has

$$D_{\dot\gamma}\dot\gamma = \frac{d}{dt}\dot\gamma = \ddot\gamma \quad \Rightarrow \quad \nabla_{\dot\gamma}\dot\gamma = \ddot\gamma_\top.$$

(Note the double role the velocity vectors of γ are playing here: they are the vector field that we differentiate and the direction in which we differentiate.) If γ has unit speed, then

$$\ddot\gamma = \kappa_n\nu + \kappa_g \nu\times\dot\gamma \quad \Rightarrow \quad \ddot\gamma_\top = \kappa_g\nu\times\dot\gamma \quad \Rightarrow \quad \|\ddot\gamma_\top\| = |\kappa_g|,$$

which proves the first part of the lemma.

If γ is a geodesic, then $\kappa_g = 0$, which by the first part of the lemma implies $\nabla_{\dot\gamma}\dot\gamma = 0$ for a unit-speed parametrization of γ. But for a constant-speed reparametrization the derivative $\nabla_{\dot\gamma}\dot\gamma$ gets multiplied with the square of the velocity, so this vector remains zero. In the opposite direction, if $\nabla_{\dot\gamma}\dot\gamma = 0$, then

$$\langle\dot\gamma,\ddot\gamma\rangle = \langle\dot\gamma,\ddot\gamma_\top\rangle = \langle\dot\gamma,\nabla_{\dot\gamma}\dot\gamma\rangle = 0 \quad \Rightarrow \quad \|\dot\gamma\| = \text{const}\,.$$

Thus κ_g is a scalar multiple of $\|\nabla_{\dot\gamma}\dot\gamma\|$ and hence zero. $\square$

The main theorem of this section is the following.

Theorem 4.30. *The covariant derivative is intrinsic.*

First one has to understand what does intrinsic mean in this case. The covariant derivative is an operation on tangent vector fields, so we must interpret tangent vectors intrinsically. First they seem to be something extrinsic because they lie in $\mathbb{R}^3$. But they are the velocity vectors of curves lying in M, so one could describe a tangent vector at p as an equivalence class of curves through p under a certain equivalence relation. Informally speaking, tangent vectors are infinitely short curves on M, which makes then intrinsic.

Formally, every surface patch σ allows to bring all intrinsic geometry of M (of the part covered by the patch) onto the parameter domain U. The intrinsic geometry is then expressed by a family of quadratic forms I_σ depending on (u, v). A tangent vector field on M corresponds to a tangent vector field on U: the vectors $\sigma_u, \sigma_v \in T_pM \subset \mathbb{R}^3$ correspond to the basis vectors in U. Because of this the theorem can be reformulated as follows.

Theorem 4.30 (explained)**.** *The components of the vector field $\nabla_X Y$ in the basis (σ_u, σ_v) defined by a surface patch σ depend only on the components of X and Y and on the coefficients of the first fundamental form I_σ.*

In particular, if $F\colon M \to N$ is an isometry, then one has

$$dF(\nabla_X Y) = \nabla_{dF(X)} dF(Y).$$

For the proof of Theorem 4.30 we will need the following properties of the covariant derivative.

Lemma 4.31. *The covariant derivative has the following properties.*

1. *The map $(X, Y) \mapsto \nabla_X Y$ is $C^\infty(M)$-linear in X:*

$$\nabla_{X_1+X_2} Y = \nabla_{X_1} Y + \nabla_{X_2} Y, \quad \nabla_{fX} Y = f\, \nabla_X Y.$$

2. *The map $(X, Y) \mapsto \nabla_X Y$ is additive in Y and satisfies the product rule with respect to multiplication of Y with a function:*

$$\nabla_X (Y_1 + Y_2) = \nabla_X Y_1 + \nabla_X Y_2, \quad \nabla_X (fY) = (D_X f) Y + f\, \nabla_X Y.$$

3. *For any three tangent vector fields X, Y, Z he following product rule holds:*

$$D_X \langle Y, Z\rangle = \langle \nabla_X Y, Z\rangle + \langle Y, \nabla_X Z\rangle$$

Proof. Properties 1. and 2. follow from Lemma 3.7 by taking the tangential parts under the assumption that Y is a tangent vector field.

Property 3. follows from the first product rule in Lemma 3.8 and the fact that $\langle \nabla_X Y, Z\rangle = \langle D_X Y, Z\rangle$ for a tangent vector field Z.

□

Sometimes it is convenient to use ∇_X instead of D_X for the directional derivative of a function, so that the product formulas become more uniform:

$$\nabla_X(fY) = (\nabla_X f)Y + f\nabla_X Y, \quad \nabla_X\langle Y, Z\rangle = \langle \nabla_X Y, Z\rangle + \langle Y, \nabla_X Z\rangle.$$

Let us now prove the main theorem.

Proof of Theorem 4.30. Take a surface patch $\sigma\colon U \to \mathbb{R}^3$ for M and let

$$X = X^1\sigma_u + X^2\sigma_v, \quad Y = Y^1\sigma_u + Y^2\sigma_v,$$

where X^1, X^2, Y^1, Y^2 are smooth functions on U. By points 1. and 2. of Lemma 4.31 one has

$$\nabla_X Y = X^1\nabla_u Y + X^2\nabla_v Y,$$

where we denote

$$\nabla_u Y := \nabla_{\sigma_u} Y, \quad \nabla_v Y := \nabla_{\sigma_v} Y.$$

By point 2. of Lemma 4.31 one has

$$\nabla_u Y = \frac{\partial Y^1}{\partial u}\sigma_u + Y^1\,\nabla_u\sigma_u + \frac{\partial Y^2}{\partial u}\sigma_v + Y^2\,\nabla_u\sigma_v,$$

and a similar formula for $\nabla_v Y$. It remains to express $\nabla_u\sigma_u$ and the like in terms of the first fundamental form. The vector $\nabla_u\sigma_u$ is the tangential part of $D_u\sigma_u = \sigma_{uu}$. Thus one has

$$\nabla_u\sigma_u = (\sigma_{uu})_\top \quad \text{etc.}$$

Let

$$\begin{aligned} \sigma_{uu} &= \Gamma^1_{11}\sigma_u + \Gamma^2_{11}\sigma_v + (\sigma_{uu})_\perp \\ \sigma_{uv} &= \Gamma^1_{12}\sigma_u + \Gamma^2_{12}\sigma_v + (\sigma_{uv})_\perp \\ \sigma_{vv} &= \Gamma^1_{22}\sigma_u + \Gamma^2_{22}\sigma_v + (\sigma_{vv})_\perp \end{aligned} \tag{16}$$

Our goal is to show that the coefficients Γ^k_{ij} depend only on the coefficients of the first fundamental form. Take the inner product of the first equation with σ_u:

$$\begin{aligned} \langle\sigma_{uu}, \sigma_u\rangle &= \Gamma^1_{11}E + \Gamma^2_{11}F \\ \langle\sigma_{uu}, \sigma_v\rangle &= \Gamma^1_{11}F + \Gamma^2_{11}G \end{aligned}$$

This is a system of linear equations on Γ^1_{11} and Γ^2_{11} with a non-degenerate matrix $\begin{pmatrix} E & F \\ F & G \end{pmatrix}$. If we are able to express the left hand sides in terms of E, F, G, then we are done. We did something similar in Section 4.6. From the equations

$$\langle\sigma_{uu}, \sigma_u\rangle = \frac{1}{2}E_u, \quad \langle\sigma_{vv}, \sigma_v\rangle = \frac{1}{2}G_v,$$
$$\langle\sigma_{uv}, \sigma_u\rangle = \frac{1}{2}E_v, \quad \langle\sigma_{uv}, \sigma_v\rangle = \frac{1}{2}G_u,$$
$$\langle\sigma_{uu}, \sigma_v\rangle + \langle\sigma_{uv}, \sigma_u\rangle = F_u,$$
$$\langle\sigma_{vv}, \sigma_u\rangle + \langle\sigma_{uv}, \sigma_v\rangle = F_v$$

one gets

$$\langle \sigma_{uu}, \sigma_u \rangle = \frac{1}{2} E_u, \quad \langle \sigma_{uu}, \sigma_v \rangle = F_u - \frac{1}{2} E_v.$$

Thus the above linear system takes the form

$$\begin{pmatrix} E & F \\ F & G \end{pmatrix} \begin{pmatrix} \Gamma^1_{11} \\ \Gamma^2_{11} \end{pmatrix} = \begin{pmatrix} \frac{1}{2} E_u \\ F_u - \frac{1}{2} E_v \end{pmatrix}.$$

One obtains similar systems for other Γ^z_{xy}, which shows that all of these coefficients are intrinsic. □

The coefficients Γ^z_{xy} in (16) are called *Christoffel symbols*. The complete set of equations allowing to express the Christoffel symbols in terms of the first fundamental form is

$$\begin{pmatrix} E & F \\ F & G \end{pmatrix} \begin{pmatrix} \Gamma^1_{11} & \Gamma^1_{12} & \Gamma^1_{22} \\ \Gamma^2_{11} & \Gamma^2_{12} & \Gamma^2_{22} \end{pmatrix} = \frac{1}{2} \begin{pmatrix} E_u & E_v & 2F_v - G_u \\ 2F_u - E_v & G_u & G_v \end{pmatrix} \tag{17}$$

Since $\sigma_{uv} = \sigma_{vu}$, it makes sense to introduce also the Christoffel symbols $\Gamma^1_{21} = \Gamma^1_{12}$ and $\Gamma^2_{21} = \Gamma^2_{21}$.

From the formulas at the beginning of the proof of Theorem 4.30 one derives the following formula for the covariant derivative of a vector field:

$$\begin{aligned} \nabla_X Y = \left(D_X Y^1 + \sum_{i,j} \Gamma^1_{ij} X^i Y^j \right) \sigma_u + \left(D_X Y^2 + \sum_{i,j} \Gamma^2_{ij} X^i Y^j \right) \sigma_v \\ = (D_X Y^1) \sigma_u + (D_X Y^2) \sigma_v + \sum_{i,j,k} \Gamma^k_{ij} X^i Y^j \sigma_k \end{aligned} \tag{18}$$

In combination with Lemma 4.29 this gives an alternative proof of geodesic equations from Theorem 4.21. It suffices to substitute

$$X = Y = \dot{\gamma} = \dot{u} \sigma_u + \dot{v} \sigma_v$$

into the formula above and to equate to zero the coefficients at σ_u and σ_v.

4.9 Parallel transport

Let $\gamma \colon [a, b] \to M$ be a smooth curve on a smooth surface M. Recall that a tangent vector field along γ is a smooth map $X \colon [a, b] \to \mathbb{R}^3$ such that $X(t) \in T_{\gamma(t)} M$.

Definition 4.32. *A tangent vector field X along γ is called* parallel *if one has $\nabla_{\dot{\gamma}} X = 0$ at every point of γ.*

Example 4.33. The velocity vector field of a curve is parallel if and only if this curve is a geodesic.

Lemma 4.34. *For every smooth curve $\gamma \colon [a, b] \to M$ and every vector $X_0 \in T_{\gamma(a)} M$ there is a unique tangent vector field X along γ that is parallel and satisfies $X(a) = X_0$.*

Proof. By equation (18), the condition $\nabla_{\dot\gamma} X = 0$ is equivalent to a system of first order linear differential equations with variable coefficients

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

By the general theory, this system has a unique solution on the interval $[a, b]$. □

Let $p, q \in M$ be two arbitrary points, and let $\gamma\colon [a, b] \to M$ be a smooth curve connecting these points: $\gamma(a) = p, \gamma(b) = q$. Then the above lemma defines a map

$$\Pi_\gamma\colon T_pM \to T_qM$$

which sends every vector $X \in T_pM$ to the vector $Y \in T_qM$ such that $Y = Z(b)$ for the parallel vector field Z along γ with the initial condition $Z(a) = X$. The map Π_γ is called *parallel transport* along γ. Observe that for different curves γ from p to q the maps Π_γ are different.

Lemma 4.35. *The parallel transport map is an isometry of inner product spaces, that is $\langle \Pi_\gamma(X), \Pi_\gamma(Y)\rangle = \langle X, Y\rangle$ for all $X, Y \in T_pM$.*

Proof. Exercise. □

The parallel transport has the following mechanical interpretation. Assume that M is the boundary of a convex body. Put the body on a plane so that the plane touches it at the point p. Then roll the body along the curve γ (which is drawn on its surface) without sliding and skidding. In the initial and in the final position the plane of support can be identified with the tangent planes T_pM and T_qM, respectively. The parallel transport along γ then corresponds to the translation of the plane of support which moves the initial point of contact to the final point of contact. From a different point of view one can imagine the tangent planes to M rolling along the curve γ.

4.10 Theorema Egregium

Theorem 4.36 (Gauss' Theorema Egregium)**.** *The Gauss curvature is intrinsic, that is it can be expressed in terms of the coefficients of the first fundamental form. In particular, the Gauss curvature is preserved by isometries.*

By contrast, the mean curvature is not preserved by all isometries. For example, a developable surface is isometric to a piece of the plane. The plane has both Gauss and mean curvature zero. A developable surface has one zero principal curvature and one non-zero (in general), thus zero Gauss and non-zero mean curvature. Bending surfaces of revolution provides another set of examples.

We prove Theorema Egregium by computation in coordinates, which was also historically the first argument. Later we return to it with a coordinate-free approach.

Recall from Lemma 3.20 and Section 3.4 the formula for the Gauss curvature in terms of the coefficients of fundamental forms:

$$K = \frac{\det II}{\det I} = \frac{LN - M^2}{EG - F^2}.$$

Also recall that

$$\langle \sigma_{uu}, \nu \rangle = L, \quad \langle \sigma_{uv}, \nu \rangle = M, \quad \langle \sigma_{vv}, \nu \rangle = N, \tag{19}$$

where ν is a local unit normal field.

Proof of Theorem 4.36. Due to (19) equations (16) become

$$\begin{aligned}
\sigma_{uu} &= \Gamma^1_{11}\sigma_u + \Gamma^2_{11}\sigma_v + L\nu \\
\sigma_{uv} &= \Gamma^1_{12}\sigma_u + \Gamma^2_{12}\sigma_v + M\nu \\
\sigma_{vv} &= \Gamma^1_{22}\sigma_u + \Gamma^2_{22}\sigma_v + N\nu
\end{aligned}$$

The plan is simple: take the v-derivative of the first equation and the u-derivative of the second equation and use $(\sigma_{uu})_v = (\sigma_{uv})_u$. The differentiated right hand sides of the first and the second equations will be decomposed in the basis $(\sigma_u, \sigma_v, \nu)$, and the coefficients at σ_u will be compared.

Differentiate the first equation:

$$(\sigma_{uu})_v = (\Gamma^1_{11})_v\sigma_u + (\Gamma^2_{11})_v\sigma_v + L_v\nu + \Gamma^1_{11}\sigma_{uv} + \Gamma^2_{11}\sigma_{vv} + L\nu_v. \tag{20}$$

We know the decomposition of σ_{uv} and σ_{vv} in the basis $(\sigma_u, \sigma_v, \nu)$. For $\nu_v = -S(\sigma_v)$ use the fact that the matrix of the shape operator S is $I^{-1}II$ (Lemma 3.14):

$$S = \frac{1}{EG - F^2}\begin{pmatrix} GL - FM & GM - FN \\ EM - FL & EN - FM \end{pmatrix}$$

It follows that

$$\begin{aligned}
\nu_u &= \frac{FM - GL}{EG - F^2}\sigma_u + \frac{FL - EM}{EG - F^2}\sigma_v \\
\nu_v &= \frac{FN - GM}{EG - F^2}\sigma_u + \frac{FM - EN}{EG - F^2}\sigma_v
\end{aligned}$$

Substitute into the right hand side of (20) the decompositions of σ_{uv}, σ_{vv}, and ν_v:

$$\begin{aligned}
(\sigma_{uu})_v = &\left((\Gamma^1_{11})_v + \Gamma^1_{11}\Gamma^1_{12} + \Gamma^2_{11}\Gamma^1_{22} + L\frac{FN - GM}{EG - F^2} \right)\sigma_u \\
&+ \left((\Gamma^2_{11})_v + \Gamma^1_{11}\Gamma^2_{12} + \Gamma^2_{11}\Gamma^2_{22} + L\frac{FM - EN}{EG - F^2} \right)\sigma_v \\
&+ \left(L_v + \Gamma^1_{11}M + \Gamma^2_{11}N \right)\nu
\end{aligned} \tag{21}$$

Similarly, compute

$$\begin{aligned}
(\sigma_{uv})_u &= (\Gamma^1_{12})_u\sigma_u + (\Gamma^2_{12})_u\sigma_v + M_u\nu + \Gamma^1_{12}\sigma_{uu} + \Gamma^2_{12}\sigma_{uv} + M\nu_u \\
&= \left((\Gamma^1_{12})_u + \Gamma^1_{12}\Gamma^1_{11} + \Gamma^2_{12}\Gamma^1_{12} + M\frac{FM-GL}{EG-F^2} \right)\sigma_u \\
&+ \left((\Gamma^2_{12})_u + \Gamma^1_{12}\Gamma^2_{11} + \Gamma^2_{12}\Gamma^2_{12} + M\frac{FL-EM}{EG-F^2} \right)\sigma_v \\
&+ \left(M_u + \Gamma^1_{12}L + \Gamma^2_{12}M \right)\nu
\end{aligned} \tag{22}$$

Equating the coefficients at σ_u one gets

$$\begin{aligned}
(\Gamma^1_{11})_v + \Gamma^1_{11}\Gamma^1_{12} + \Gamma^2_{11}\Gamma^1_{22} + L\frac{FN-GM}{EG-F^2} &= (\Gamma^1_{12})_u + \Gamma^1_{12}\Gamma^1_{11} + \Gamma^2_{12}\Gamma^1_{12} + M\frac{FM-GL}{EG-F^2} \\
\Rightarrow (\Gamma^1_{11})_v - (\Gamma^1_{12})_u + \Gamma^2_{11}\Gamma^1_{22} - \Gamma^2_{12}\Gamma^1_{12} &= M\frac{FM-GL}{EG-F^2} - L\frac{FN-GM}{EG-F^2} \\
&= F\frac{M^2-LN}{EG-F^2} = -F\frac{\det II}{\det I} = -FK
\end{aligned}$$

This proves the theorem, since it allows to express the Gauss curvature K in terms of the coefficients of the first fundamental form. ☐

Let us also equate the coefficients at σ_v in (21) and (22):

$$\begin{aligned}
(\Gamma^2_{11})_v + \Gamma^1_{11}\Gamma^2_{12} + \Gamma^2_{11}\Gamma^2_{22} + L\frac{FM-EN}{EG-F^2} &= (\Gamma^2_{12})_u + \Gamma^1_{12}\Gamma^2_{11} + \Gamma^2_{12}\Gamma^2_{12} + M\frac{FL-EM}{EG-F^2} \\
\Rightarrow (\Gamma^2_{11})_v - (\Gamma^2_{12})_u + \Gamma^1_{11}\Gamma^2_{12} + \Gamma^2_{11}\Gamma^2_{22} - \Gamma^1_{12}\Gamma^2_{11} - (\Gamma^2_{12})^2 &= M\frac{FL-EM}{EG-F^2} - L\frac{FM-EN}{EG-F^2} \\
&= E\frac{LN-M^2}{EG-F^2} = EK
\end{aligned}$$

The two expressions of the Gauss curvature in intrinsic terms we obtained are

$$\begin{aligned}
K &= \frac{(\Gamma^1_{12})_u - (\Gamma^1_{11})_v + \Gamma^2_{12}\Gamma^1_{12} - \Gamma^2_{11}\Gamma^1_{22}}{F} \\
K &= \frac{(\Gamma^2_{11})_v - (\Gamma^2_{12})_u + \Gamma^1_{11}\Gamma^2_{12} + \Gamma^2_{11}\Gamma^2_{22} - \Gamma^1_{12}\Gamma^2_{11} - (\Gamma^2_{12})^2}{E}
\end{aligned} \tag{23}$$

These are called *Gauss equations.* There are two more Gauss equations obtained by exchanging u and v (and arising from $(\sigma_{uv})_v = (\sigma_{vv})_u$.

The *Baltzer–Brioschi formula* expresses the Gauss curvature K directly in terms of E, F, G and their derivatives:

$$K = \frac{\begin{vmatrix} -\frac{1}{2}E_{vv} + F_{uv} - \frac{1}{2}G_{uu} & \frac{1}{2}E_u & F_u - \frac{1}{2}E_v \\ F_v - \frac{1}{2}G_u & E & F \\ \frac{1}{2}G_v & F & G \end{vmatrix} - \begin{vmatrix} 0 & \frac{1}{2}E_v & \frac{1}{2}G_u \\ \frac{1}{2}E_v & E & F \\ \frac{1}{2}G_u & F & G \end{vmatrix}}{(EG-F^2)^2}$$

One can derive this from the Gauss formulas and the formulas (17) for the Christoffel symbols, but there is also a shorter direct proof, see [BL73, §45].

Remark 4.37. We did not equate the coefficients at ν in (21) and (22). Doing this also for $(\sigma_{uv})_v = (\sigma_{vv})_u$ one obtains two equations

$$\begin{aligned} L_v - M_u &= L\Gamma^1_{12} + M(\Gamma^2_{12} - \Gamma^1_{11}) - N\Gamma^2_{11} \\ M_v - N_u &= L\Gamma^1_{22} + M(\Gamma^2_{22} - \Gamma^2_{12}) - N\Gamma^2_{12} \end{aligned} \tag{24}$$

called the *Codazzi–Mainardi equations.* We will revisit them later.

For special coordinate systems the formulas for the Gauss curvature take simpler forms.

Theorem 4.38. *1. If the coordinate system is orthogonal: $F = 0$, then one has*

$$K = -\frac{1}{2\sqrt{EG}}\left(\frac{\partial}{\partial u}\left(\frac{G_u}{\sqrt{EG}}\right) + \frac{\partial}{\partial v}\left(\frac{E_v}{\sqrt{EG}}\right)\right)$$

2. *If the parametrization σ is conformal: $I_\sigma = e^{2\varphi(u,v)}(du^2 + dv^2)$, then*

$$K = -e^{-2\varphi}\Delta\varphi$$

where $\Delta f = \frac{\partial^2 f}{\partial u^2} + \frac{\partial^2 f}{\partial v^2}$.

3. *If the first fundamental form is $du^2 + G(u,v)dv^2$ (surfaces of revolution with arc-length parametrized meridian or geodesic coordinates), then*

$$K = -\frac{1}{\sqrt{G}}\frac{\partial^2\sqrt{G}}{\partial u^2}.$$

Proof. 1. In the case $F = 0$ the Christoffel symbols as computed from (17) are:

$$\Gamma^1_{11} = \frac{E_u}{2E}, \quad \Gamma^2_{11} = -\frac{E_v}{2G}, \quad \Gamma^1_{12} = \frac{E_v}{2E}, \quad \Gamma^2_{12} = \frac{G_u}{2G}, \quad \Gamma^1_{22} = -\frac{G_u}{2E}, \quad \Gamma^2_{22} = \frac{G_v}{2G}$$

By substituting into the second of the Gauss equations (23) and computing for a while one obtains the formula.

Points 2. and 3. follow from 1. by the appropriate substitutions. □

4.11 Gauss-Bonnet Theorem

Lecture 10

Gauss-Bonnet theorem is one of the most elegant and deep results of the classical surface theory. It relates the integral of the Gaussian curvature over a closed surface with the topology of the surface. A special case of the Gauss-Bonnet theorem was mentioned in Remark 3.29: the integral of the Gaussian curvature over a strictly convex surface is equal to 4π. We will prove that the integral is 4π for any smooth surface homeomorphic to the sphere, and moreover

$$\int_M K\, dA = 2\pi(2 - 2g)$$

for any smooth surface homeomorphic to the "sphere with g handles", see Figure 30.

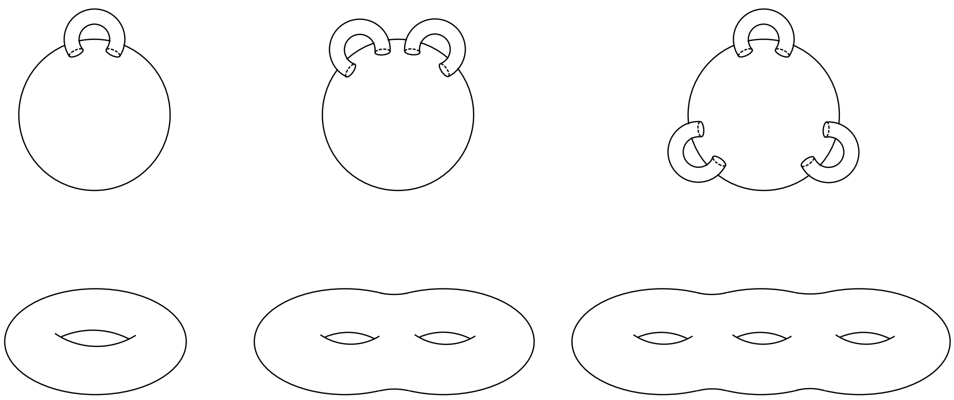

Figure 30: Spheres with one, two, three handles.

Remark 4.39. Every closed surface in $\mathbb{R}^3$ is homeomorphic to the sphere with g handles. The number g is called the *genus* of the surface. As an exercise, imagine the cube with three holes drilled through the center, each hole connecting a pair of opposite faces (a polyhedral model: from a 3×3-cube remove the central cube and six cubes at the centers of the faces) and compute the genus of its surface.

The proof of the Gauss-Bonnet theorem will be done in several steps, the result of each step interesting on its own. Our proof follows [Pre10].

The first step is to integrate the Gaussian curvature over a small disk. Consider a surface patch $\sigma\colon U \to \mathbb{R}^3$ and a disk $D \subset \sigma(U)$ bounded by a curve γ. Choose a unit normal field ν along D. We say that γ is *positively oriented* with respect to ν if the vector $\nu \times \dot\gamma$ points inside D (in other words, D lies on your left hand side if you walk along γ with the normal field pointing upwards).

Theorem 4.40. *Let γ be a simple closed curve in the image of a surface patch, positively oriented with respect to a unit normal field ν and bounding a disk D. Then one has*

$$\int_\gamma \kappa_g \, dL = 2\pi - \int_D K \, dA,$$

where κ_g is the geodesic curvature of γ with respect to the normal field ν.

In what follows, we assume γ to be parametrized with the unit speed.

Example 4.41. If the surface is developable, then because of the invariance of κ_g and K under isometry one can assume $D \subset \mathbb{R}^2$. The Gaussian curvature is then of course equal to 0 and the theorem follows from Hopf's Umlaufsatz, Theorem 1.36.

In our proof of Theorem 4.40 we will make use of a *moving orthonormal frame*. We have already chosen a unit normal field ν along D. Complement it in every point $p \in D$ to an orthonormal basis (e_1, e_2, ν) of $\mathbb{R}^3$ in such a way that e_1 and e_2 form smooth tangent vector fields. A moving frame can be constructed by applying the Gram-Schmidt process to the coordinate vector fields σ_u, σ_v at all points of D at the same time:

$$e_1 = \frac{\sigma_u}{\|\sigma_u\|}, \quad e_2 = \frac{\sigma_v - \langle \sigma_v, e_1 \rangle e_1}{\|\sigma_v - \langle \sigma_v, e_1 \rangle e_1\|}.$$

We will assume (e_1, e_2, ν) to be positively oriented.

There is a natural moving orthonormal frame along a curve γ, namely $(\dot\gamma, \nu \times \dot\gamma, \nu)$. But this frame cannot be extended to a frame on D.

Plan of the proof of Theorem 4.40: express κ_g and K in terms of the derivatives of the moving frame, then apply Green's theorem.

Lemma 4.42. *Along the curve γ, let θ be the signed angle from e_1 to $\dot\gamma$. Then one has*

$$\kappa_g = \dot\theta - \langle e_1, \dot e_2 \rangle.$$

Proof. By definition $\kappa_g = \langle \nu \times \dot\gamma, \ddot\gamma \rangle$. Let us express the right hand side in terms of the moving frame.

By definition of θ one has

$$\dot\gamma = e_1 \cos\theta + e_2 \sin\theta,$$

which implies

$$\nu \times \dot\gamma = -e_1 \sin\theta + e_2 \cos\theta.$$

On the other hand,

$$\ddot\gamma = \dot e_1 \cos\theta + \dot e_2 \sin\theta + \dot\theta(-e_1 \sin\theta + e_2 \cos\theta).$$

Taking into account the relations

$$\langle e_1, e_2 \rangle = \langle e_1, \dot e_1 \rangle = \langle e_2, \dot e_2 \rangle = 0, \quad \langle \dot e_1, e_2 \rangle + \langle e_1, \dot e_2 \rangle = 0$$

we compute

$$\langle \nu \times \dot\gamma, \ddot\gamma \rangle = -\langle e_1, \dot e_2 \rangle \sin^2\theta + \dot\theta \sin^2\theta + \langle \dot e_1, e_2 \rangle \cos^2\theta + \dot\theta \cos^2\theta = \dot\theta - \langle e_1, \dot e_2 \rangle.$$

□

Lemma 4.43. *For every positively oriented orthonormal frame (e_1, e_2) one has*

$$\langle (e_1)_u, (e_2)_v \rangle - \langle (e_1)_v, (e_2)_u \rangle = K\sqrt{\det I}.$$

Proof. Decompose the coordinate derivatives of e_1 and e_2 in our orthonormal frame:

$$\begin{aligned}
(e_1)_u &= a_{1u} e_1 + b_{1u} e_2 + c_{1u} \nu \\
(e_1)_v &= a_{1v} e_1 + b_{1v} e_2 + c_{1v} \nu \\
(e_2)_u &= a_{2u} e_1 + b_{2u} e_2 + c_{2u} \nu \\
(e_2)_v &= a_{2v} e_1 + b_{2v} e_2 + c_{2v} \nu
\end{aligned}$$

Due to $\langle (e_1)_u, e_1\rangle = 0$ etc. one has

$$a_{1u} = a_{1v} = b_{2u} = b_{2v} = 0.$$

It follows that

$$\langle (e_1)_u, (e_2)_v\rangle - \langle (e_1)_v, (e_2)_u\rangle = c_{1u}c_{2v} - c_{1v}c_{2u}.$$

This can be transformed further as

$$\begin{aligned} c_{1u}c_{2v} - c_{1v}c_{2u} &= \langle \nu, (e_1)_u\rangle\langle \nu, (e_2)_v\rangle - \langle \nu, (e_2)_u\rangle\langle \nu, (e_1)_v\rangle \\ &= \langle \nu_u, e_1\rangle\langle \nu_v, e_2\rangle - \langle \nu_u, e_2\rangle\langle \nu_v, e_1\rangle = \langle \nu_u \times \nu_v, e_1 \times e_2\rangle = \langle \nu_u \times \nu_v, \nu\rangle \end{aligned}$$

and one gets the formula of the lemma by noting that

$$\nu_u \times \nu_v = S(\sigma_u) \times S(\sigma_v) = \det S \cdot \sigma_u \times \sigma_v = K \cdot \sigma_u \times \sigma_v = K\sqrt{\det I} \cdot \nu.$$

□

Lemma 4.44. *One has*

$$\int_\gamma \dot\theta \, dL = 2\pi.$$

Sketch of proof. This can be reduced to Hopf's Umlaufsatz: since $\theta(t)$ is well-defined up to an integer multiple of 2π, it changes by $2\pi k$ as we run along γ, which means that the integral of $\dot\theta$ equals $2\pi k$. Now, shrink the curve γ inside D to a point. The integral changes continuously, and since it is a multiple of 2π it cannot change at all. When the curve is very small it looks like a curve in the plane, the vector fields e_1 and e_2 are almost constant, and therefore the integral of $\dot\theta$ equals 2π. □

Proof of Theorem 4.40. By Lemmas 4.42 and 4.44 one has

$$\int_\gamma \kappa_g \, dL = 2\pi - \int_\gamma \langle e_1, \dot e_2\rangle \, dL.$$

We claim that $\int_\gamma \dot\theta \, dL = 2\pi$.

It remains to compute the last integral. By the chain rule one has $\dot e_2 = (e_2)_u \dot u + (e_2)_v \dot v$, which implies

$$\int_\gamma \langle e_1, \dot e_2\rangle \, dL = \int_{\sigma^{-1}\circ\gamma} \langle e_1, (e_2)_u\rangle \, du + \langle e_1, (e_2)_v\rangle \, dv.$$

By Green's theorem one has

$$\begin{aligned} &\int_{\sigma^{-1}\circ\gamma} \langle e_1, (e_2)_u\rangle \, du + \langle e_1, (e_2)_v\rangle \, dv \\ &\qquad = \int_{\sigma^{-1}(D)} (\langle e_1, (e_2)_v\rangle_u - \langle e_1, (e_2)_u\rangle_v) \, dudv \\ &\qquad\qquad = \int_{\sigma^{-1}(D)} (\langle (e_1)_u, (e_2)_v\rangle - \langle (e_1)_v, (e_2)_u\rangle) \, dudv. \end{aligned}$$

Now apply Lemma 4.43:

$$\int_{\sigma^{-1}(D)} (\langle (e_1)_u, (e_2)_v\rangle - \langle (e_1)_v, (e_2)_u\rangle)\, dudv = \int_{\sigma^{-1}(D)} K\sqrt{\det I}\, dudv = \int_D K\, dA,$$

and the theorem is proved. □

We now generalize Theorem 4.40 to piecewise smooth curves.

Theorem 4.45. *Let C be a non-self-intersecting curvilinear polygon bounding a disk D in the image of a surface patch.*

Then one has

$$\int_C \kappa_g\, dL + \sum_{i=1}^n \delta_i = 2\pi - \int_D K\, dA,$$

where κ_g is the geodesic curvature of C, computed for a positive orientation of C, and $\delta_i \in (-\pi, \pi)$ is the exterior angle at the i-th vertex of C.

A *curvilinear polygon* is a cyclically ordered set of smooth curves such that the starting point of each curve coincides with the endpoint of the previous curve. These points are called the vertices, and the exterior angle at a vertex is the angle from the velocity vector of the previous curve to the velocity vector of the next curve, see Figure 31.

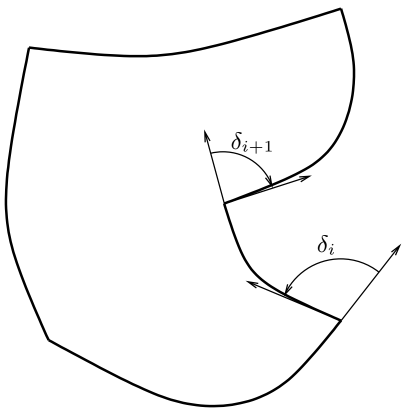

Figure 31: Exterior angles of a curvilinear polygon.

Proof. Use the same argument as in the proof of Theorem 4.40, but instead of Lemma 4.44 prove

$$\int_C \dot\theta\, dL + \sum_{i=1}^n \delta_i = 2\pi.$$

The left hand side is the total turning angle of the oriented tangent, where the direction of the tangent has some discontinuities. One can reduce this equation to the equation for smooth curves by "cutting the corners": approximate the polygon by a smooth curve and show that the total curvature of the i-th round corner is equal to δ_i. □

In order to formulate and prove the Gauss-Bonnet theorem for closed surfaces we need a couple of new notions.

Definition 4.46. *A* polygonal subdivision *of a closed smooth surface M is a finite set of points on M together with non-intersecting simple curves between these points such that the union of all curves subdivides the surface into a set of topological disks.*

The points, curves, and the disks are called vertices, edges*, and* faces *of the subdivision.*

Definition 4.47. *Let $(M, \mathcal{T})$ be a surface equipped with a subdivision $\mathcal{T}$. Its* Euler characteristic *is defined as*

$$\chi(M, \mathcal{T}) = V - E + F,$$

where V, E, and F are the numbers of the vertices, edges, and faces of the subdivision.

Theorem 4.48. *The Euler characteristic does not depend on the choice of the subdivision of a surface M.*

We do not prove this theorem.

Lemma 4.49. *The Euler characteristic of a sphere with g handles is equal to $2 - 2g$.*

Proof. This can be proved by induction. By looking at your favorite subdivision of the sphere you find that $\chi(\mathbb{S}^2) = 2$. For the induction step, handle attachment, refine the subdivision so that to have a pair of disjoint triangles, then replace these triangles by three quadrilaterals as shown in Figure 32. After this operation E increases by 3 and F increases by 1, therefore χ decreases by 2. □

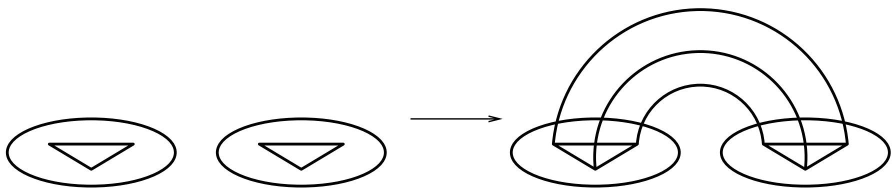

Figure 32: Handle attachment.

Theorem 4.50. *For every closed smooth surface $M \subset \mathbb{R}^3$ the integral of the Gauss curvature is equal to 2π times the Euler characteristic of M:*

$$\int_M K\, dA = 2\pi\chi(M) = 2\pi(2 - 2g).$$

Proof. Let $D_1, \ldots, D_m$ be the faces of a subdivision of M, and let n_i be the number of vertices of D_i. Write the formula of Theorem 4.45 for each D_i while expressing the exterior angle δ_{iv} at a vertex v of D_i in terms of the corresponding interior angle $\alpha_{iv} = \pi - \delta_{iv}$:

$$\int_{C_i} \kappa_g \, dL = \sum_{v \in D_i} \alpha_{iv} - (n_i - 2)\pi - \int_{D_i} K \, dA.$$

Summing these equations up we get the following expression for the integral of the Gauss curvature:

$$\int_M K \, dA = -\sum_{i=1}^m \int_{C_i} \kappa_g \, dL + \sum_{i=1}^m \sum_{v \in D_i} \alpha_{iv} - \pi \sum_{i=1}^m n_i + 2\pi F.$$

The first sum on the right hand side vanishes, because every edge belongs to two faces, thus its geodesic curvature is integrated twice but for different orientations of this edge, and changing the orientation changes the sign of the geodesic curvature.

Changing the order of summation in the double sum we get the sum over all vertices of the subdivision of the sums of all angles around a vertex, that is 2π times the number of vertices.

The third sum on the right hand side equals to $2E$ because every edge belongs to two faces and is therefore counted twice.

As a result we get

$$\int_M K \, dA = 2\pi V - 2\pi E + 2\pi F = 2\pi\chi(M)$$

and the theorem is proved. □

Remark 4.51. Note that Theorem 4.50 implies Theorem 4.48.

5 Smooth manifolds

Riemannian geometry takes the intrinsic geometry of a smooth surface as inspiration: one concentrates on the geometric properties which can be extracted from the length measurement on the surface. And there is quite a lot to extract, as we have seen in Section 4. But then, since the ambient space is of no use, we do not care if it exists. Imagine an abstract surface with length measurement given by sort of first fundamental form, that is with a positive definite quadratic form attached to each point. This generalization of the first fundamental form is called a *Riemannian metric.*

Example 5.1. The family of quadratic forms $d\rho^2 + \sinh^2 \rho \, d\theta^2$, where (ρ, θ) are polar coordinates, is a Riemannian metric in the plane. If we put ρ in place of $\sinh \rho$, then we get the usual Euclidean metric; if we use $\sin \rho$ and restrict to $\theta \in [0, 2\pi]$, then we get the metric on the unit sphere in $\mathbb{R}^3$. The above expression cannot be realized as the first fundamental form of a surface in $\mathbb{R}^3$ (to be more exact, one can do this locally, but not globally, that is for the whole (ρ, θ)-plane). Nevertheless it allows to measure lengths and areas by the usual formulas and leads to a very nice and famous geometry.

Example 5.2. The topological space $\mathbb{S}^1 \times \mathbb{S}^1$ (the Cartesian product of two circles) is called the torus (it is obviously homeomorphic to a torus of revolution which we met at some point). On this space, consider the Riemannian metric $d\alpha^2 + d\beta^2$. Locally, this is the usual Euclidean metric and therefore has Gaussian curvature 0. If we would have taken the cylinder $\mathbb{S}^1 \times \mathbb{R}$ instead of the torus, we could have realized it as a surface in $\mathbb{R}^3$ with exactly this first fundamental form. But there is no smooth torus in $\mathbb{R}^3$ with this first fundamental form.

There is no need to restrict oneself to surfaces, one can study their higher-dimensional generalizations, called manifolds. The general theory of relativity claims that we are living in a four-dimensional Lorentzian manifold. Lorentzian geometry is a simple generalization of the Riemannian geometry: instead of a family of positive definite quadratic forms one considers a family of quadratic forms of signature $(-,+,+,+)$. Interestingly, the general relativity was discovered about 60 years after the invention of Riemannian geometry. Thus, an abstract mathematical theory has preceded a physical theory later confirmed by experiments. In the words of the physicist Eugene Wigner, this as an instance of “the unreasonable effectiveness of mathematics in the natural sciences”.

In the above outline we left some points unclarified. What is an abstract surface and what is a manifold? We also said that a Riemannian metric is a family of quadratic forms. These quadratic forms must live in some vector spaces, so we should define tangent spaces to abstract surfaces and manifolds.

5.1 Topological and smooth manifolds

For this Section see [Laf15, Chapter 2].

Definition 5.3. *An n-dimensional* topological manifold *M is a Hausdorff, second-countable, and connected topological space such that every point of M has an open neighborhood homeomorphic to an open subset of $\mathbb{R}^n$.*

We have no time to go into the Hausdorff and the second-countability conditions. They are needed to avoid some pathologies, the typical examples being the line with two origins and a long line.

The connectedness assumption is sometimes omitted. The connected components of a “disconnected manifold” are also manifolds, and can be studied separately.

Note that we study only manifolds without boundary. For example, a closed disk is not a manifold in the above sense: points on its boundary do not have appropriate neighborhoods.

Example 5.4. A smooth surface in $\mathbb{R}^3$ is a 2-dimensional topological manifold. Indeed, the diffeomorphism Φ from Definition 2.1 is in particular a homeomorphism, and so is its restriction to $V \cap M$. Thus $V \cap M$ is homeomorphic to an open subset of $\mathbb{R}^2$.

Two-dimensional manifolds are called (abstract) surfaces.

Example 5.5. Take a square and glue its sides in pairs as shown in Figure 33, left: the sides marked by the same letter are glued in accordance with the directions of arrows. (The gluing operation can be defined formally, through the quotient topology, but we do not go into details.) As can be easily recognized, the result is a torus.

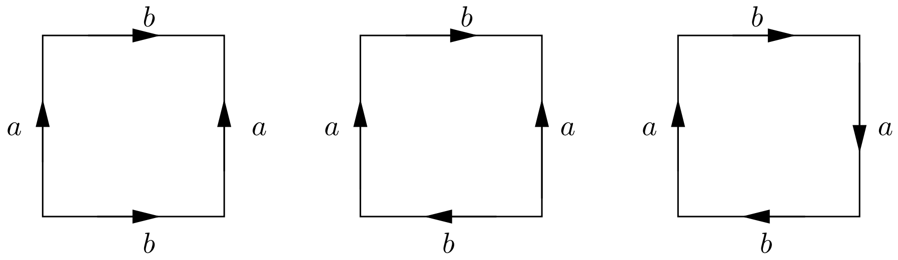

Figure 33: The torus, the Klein bottle, and the projective plane.

If one reverses the gluing orientation for one pairs of sides as shown in Figure 33, middle, then the result is also a surface, called the *Klein bottle*. You can try to visualize it by first gluing the a-sides. The result is a cylinder. Now you have to glue the bases of this cylinder, but in a wrong way. The result is an abstract surface (indeed, any point on the line of gluing has a neighborhood which is homeomorphic to an open subset of the plane), but it cannot be realized as a subset of $\mathbb{R}^3$.

If one reverses the relative orientation for both pairs of opposite sides, then the result is even more difficult to visualize. This surface is called the *projective plane*, and it is indeed homeomorphic to the quotient $(\mathbb{R}^3 \setminus \{0\})/x \sim \lambda x$.

Similarly to a square one can take any polygon with an even number of sides and identify sides in pairs. The result is always a surface.

Remark 5.6. One can show that each compact orientable surface is homeomorphic to a sphere with handles, and each compact non-orientable surface is homeomorphic to a projective plane with handles (the Klein bottle is the projective plane with one handle). Classification of 3-dimensional manifolds is much more complicated.

Definition 5.7. *A* chart *of a topological manifold M is a pair (U, φ) consisting of an open subset U of M and a homeomorphism φ from U to an open subset of $\mathbb{R}^n$.*

A chart is also called a *local coordinate system.*

This reminds very much of surface patches for surfaces in $\mathbb{R}^3$. The inverses of surface patches are charts. By definition of a manifold, every point lies in the domain of some chart.

Definition 5.8. *An* atlas *of a topological manifold M is a family (U_i, φ_i) of charts such that the domains U_i cover M.*

For example, one can construct atlases for the torus and for the Klein bottle made of two charts each, and an atlas for the projective plane made of three charts.

Definition 5.9. *Let (U, φ) and (V, ψ) be two charts of a topological manifold M such that $U \cap V \neq \emptyset$. Then the map*

$$\psi \circ \varphi^{-1} \colon \varphi(U \cap V) \to \psi(U \cap V)$$

is called the transition map *from φ to ψ.*

This is again similar to the transition map between surface patches. See Figure 34, which is a recycled Figure 20.

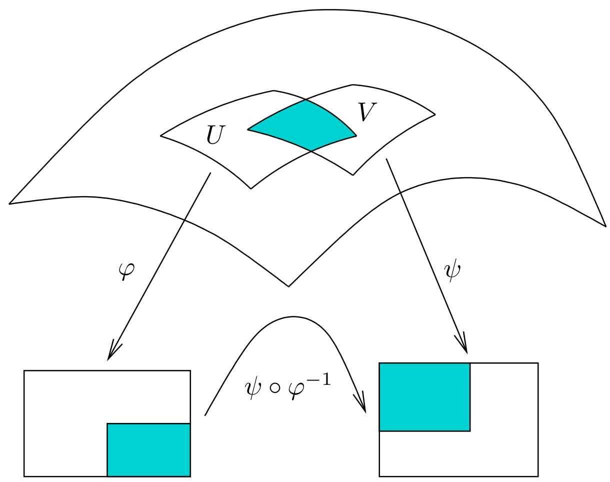

Figure 34: Transition map between two charts of a manifold.

While U and V live in some abstract topological space, their transition map is a homeomorphism between open subsets of $\mathbb{R}^n$. An atlas whose transition maps are special in some way introduces a structure on the manifold on the top of the topological structure. For example, if all transition maps between charts of a given atlas are orientation-preserving at all points, then this atlas defines an *orientation* of a manifold. Not all manifolds are orientable, for example the Klein bottle and the projective plane are not.

Definition 5.10. *Two charts on a topological manifold M are called* compatible *if their transition map is a diffeomorphism.*

A smooth atlas *on a topological manifold M is an atlas whose charts are pairwise compatible.*

By smooth we always understand C^∞-smooth, likewise a diffeomorphism for us is a C^∞-diffeomorphism.

Definition 5.11. *A smooth atlas on a topological manifold M is called* maximal *if it contains every chart compatible with the atlas charts. A maximal smooth atlas is also called a* smooth structure.

A smooth manifold *is a topological manifold equipped with a smooth structure.*

A maximal atlas contains absurdly many charts, however this is a convenient way to distinguish between different smooth structures. Any smooth atlas $\mathcal{A}$ defines a smooth structure: there is a unique maximal atlas containing $\mathcal{A}$. It is easy to show that two different atlases define the same smooth structure if and only if every chart from the first atlas is compatible with every chart from the second atlas.

Lemma 2.8 implies that any smooth surface in $\mathbb{R}^3$ is a smooth 2-dimensional manifold.

Definition 5.12. *Let M and N be two smooth manifolds. A continuous map F from M to N is said to be* smooth *at the point $p \in M$ if for some (and then any) pair of charts (U, φ) for M and (V, ψ) for N such that $p \in U$ and $f(p) \in V$ the map $\psi \circ F \circ \varphi^{-1}$ is smooth at p.*

If the composition $\psi \circ F \circ \varphi^{-1}$ is smooth at p for some φ and ψ, then it remains smooth with φ replaced by φ' and ψ replaced by ψ' because the transition maps from φ' to φ and from ψ to ψ' are smooth.

What we are now doing repeats Section 2.7 very closely. One can now define smooth functions and smooth curves, but these are actually already defined as a special case of Definition 5.12. Namely, $\mathbb{R}$ is a smooth 1-dimensional manifold with a canonical smooth structure defined by the atlas consisting of a single chart, the identity map $\mathbb{R} \to \mathbb{R}$. Smooth maps $M \to \mathbb{R}$ are called *smooth functions* on M, and smooth maps $\mathbb{R} \to M$ are called *smooth curves* in M. One can also consider curves $I \to M$, where $I \subset \mathbb{R}$ is an open interval with the smooth structure determined by the identity map $I \to I \subset \mathbb{R}$.

Similarly to Section 2.7 one proves the following theorem.

Theorem 5.13. *Composition of smooth maps is a smooth map.*

And one gives the following definition.

Definition 5.14. *A* diffeomorphism *between two smooth manifolds is a smooth bijection with a smooth inverse.*

5.2 Submanifolds

Lecture 11

Definition 5.15. *A subset M of a smooth n-dimensional manifold N is called an m-dimensional* smooth submanifold *of N if M is connected and for every $p \in M$ there is an open subset V of N containing p, an open subset W of $\mathbb{R}^n$, and a diffeomorphism $\Phi\colon V \to W$ such that*

$$\Phi(V \cap M) = W \cap (\mathbb{R}^m \times \{0\}).$$

See Figure 1.5 for an illustration.

In particular, $V \cap M$ is homeomorphic to an open subset of $\mathbb{R}^m$. This defines an atlas and makes M to a topological manifold. The transition maps of this atlas are smooth, thus M is a smooth manifold.

In particular, smooth surfaces in $\mathbb{R}^3$ are smooth submanifolds of $\mathbb{R}^3$.

Definition 5.16. *A smooth map $F\colon M \to N$ between smooth manifolds is called an* embedding *if $F(M)$ is a submanifold of N and F is a diffeomorphism of M onto $F(M)$.*

Theorem 5.17. *(Whitney) Every compact n-dimensional manifold can be embedded into $\mathbb{R}^{2n}$.*

We do not prove this theorem. The proof of a slightly weaker statement about the embeddability into $\mathbb{R}^{2n+1}$ can be found in [Laf15, pp. 99–102], and this is quite a nice argument.

Remark 5.18. The projective plane cannot be embedded into $\mathbb{R}^3$, but can be embedded into $\mathbb{R}^4$ by the Whitney embedding theorem. There is a nice embedding of the projective plane into $\mathbb{R}^5$, the Veronese embedding, which can be further modified to yield an embedding into $\mathbb{R}^4$.

Thus every smooth manifold can be viewed as a submanifold of some Euclidean space.

5.3 Tangent space and the differential of a smooth map

For a surface in $\mathbb{R}^3$, we defined tangent vectors as the velocity vectors of smooth curves in the surface. Now velocity vectors do not make sense, but the curves will still be helpful.

Let $p \in M$ be a point in a smooth manifold. Denote by $\mathcal{C}_p^M$ the set of all smooth curves $\gamma\colon I \to M$ such that $0 \in I$ and $\gamma(0) = p$.

Definition 5.19. *Let us call two curves $\gamma_1\colon I_1 \to M$ and $\gamma_2\colon I_2 \to M$* tangent *at p if $\gamma_1, \gamma_2 \in \mathcal{C}_p^M$ and there is a chart (U, φ) such that $p \in U$ and*

$$\left.\frac{d}{dt}\right|_{t=0} (\varphi \circ \gamma_1) = \left.\frac{d}{dt}\right|_{t=0} (\varphi \circ \gamma_2).$$

This condition is independent of the choice of a chart: if (V, ψ) is another chart with $p \in V$, then the chain rule for a composition $\mathbb{R} \to \mathbb{R}^n \to \mathbb{R}^n$ yields

$$\left.\frac{d}{dt}\right|_{t=0} (\psi \circ \gamma_i) = d(\psi \circ \varphi^{-1})_{\psi(p)} \cdot \left.\frac{d}{dt}\right|_{t=0} (\varphi \circ \gamma_i).$$

It follows that being tangent at p is an equivalence relation on $\mathcal{C}_p^M$.

Definition 5.20. *Let M be a smooth manifold, and let $p \in M$. A* tangent vector *to M at p is an equivalence class of the equivalence relation above. The set of all tangent vectors to M at p is denoted T_pM and called the* tangent space *to M at p.*

We now want to equip T_pM with a vector space structure. For any chart (U, φ) with $p \in U$, consider the map

$$\theta_\varphi\colon T_pM \to \mathbb{R}^n, \quad \theta_\varphi(X) = \left.\frac{d}{dt}\right|_{t=0} (\varphi \circ \gamma), \tag{25}$$

where γ is any representative of the tangent vector X. This map is well-defined and injective by definition of tangent vectors. The map θ_φ is also surjective: for any $v \in \mathbb{R}^n$, the equivalence class of the curve $t \mapsto \varphi^{-1}(\varphi(p) + tv)$ is mapped to v. Thus the map (25) is a bijection.

Define the vector space structure on T_pM by inducing it from $\mathbb{R}^n$ through the map (25):

$$X + Y := \theta_\varphi^{-1}(\theta_\varphi(X) + \theta_\varphi(Y)), \quad \lambda X := \theta_\varphi^{-1}(\lambda\theta_\varphi(X)).$$

One has to check that this is independent of the choice of a chart. But for any other chart (V, ψ) with $p \in V$ the composition

$$\theta_\psi \circ \theta_\varphi^{-1} = d(\psi \circ \varphi^{-1})_{\psi(p)}$$

is a linear map $\mathbb{R}^n \to \mathbb{R}^n$. Therefore it does not matter whether we add elements of T_pM by adding their images under θ_φ or under θ_ψ.

Definition 5.21. *Let $F\colon M \to N$ be a smooth map between smooth manifolds (of dimensions m and n respectively). The* differential *of F at a point $p \in M$ is the map*

$$dF_p\colon T_pM \to T_{F(p)}N$$

obtained from the map

$$\mathcal{C}_p^M \to \mathcal{C}_{F(p)}^N, \quad \gamma \mapsto F \circ \gamma$$

by descending to the equivalence classes.

For the differential to be well-defined one has to show that if two curves $\gamma_1, \gamma_2 \in \mathcal{C}_p^M$ are tangent, then the curves $F \circ \gamma_1, F \circ \gamma_2 \in \mathcal{C}_{F(p)}^N$ are also tangent. This follows from the chain rule:

$$\left.\frac{d}{dt}\right|_{t=0} (\psi \circ F \circ \gamma_i) = d(\psi \circ F \circ \varphi^{-1}) \cdot \left.\frac{d}{dt}\right|_{t=0} (\varphi \circ \gamma_i).$$

The above formula also shows that the differential is a linear map. Indeed, in the commutative diagram

$$\begin{tikzcd} T_pM \arrow[r, "dF_p"] \arrow[d, "\theta_\varphi"'] & T_{F(p)}N \arrow[d, "\theta_\psi"'] \\ \mathbb{R}^m \arrow[r, "d(\psi\circ F\circ\varphi^{-1})"] & \mathbb{R}^n \end{tikzcd}$$

the maps θ_φ and θ_ψ are linear isomorphisms by definition, and $d(\psi \circ F \circ \varphi^{-1})$ is a linear map.

Extending the above commutative diagram one can prove the chain rule for the differentials.

Theorem 5.22. *Let $F\colon M \to N$ and $G\colon N \to P$ be smooth maps between smooth manifolds. Then one has*

$$d(G \circ F)_p = dG_{F(p)} \circ dF_p.$$

5.4 Vector fields

Let us now consider not a single vector at an individual point p, but a vector field, a choice of a tangent vector X_p at every point $p \in M$. The vector X_p must depend smoothly on the point p. To explain what the smooth dependence means, we as always use charts. Recall the map θ_φ which identifies T_pM with $\mathbb{R}^n$, see (25). There is such a map for every $p \in U$, so together they yield

$$\theta_\varphi \colon \bigcup_{p\in U} T_pM \to \mathbb{R}^n. \tag{26}$$

Definition 5.23. *The* tangent bundle *of a smooth manifold M is the union of all tangent spaces to M:*

$$TM := \bigcup_{p\in M} T_pM.$$

A vector field *on a smooth manifold M is a map $X \colon M \to TM$ such that the image of p belongs to T_pM.*

A vector field X is called smooth *if for every chart (U, φ) of M the map*

$$\theta_\varphi \circ X \colon U \to \mathbb{R}^n,$$

where θ_φ is as in (26), is smooth.

Remark 5.24. The tangent bundle TM is a smooth manifold of dimension twice that of M (a smooth atlas can be introduced with the help of the maps θ_φ). In these terms, a smooth vector field is a smooth map $X \colon M \to TM$ which sends every p in T_pM. (Maps with the latter property are also called *sections* of the projection $TM \to M$.)

All vector fields will be assumed smooth in the sequel.

5.5 Vectors and vector fields as derivations

Let M be a smooth manifold, p a point in M, and X_p a tangent vector to M at p. As in Section 3.1, one defines the derivative of a smooth function f in the direction X as

$$D_{X_p}f := \left.\frac{d}{dt}\right|_{t=0} f(\gamma(t)),$$

where $\gamma \in \mathcal{C}_p^M$ is any representative of X_p. In the same way as before (except that the place of a surface patch is taken by a chart) one proves that $D_{X_p}f$ is well defined and that for any function f the map $X_p \mapsto D_{X_p}f$ is linear.

Remark 5.25. The function f can be viewed as a smooth map from M to a 1-dimensional smooth manifold $\mathbb{R}$. As such, it has a differential df_p at the point p, which is a linear map

$$df_p \colon T_pM \to T_{f(p)}\mathbb{R} \cong \mathbb{R}.$$

(From the definition of the tangent space given in Section 5.3 it follows that the tangent space to $\mathbb{R}^n$ is canonically isomorphic to $\mathbb{R}^n$.) One can prove that

$$D_{X_p} f = df_p(X_p).$$

Now let us fix a vector $X_p \in T_pM$ and consider the map

$$D_{X_p} \colon C^\infty(M) \to \mathbb{R}, \quad f \mapsto D_{X_p} f.$$

Definition 5.26. *A* derivation at a point $p \in M$ *is an* $\mathbb{R}$*-linear map* $D_p \colon C^\infty(M) \to \mathbb{R}$ *such that for any* $f, g \in C^\infty(M)$ *one has*

$$D_p(fg) = D_p f \cdot g(p) + f(p) \cdot D_p g. \tag{27}$$

It is easy to see that for every $X_p \in T_pM$ the map D_{X_p} is a derivation at p. What is less obvious is the following theorem.

Theorem 5.27. *For every derivation* D_p *at* p *there is a vector* $X_p \in T_pM$ *such that* $D_p = D_{X_p}$.

Lemma 5.28. *For every derivation* D_p *one has* $D_p c = 0$ *for every constant function* c.

Proof. Due to the $\mathbb{R}$-linearity of D_p it suffices to prove $D_p 1 = 0$, where 1 stands for the constant function equal to 1. This follows by setting $f = g = 1$ in the product formula:

$$D_p(1 \cdot 1) = D_p 1 \cdot 1 + 1 \cdot D_p 1 \Rightarrow D_p 1 = 2D_p 1 \Rightarrow D_p 1 = 0.$$

$\square$

Proof of Theorem 5.27. Let us show that $D_p f$ depends only on the values of f in an arbitrarily small neighborhood of p. By the additivity of D_p it suffices to show that if $f(x) = 0$ for all $x \in U$ for some open set $U \ni p$, then $D_p f = 0$. Take a function g such that $g(p) = 1$ and $g(x) = 0$ outside of U. Then $fg = 0$, and one has

$$0 = D_p(fg) = D_p f \cdot g(p) + f(p) \cdot D_p g = D_p f.$$

Due to the local nature of D_p one can take any chart (U, φ) around p and move the problem to $\mathbb{R}^n$: we have to show that every derivation $C^\infty(\mathbb{R}^n) \to \mathbb{R}$ at 0 (wlog $\varphi(p) = 0$) is a directional derivative at 0. This is done with the help of the Hadamard lemma: every smooth function f can be written as

$$f(x) = f(0) + \sum_{i=1}^n x_i h_i(x),$$

where h_i are smooth and $h_i(0) = \frac{\partial f}{\partial x_i}$. Observing that every derivation sends a constant function to 0, one gets

$$D_p f = \sum_{i=1}^n D_p(x_i) h_i(0) = \sum_{i=1}^n D_p(x_i) \frac{\partial f}{\partial x_i} = D_{X_p} f,$$

where $X_p = \sum_{i=1}^n D_p(x_i) e_i$, and the theorem is proved. $\square$

All right, if we can interpret tangent vectors as derivations, what is the corresponding interpretation of vector fields? The answer is as follows.

Definition 5.29. *A* derivation *(or a* global derivation*) on a smooth manifold M is an $\mathbb{R}$-linear map $D\colon C^\infty(M) \to C^\infty(M)$ such that*

$$D(fg) = Df \cdot g + f \cdot Dg. \tag{28}$$

Comparing formula (28) with formula (27) one sees that a global derivation is a family of derivations at points of M (evaluate for this both sides of (28) at a point p). Besides, since for every smooth f the function Df must also be smooth, this family of pointwise derivations is "smooth" in some sense.

This suggests the following theorem.

Theorem 5.30. *The set of smooth vector fields on M is in a one-to-one correspondence with the set of derivations on M. The derivation D_X corresponding to a vector field X acts on smooth functions as follows:*

$$(D_X f)(p) := D_{X_p} f.$$

It is clear that D_X is a global derivation: that $D_{X_p} f$ is a smooth function of p can be proved by considering a chart. Less obvious is that any global derivation comes from a vector field. This is an analog of Theorem 5.27 and is proved in a similar way.

5.6 Commutator of vector fields

Two derivations $D_1, D_2\colon C^\infty(M) \to C^\infty(M)$ can be composed, but the resulting map is usually not a derivation. Indeed, $D_1 \circ D_2$ is $\mathbb{R}$-linear, but does not necessarily satisfy the product rule. Instead, one can do the following.

Theorem-Definition 5.31. *Let D_1 and D_2 be two derivations. Denote*

$$[D_1, D_2] := D_1 \circ D_2 - D_2 \circ D_1$$

and call it the commutator *of D_1 and D_2. Then $[D_1, D_2]$ is also a derivation.*

Proof. The $\mathbb{R}$-linearity is obvious, and the product rule is proved by a direct computation:

$$\begin{aligned}
[D_1, D_2](fg) &= D_1(D_2(fg)) - D_2(D_1(fg)) \\
&= D_1(D_2 f \cdot g + f \cdot D_2 g) - D_2(D_1 f \cdot g + f \cdot D_1 g) \\
&= (D_1 D_2 f \cdot g + D_2 f \cdot D_1 g + D_1 f \cdot D_2 g + f \cdot D_1 D_2 g) - (D_2 D_1 f \cdot g + D_1 f \cdot D_2 g + D_2 f \cdot D_1 g + f \cdot D_2 D_1 g) \\
&= (D_1 \circ D_2 - D_2 \circ D_1) f \cdot g + f \cdot (D_1 \circ D_2 - D_2 \circ D_1) g.
\end{aligned}$$

□

Since derivations on $C^\infty(M)$ bijectively correspond to vector fields on M, one can give the following definition.

Definition 5.32. *Let X, Y be two vector fields on a manifold M. The vector field corresponding to the commutator of derivations D_X and D_Y is called the* commutator *of X and Y or the* Lie bracket *of X and Y:*

$$D_{[X,Y]}f = D_X D_Y f - D_Y D_X f.$$

Lemma 5.33. *The commutator of vector fields has the following properties.*

1. *It is $\mathbb{R}$-bilinear:*

$$\begin{aligned} [X_1 + X_2, Y] &= [X_1, Y] + [X_2, Y], & [\lambda X, Y] &= \lambda[X, Y] \\ [X, Y_1 + Y_2] &= [X, Y_1] + [X, Y_2], & [X, \lambda Y] &= \lambda[X, Y] \end{aligned}$$

2. *It is antisymmetric:*

$$[X, Y] = -[Y, X]$$

3. *The map $Y \mapsto [X, Y]$ satisfies a Leibnitz property with respect to multiplication of Y by a function, and the map $X \mapsto [X, Y]$ satisfies an anti-Leibnitz property:*

$$\begin{aligned} [X, gY] &= D_X g \cdot Y + g \cdot [X, Y] \\ [fX, Y] &= -D_Y f \cdot X + f \cdot [X, Y] \end{aligned}$$

4. *It satisfies the Jacobi identity:*

$$[X, [Y, Z]] + [Z, [X, Y]] + [Y, [Z, X]] = 0$$

Proof. Exercise. □

Remark 5.34. The vector fields thus form a Lie algebra (of infinite dimension). There is a Lie group for this algebra: the group of diffeomorphisms of M. Intuitively, if one takes a diffeomorphism close to the identity, there is a corresponding vector field made of vectors pointing from points to their images. There is also the exponential map from the algebra of vector fields to the group of diffeomorphisms. Namely, a vector field generates a so-called *flow*: every point p moves along a trajectory $\Phi_t(p)$ satisfying $\frac{d}{dt}\Phi_t(p) = X(\Phi_t(p))$. This defines a family of diffeomorphisms $\Phi_t \colon M \to M$, at least for all small t. The flows provide a geometric interpretation to the commutator, $[X, Y]$ is related to $\lim_{t,s\to 0} \frac{1}{ts}(\Phi_t \circ \Psi_s - \Psi_s \circ \Phi_t)$, where Φ and Ψ are the flows of X and Y. That is, the commutator measures the non-commutativity of the flows.

Let us now compute the commutator in coordinates. A chart $\varphi \colon U \to \mathbb{R}^n$ on a manifold M defines *coordinate functions* $x^i \colon U \to \mathbb{R}$ (this is an abuse of notation, more correct would be to write $x^i \circ \varphi$, where $x^1, \ldots, x^n$ are the coordinates on $\mathbb{R}^n$). Through a point in U with coordinates $(x_0^1, \ldots, x_0^n)$ go n *coordinate curves*

$$t \mapsto (x_0^1, \ldots, t, \ldots, x_0^n).$$

The tangent vectors to these curves form n *coordinate vector fields* in U which are denoted by

$$\partial_i \text{ or } \frac{\partial}{\partial x^i}.$$

The reason for this notation is that the derivation corresponding to the i-th coordinate vector field associates to a function $f\colon U \to \mathbb{R}$ the partial derivative $\frac{\partial(f\circ\varphi^{-1})}{\partial x^i}$. The combination $\partial_i \circ \partial_j - \partial_j \circ \partial_i$ of these derivations sends every function to zero because partial derivatives commute. Thus we arrive to the following conclusion.

Lemma 5.35. *The coordinate vector fields of any chart commute.*

The coordinate vector fields of a chart (or of a local diffeomorphism) form a basis of the tangent space at every point $p \in U$. Thus every vector field can be decomposed with respect to the coordinate vector fields. Let us take any two vector fields

$$X = \sum_{i=1}^n v^i \frac{\partial}{\partial x^i}, \quad Y = \sum_{i=1}^n w^i \frac{\partial}{\partial x^i}, \quad v^i, w^i \in C^\infty(U)$$

and compute their commutator. For any function $f \in C^\infty(U)$ one has

$$Y(f) = \sum_{j=1}^n w^j \frac{\partial f}{\partial x^j},$$

$$X(Y(f)) = \sum_{i=1}^n v^i \frac{\partial}{\partial x^i} \left(\sum_{j=1}^n w^j \frac{\partial f}{\partial x^j} \right) = \sum_{i,j} v^i \frac{\partial w^j}{\partial x^i} \frac{\partial f}{\partial x^j} + \sum_{i,j} v^i w^j \frac{\partial^2 f}{\partial x^i \partial x^j}.$$

(An abuse of notation: formally correct would be to write $\frac{\partial(f\circ\varphi^{-1})}{\partial x^j}$ and $\frac{\partial(w^j\circ\varphi^{-1})}{\partial x^i}$. One can also say that we are identifying U with its image in $\mathbb{R}^n$ via φ, if φ is injective.)

There is a similar expression for $Y(X(f))$, with v and w exchanged. Subtracting one from the other we see that the terms involving the second derivatives of f cancel:

$$X(Y(f)) - Y(X(f)) = \sum_{i,j} \left(v^i \frac{\partial w^j}{\partial x^i} - w^i \frac{\partial v^j}{\partial x^i} \right) \frac{\partial f}{\partial x^j}.$$

Thus one has

$$[X, Y] = \sum_{j=1}^n \sum_{i=1}^n \left(v^i \frac{\partial w^j}{\partial x^i} - w^i \frac{\partial v^j}{\partial x^i} \right) \frac{\partial}{\partial x^j}. \tag{29}$$

This formula can be used as a definition of the commutator (thus avoiding the approach through global derivations), but requires one more step: one has to prove that $[X, Y]$ is well-defined.

Another method of computation of the commutator is to use Lemma 5.33. For example,

$$[\partial_x, x\partial_y] = \partial_x x \cdot \partial_y + x \cdot [\partial_x, \partial y] = \partial_y.$$

Example 5.36. In the domain $(0, +\infty) \times \mathbb{R}$ with the coordinates (ρ, θ) consider the vector fields

$$X = \cos\theta \frac{\partial}{\partial\rho} - \frac{\sin\theta}{\rho}\frac{\partial}{\partial\theta}, \quad Y = \sin\theta \frac{\partial}{\partial\rho} + \frac{\cos\theta}{\rho}\frac{\partial}{\partial\theta}.$$

After a lengthy computation one arrives at $[X, Y] = 0$. This is not surprising in view of Lemma 5.35 since $X = \frac{\partial}{\partial x}$ and $Y = \frac{\partial}{\partial y}$ for $(x, y) = (\rho\cos\theta, \rho\sin\theta)$. In other words, X and Y are the coordinate vector fields of the coordinate system on $(0, +\infty) \times \mathbb{R}$ obtained by mapping $(0, +\infty) \times \mathbb{R}$ to $\mathbb{R}^2$ through the above formulas. Note that the map is not injective: it wraps the half-plane $(0, +\infty) \times \mathbb{R}$ around the origin of $\mathbb{R}^2$ infinitely many times.

5.7 Tensor product of vector spaces

Lecture 12

The notion of a tensor unifies the concepts of a vector, a linear functional, a linear homomorphism, and a bilinear form. The most direct approach to tensors is to represent them as a multiindexed collection of numbers: vectors and covectors require one index and thus have n components; linear endomorphisms and bilinear forms require two indices, thus have n^2 components which can be written in a matrix; tensors of order r have r indices, that is n^r components. But one gets a broader perspective by starting from a coordinate-free viewpoint.

First, let us recall and possibly update our linear algebra. An isomorphism between two vector spaces is called *canonical* if it does not depend on the choice of bases. Any two n-dimensional vector spaces are isomorphic to each other and to $\mathbb{R}^n$ in particular, but canonical isomorphisms are rare. If V is an n-dimensional vector space, then its dual

$$V^* = \{\alpha \colon V \to \mathbb{R} \mid \alpha \text{ is linear}\}$$

is also an n-dimensional vector space. To every basis $(e_1, \ldots, e_n)$ of V there corresponds the *dual basis* $(\eta^1, \ldots, \eta^n)$ of V^* defined by

$$\eta^i(e_j) = \delta^i_j.$$

One gets an isomorphism between V and V^* by sending e_i to η^i, but if you take a different basis of V and send its elements to the elements of its dual basis, then you get a different map $V \to V^*$.

The spaces V and V^* are not canonically isomorphic, but there is a *canonical pairing* between them, that is a bilinear map $V^* \times V \to \mathbb{R}$ independent of the choice of bases. This map sends (α, v) to $\alpha(v)$, and in order to stress the equal status of V and V^* we use the notation

$$\langle \alpha, v \rangle := \alpha(v).$$

The canonical pairing leads to a canonical isomorphism $(V^*)^* \equiv V$.

A choice of a basis of V allows to represent every linear endomorphism $V \to V$ by a $n \times n$-matrix A, so that the map is written in coordinates as $x \mapsto Ax$. Every bilinear form $V \times V \to \mathbb{R}$ can also be represented by a matrix B, so that the form acts as $(x, y) \mapsto x^\top By$. Thus both the space of linear endomorphisms and the space of bilinear forms are isomorphic to $\mathbb{R}^{n^2}$, but these isomorphisms are not canonical: when you change the basis, the matrix does change. One might hope that if an endomorphism and a bilinear form are represented by the same matrix for some choice of a basis, then they are represented by the same matrix for any choice of a basis. This would establish a canonical isomorphism between endomorphisms and forms. But this hope fades when one remembers that the matrices of endomorphisms and bilinear forms behave differently under the choice of a basis:

$$A \mapsto C^{-1}AC, \quad B \mapsto C^\top BC.$$

We will now define the tensor product of vector spaces, and will do it in several different ways. In the following, all vector spaces are real and finite-dimensional.

Definition 5.37. *Let V, W, Z be vector spaces. A map $h \colon V \times W \to Z$ is called* bilinear *if it satisfies*

$$\begin{aligned} h(v_1 + v_2, w) &= h(v_1, w) + h(v_2, w), \\ h(v, w_1 + w_2) &= h(v, w_1) + h(v, w_2), \\ h(\lambda v, w) &= \lambda h(v, w) = h(v, \lambda w). \end{aligned}$$

A bilinear map to $\mathbb{R}$ is called a bilinear form.

Definition 5.38. *The* tensor product *of V and W is a vector space $V \otimes W$ such that there is a canonical correspondence between bilinear maps $V \times W \to Z$ and linear maps $V \otimes W \to Z$.*

More exactly, the tensor product of V and W is a vector space $V \otimes W$ together with a map $\varphi \colon V \times W \to V \otimes W$ such that any bilinear map $h \colon V \times W \to \mathbb{R}$ factors through φ in a unique way, that is there is a unique linear map $\widetilde{h} \colon V \otimes W \to Z$ such that $h = \widetilde{h} \circ \varphi$, see the diagram below.

$$\begin{tikzcd} V \times W \arrow[r, "\varphi"] \arrow[dr, "h"'] & V \otimes W \arrow[d, dashed, "\widetilde{h}"] \\ & Z \end{tikzcd}$$

Note that there is a linear structure on $V \times W$ well known to you, that of the direct sum $V \oplus W$. But linear maps $V \oplus W \to Z$ are different from the bilinear maps $V \times W \to Z$ (check this!).

The above definition says a lot and at the same time nothing about the vector space $V \otimes W$. Why does such a space exist? The next is a constructive definition of $V \otimes W$.

Definition 5.39. *For any set X denote by $\mathbb{R}^X$ the set of all formal finite linear combinations of elements of X:*

$$\mathbb{R}^X = \left\{ \sum_{i=1}^{N} \lambda_i x_i \mid x_i \in X, \lambda_i \in \mathbb{R} \right\}.$$

The set $\mathbb{R}^X$ *has an obvious structure of a vector space, called the* free vector space *generated by* X.

There is a natural embedding $X \to \mathbb{R}^X$, and the elements of X form a basis of $\mathbb{R}^X$.

Definition 5.40. *The* tensor product $V \otimes W$ *is the quotient space* $\mathbb{R}^{V\times W}/\Omega$, *where* Ω *is the linear subspace spanned by all elements of* $\mathbb{R}^{V\times W}$ *of the form*

$$(v_1 + v_2, w) - (v_1, w) - (v_2, w),$$
$$(v, w_1 + w_2) - (v, w_1) - (v, w_2),$$
$$(\lambda v, w) - \lambda(v, w), \quad (v, \lambda w) - \lambda(v, w).$$

The equivalence class of (v, w) is denoted by $v \otimes w$. Thus $V \otimes W$ can be described as the set of formal sums

$$\sum_{k=1}^{N} \lambda_k (v_k \otimes w_k),$$

allowed to be transformed according to the rules coming from the linear subspace Ω:

$$(v_1 + v_2) \otimes w = v_1 \otimes w + v_2 \otimes w \quad \text{etc.}$$

In the exercises you are asked to show that the construction of Definition 5.40 satisfies the property of Definition 5.38.

In Definition 5.40 we have the quotient of a huge vector space by a huge subspace. It turns out that this quotient is finite-dimensional.

Lemma 5.41. *If* $\{e_1, \ldots, e_m\}$ *is a basis of* V, *and* $\{f_1, \ldots, f_n\}$ *is a basis of* W, *then the equivalence classes of pairs* $\{(e_i, f_j) \mid i = 1, \ldots, m, \ j = 1, \ldots, n\}$ *form a basis of* $V \otimes W$. *In particular,* $\dim(V \otimes W) = \dim V \cdot \dim W$.

Proof. Indeed, any vector $(v, w) = \left(\sum v^i e_i, \sum w^j f_j\right)$ is equivalent modulo Ω to $\sum v^i w^j (e_i, f_j)$, and the same holds for any linear combination $\sum \lambda_k (v_k, w_k)$. $\square$

Our last definition of the tensor product is as follows.

Definition 5.42. *The* tensor product $V \otimes W$ *is the space of all bilinear maps from* $V^* \times W^*$ *to* $\mathbb{R}$, *equipped with the natural linear structure:* $(h_1 + h_2)(v, w) = h_1(v, w) + h_2(v, w)$ *and* $(\lambda h)(v, w) = \lambda h(v, w)$.

To establish a link with the previous definition, associate with an element $v \otimes w \in V \times W$ the map

$$V^* \times W^* \to \mathbb{R}, \quad (\alpha, \beta) \mapsto \langle v, \alpha \rangle \langle w, \beta \rangle,$$

where $\langle \cdot, \cdot \rangle$ denotes the canonical pairing between a vector space and its dual. It is easy to check that the above map is bilinear. Similarly one associates a map $V^* \times W^* \to \mathbb{R}$ to any formal linear combination $\sum \lambda_k v_k \otimes w_k$ and finds out that $(v_1 + v_2) \otimes w$ and $v_1 \otimes w + v_2 \otimes w$ result in the same map etc.

If one applies Definition 5.42 to the spaces V^* and W^*, then one gets that $V^* \otimes W^*$ is the space of bilinear maps from $V \times W$ to $\mathbb{R}$.

Lemma 5.43. *There is a canonical isomorphism* $(V \otimes W)^* \cong V^* \otimes W^*$.

A canonical isomorphism is one which is independent of a choice of a basis. For example, V and V^* are isomorphic, but not canonically.

Proof. Elements of $(V \otimes W)^*$ are linear maps $V \otimes W \to \mathbb{R}$, and elements of $V^* \otimes W^*$ are bilinear maps $V \times W \to \mathbb{R}$ (by Definition 5.42). By Definition 5.38 this is the same. □

Lemma 5.44. *There is a canonical isomorphism*

$$V^* \otimes W \cong \mathrm{Hom}(V, W).$$

Proof. By Definition 5.42, elements of $V^* \otimes W$ are bilinear maps $V \times W^* \to \mathbb{R}$. To any such map h and any vector $v \in V$ associate an element $\overline{h}(v) \in W$ by the rule

$$\langle \overline{h}(v), \beta \rangle := h(v, \beta).$$

This defines $\overline{h}(v)$ correctly and uniquely, because a vector in W is determined uniquely by its linear action on the dual space W^*, and the action of $\overline{h}(v)$ defined above is linear due to the linearity of h in its second argument. The linearity of h in its first argument ensures that the map $\overline{h} \colon V \to W$ is linear. □

It can be shown that the tensor product of vector spaces is associative:

$$(V \otimes W) \otimes X \cong V \otimes (W \otimes X),$$

where the isomorphism is canonical. The corresponding triple product $V \otimes W \otimes X$ is the space of all trilinear maps $V^* \times W^* \times X^* \to \mathbb{R}$.

5.8 Tensors over a vector space

Definition 5.45. *The space* $\underbrace{V \otimes \cdots \otimes V}_{p} \otimes \underbrace{V^* \otimes \cdots \otimes V^*}_{q}$ *is denoted by* $\mathcal{T}^p_q(V)$*, and its elements are called* tensors of type (p, q) *or* (p, q)*-tensors over* V.

Thus $(1, 0)$-tensors are vectors, $(0, 1)$-tensors are linear functionals (also called covectors), and $(0, q)$-tensors are q-linear forms, that is multilinear maps $\underbrace{V \times \cdots \times V}_{q} \to \mathbb{R}$.

Lemma 5.46. *The space of* $(1, 1)$*-tensors* $V \otimes V^*$ *is the space of linear endomorphisms of* V*. The space of* $(0, 2)$*-tensors* $V^* \otimes V^*$ *is the space of bilinear forms on* V.

Proof. Indeed, by Lemma 5.44 the space $V \otimes V^*$ is canonically isomorphic to $\mathrm{Hom}(V, V)$.

Further, by Lemma 5.43 the space $V^* \otimes V^*$ is the space of linear functionals on $V \otimes V$ which is the space of bilinear forms on V. □

Let us now describe the coordinate approach to tensors. First, we introduce/recall the Einstein summation rule. It is convenient to use lower indices for the elements of a basis $(e_1, \dots, e_n)$ and the upper indices for the coordinates of a vector. A vector then can be written as

$$\sum_{i=1}^{n} v^i e_i =: v^i e_i.$$

The Einstein rule is to sum over a pair of an upper and a lower index, if these indices are denoted by the same letter. This allows to omit the summation signs and makes the formulas shorter.

Here is an example of application of the Einstein summation rule. Let $(\eta^1, \dots, \eta^n)$ be the dual basis of V^*, that is $\langle \eta^i, e_j \rangle = \delta^i_j$. A covector can then be written as $\ell_i \eta^i$, and the canonical pairing between a vector and a covector yields

$$\langle \ell_i \eta^i, v^j e_j \rangle = \ell_i v^j \langle \eta^i, e_j \rangle = \ell_i v^j \delta^i_j = \ell_i v^i.$$

The tensors

$$e_{i_1} \otimes \dots \otimes e_{i_p} \otimes \eta^{j_1} \otimes \dots \otimes \eta^{j_q}$$

form a basis of the tensor space $\mathcal{T}^p_q(V)$. Thus every (p,q)-tensor can be written as

$$T = T^{i_1 \dots i_p}_{j_1 \dots j_q} e_{i_1} \otimes \dots \otimes e_{i_p} \otimes \eta^{j_1} \otimes \dots \otimes \eta^{j_q} = T^I_J e_I \otimes \eta^J$$

(in the last formula we use multiindices).

Let $(f_1, \dots, f_n)$ be a new basis of V, written in terms of the old basis as

$$f_i = e_j c^j_i. \tag{30}$$

In the matrix notation this is

$$\begin{pmatrix} f_1 & \cdots & f_n \end{pmatrix} = \begin{pmatrix} e_1 & \cdots & e_n \end{pmatrix} \begin{pmatrix} c^1_1 & \cdots & c^1_n \\ \vdots & \ddots & \vdots \\ c^n_1 & \cdots & c^n_n \end{pmatrix}$$

Let $(\varphi^1, \dots, \varphi^n)$ be the basis of V^* dual to $(f_1, \dots, f_n)$. Then one has

$$\langle \eta^j, f_i \rangle = \langle \eta^j, e_k c^k_i \rangle = \delta^j_k c^k_i = c^j_i = \langle c^j_l \varphi^l, f_i \rangle,$$

which implies that the new and old dual bases are related by

$$\eta^j = c^j_i \varphi^i. \tag{31}$$

Now let $(v^1, \dots, v^n)$ and $(w^1, \dots, w^n)$ be the coordinates of the same vector in the old and the new bases respectively. Then one has

$$e_j v^j = f_i w^i = (e_j c^j_i) w^i = e_j (c^j_i w^i),$$

which implies

$$v^j = c_i^j w^i.$$

One sees that the coordinates of a vector are transformed by the same rule (31) as the elements of the dual basis. A similar argument shows that the coordinates of a covector transform by the rule (30) for the elements of the basis of V. The rule (30) is called the *covariant transformation law* (*as* the basis), while (31) is called the *contravariant transformation law* (*contrarily* to the basis).

If one wishes to rewrite (30) so that it expresses the old in terms of the new (as (31) does), then one needs to introduce the inverse matrix d_j^i characterized by

$$d_j^i c_k^j = \delta_k^i = c_j^i d_k^j.$$

Then, multiplying (30) from the right with d_k^i one obtains

$$e_j = f_i d_j^i.$$

Now let us see how do the components of tensors transform. Take an $(1,1)$-tensor represented in the old an in the new bases as

$$a_j^i e_i \otimes \eta^j = b_j^i f_i \otimes \varphi^j.$$

Using the formulas that express the new bases in terms of the old we compute

$$a_j^i e_i \otimes \eta^j = a_j^i \left(f_k d_i^k \otimes c_l^j \varphi^l \right) = d_i^k a_j^i c_l^j f_k \otimes \varphi^l.$$

Renaming the indices and comparing the coefficients at the basis tensors $f_i \otimes \varphi^j$ one concludes

$$b_j^i = d_k^i a_l^k c_j^l,$$

which corresponds to the rule $B = C^{-1} A C$ for the transformation of the matrix of a linear endomorphism.

For a general tensor a similar computation yields the transformation rule

$$b_{j_1 \ldots j_q}^{i_1 \ldots i_p} = d_{k_1}^{i_1} \cdots d_{k_p}^{i_p} a_{l_1 \ldots l_q}^{k_1 \ldots k_p} c_{j_1}^{l_1} \cdots c_{j_q}^{l_q}.$$

5.9 Exterior algebra

Recall that the space

$$\mathcal{T}_q^0(V) = \underbrace{V^* \otimes \cdots \otimes V^*}_{q}$$

can be viewed as the space of all multilinear maps

$$\alpha \colon \underbrace{V \times \cdots \times V}_{q} \to \mathbb{R},$$

that is the space of multilinear forms on V.

Definition 5.47. *A multilinear form $\alpha \in \mathcal{T}_q^0(V)$ is called* skew-symmetric *if for every permutation σ of $\{1, \dots, q\}$ one has*

$$\alpha(x_{\sigma(1)}, \dots, x_{\sigma(q)}) = \operatorname{sgn}(\sigma)\alpha(x_1, \dots, x_q),$$

where $\operatorname{sgn}(\sigma) \in \{-1, 1\}$ *denotes the sign of the permutation σ.*

Example 5.48. A bilinear form α is skew-symmetric if and only if $\alpha(v, w) = -\alpha(w, v)$ for all v, w.

Lemma 5.49. *Let α be a q-linear form. Then the following are equivalent.*

1. *α is skew-symmetric.*
2. *α is alternating, i. e. $\alpha(v_1, \dots, v_q) = 0$ whenever two of v_i are equal.*
3. *$\alpha(v_1, \dots, v_q) = 0$ whenever $v_1, \dots, v_q$ are linearly dependent.*

Proof. A skew-symmetric form is alternating because permuting two equal arguments should change the value and at the same time cannot change the value of the form. An alternating bilinear form is skew-symmetric because

$$\alpha(v, w) + \alpha(w, v) = \alpha(v + w, v + w) - \alpha(v, v) - \alpha(w, w) = 0.$$

For multilinear forms the argument is similar.

Property 3 implies property 2 because a collection of vectors with two equal entries is linearly dependent. Let us show that 2 implies 3. If $v_1, \dots, v_q$ are linearly dependent, then without loss of generality $v_1 = \sum_{i=2}^q \lambda_i v_i$, and therefore

$$\alpha(v_1, \dots, v_q) = \sum_{i=2}^q \lambda_i \alpha(v_i, v_2, \dots, v_i, \dots, v_q) = 0.$$

□

We will use the words skew-symmetric and alternating as synonyms.

Remark 5.50. When the ground field has characteristic 2, skew-symmetry is equivalent to symmetry and does not imply the alternating property.

The set of all alternating q-linear forms is a linear subspace of the space of q-linear forms and is denoted by

$$\bigwedge^q V^* \subset \bigotimes^q V^*.$$

Lemma 5.51. *Let $\alpha \in \bigwedge^q V^*$ and let $(e_1, \dots, e_n)$ be a basis of V. Then one has*

$$\alpha(v_1, \dots, v_q) = \sum_{1 \le i_1 < \dots < i_q \le n} \begin{vmatrix} v_1^{i_1} & \dots & v_1^{i_q} \\ \vdots & \ddots & \vdots \\ v_q^{i_1} & \dots & v_q^{i_q} \end{vmatrix} \alpha(e_{i_1}, \dots, e_{i_q}).$$

Proof. By the multilinearity one has

$$\alpha(v_1, \ldots, v_q) = \sum_{i_1,\ldots,i_q} v_1^{i_1} \cdots v_q^{i_q} \alpha(e_{i_1}, \ldots, e_{i_q}).$$

Since α is alternating, the terms where some two of the indices i_r are equal vanish. Due to the skew-symmetry the terms $\alpha(e_{i_1}, \ldots, e_{i_q})$ with the same set of indices can be grouped together. This makes the determinant coefficients to appear. $\square$

Lemma 5.52. *One has*

$$\dim \bigwedge^q V^* = \binom{n}{q}.$$

In particular, $\dim \bigwedge^n V^* = 1$ *and* $\dim \bigwedge^r V^* = 0$ *for* $r > n$.

Proof. Indeed, by Lemma 5.51 an alternating q-form α is uniquely determined by its values on $(e_{i_1}, \ldots, e_{i_q})$. On the other hand, if one chooses arbitrary numbers as these values, then the formula from Lemma 5.51 defines an alternating q-form: permuting the arguments permutes the rows in the determinants. $\square$

The tensor product of two alternating forms is not alternating. Our next goal is to define a new product operation that respects the alternating property.

Definition 5.53. *Let α be any q-linear form. The* skew-symmetrization *of α, denoted* $\operatorname{Alt}\alpha$*, is given by*

$$(\operatorname{Alt}\alpha)(v_1, \ldots, v_q) = \frac{1}{q!} \sum_{\sigma \in S_q} \operatorname{sgn}(\sigma) \alpha(v_{\sigma(1)}, \ldots, v_{\sigma(q)}).$$

It is easy to check that $\operatorname{Alt}\alpha$ is skew-symmetric, and that if α is skew-symmetric, then $\operatorname{Alt}\alpha = \alpha$. The skew-symmetrization is thus a linear projection $\bigotimes^q V^* \to \bigwedge^q V^*$.

Definition 5.54. *The* exterior product *or the* wedge product *of $\alpha \in \bigwedge^q V^*$ and $\beta \in \bigwedge^r V^*$, denoted $\alpha \wedge \beta$, is defined as*

$$\alpha \wedge \beta = \frac{(q+r)!}{q!r!} \operatorname{Alt}(\alpha \otimes \beta).$$

Example 5.55. For $q = r = 1$ on has $(\alpha \wedge \beta)(v, w) = \alpha(v)\beta(w) - \alpha(w)\beta(v)$.

Lemma 5.56. *The exterior product is anticommutative:*

$$\beta \wedge \alpha = (-1)^{qr} \alpha \wedge \beta \quad \text{if} \quad \alpha \in \bigwedge^q V^* \text{ and } \beta \in \bigwedge^r V^*$$

and associative:

$$(\alpha \wedge \beta) \wedge \gamma = \alpha \wedge (\beta \wedge \gamma).$$

The proof consists in juggling with permutations. In particular one arrives at

$$\alpha \wedge \beta \wedge \gamma = \frac{(q+r+s)!}{q!r!s!} \operatorname{Alt}(\alpha \otimes \beta \otimes \gamma)$$

and at a similar formula for any bigger number of factors. For the exterior product of 1-forms this implies

$$(\alpha^1 \wedge \cdots \wedge \alpha^q)(v_1, \ldots, v_q) = \begin{vmatrix} \alpha^1(v_1) & \cdots & \alpha^q(v_1) \\ \vdots & \ddots & \vdots \\ \alpha^1(v_q) & \cdots & \alpha^q(v_q) \end{vmatrix},$$

which allows us to rewrite the formula of Lemma 5.51 as

$$\alpha(v_1, \ldots, v_q) = \sum_{1 \le i_1 < \cdots < i_q \le n} \alpha(e_{i_1}, \ldots, e_{i_q}) \eta^{i_1} \wedge \cdots \wedge \eta^{i_q},$$

where $(e_1, \ldots, e_n)$ and $(\eta_1, \ldots, \eta_n)$ is any dual pair of bases of V and V^*.

5.10 Tensor fields

A tensor field of type (r, s) on a smooth manifold M is a choice of a (r, s)-tensor in each tangent space, depending smoothly on the point. The smooth dependence on the point can be defined in terms of charts: the differential of a chart $\varphi \colon U \to \mathbb{R}^n$ identifies all tangent spaces T_pM, $p \in U$ with $\mathbb{R}^n$ and pushes the tensors from T_pM to $\mathbb{R}^n$.

More exactly, recall that a chart defines a basis $\frac{\partial}{\partial x^i}$ in each of the tangent spaces T_pM, $p \in U$. The dual basis of T_p^*M is formed by the differentials of the coordinate functions dx^i: it can be shown that

$$\left\langle dx^i, \frac{\partial}{\partial x^j} \right\rangle = \delta^i_j.$$

Thus an (r, s)-tensor field can be written as

$$a^{i_1 \ldots i_r}_{j_1 \ldots j_s} \frac{\partial}{\partial x^{i_1}} \otimes \cdots \otimes \frac{\partial}{\partial x^{i_r}} \otimes dx^{j_1} \otimes \cdots \otimes dx^{j_s},$$

(Einstein summation convention!), and the smooth dependence on the point means that $a^{i_1 \ldots i_r}_{j_1 \ldots j_s} \in C^\infty(M)$.

Let $y = y(x)$ be a coordinate change and let $b^{i_1 \ldots i_r}_{j_1 \ldots j_s}$ be the components of the same tensor field in the new coordinates. Then one has

$$b^{k_1 \ldots k_r}_{l_1 \ldots l_s}(y) = a^{i_1 \ldots i_r}_{j_1 \ldots j_s}(x) \frac{\partial y^{k_1}}{\partial x^{i_1}} \cdots \frac{\partial y^{k_r}}{\partial x^{i_r}} \frac{\partial x^{j_1}}{\partial y^{l_1}} \cdots \frac{\partial x^{j_s}}{\partial y^{l_s}}. \tag{32}$$

This follows from the fact that the basis change is governed by the Jacobian matrix $\frac{\partial x}{\partial y}$:

$$\frac{\partial}{\partial y^i} = \frac{\partial}{\partial x^j} \frac{\partial x^j}{\partial y^i}.$$

A tensor over V can be defined as a map associating to every basis of V an array of components that behaves under basis change in a certain way, see end of Section 5.7. Similarly, a tensor field on M is a map associating to every chart an array of functions that change under transition maps according to (32). For example, formula that we derived for the commutator

$$[X,Y] = \left(X^j \frac{\partial Y^i}{\partial x^j} - Y^j \frac{\partial X^i}{\partial x^j}\right) \frac{\partial}{\partial x^i}$$

associates to any two vector fields X and Y a well-defined vector field because applying first the coordinate change to X and Y and then the formula yields the same as first applying the formula and then the coordinate change:

$$\left(\left(X^k \frac{\partial y^j}{\partial x^k}\right) \frac{\partial}{\partial y^j} \left(Y^l \frac{\partial y^i}{\partial x^l}\right) - \left(Y^k \frac{\partial y^j}{\partial x^k}\right) \frac{\partial}{\partial y^j} \left(X^l \frac{\partial y^i}{\partial x^l}\right)\right) \frac{\partial}{\partial y^i} = \left(X^j \frac{\partial Y^k}{\partial x^j} - Y^j \frac{\partial X^k}{\partial x^j}\right) \frac{\partial y^j}{\partial x^k} \frac{\partial}{\partial y^j}.$$

Denote by $\mathcal{T}M$ the space of all vector fields and by $\mathcal{T}^*M$ the space of all covector fields on M. A covector field can be pointwise paired with a vector field: $\langle \omega, X\rangle_p = \langle \omega_p, X_p\rangle$, or, in a different notation $\omega(X)_p = \omega_p(X_p)$. This makes ω to a $C^\infty(M)$-linear map

$$\mathcal{T}M \to C^\infty(M).$$

It can be shown that conversely, every $C^\infty(M)$-linear map from the vector fields to the space of functions corresponds to a covector field. More generally, the following holds.

Fact 5.57. *Every $C^\infty(M)$-multilinear map*

$$\underbrace{\mathcal{T}^*M \times \cdots \times \mathcal{T}^*M}_{r} \times \underbrace{\mathcal{T}M \times \cdots \times \mathcal{T}M}_{s} \to C^\infty(M)$$

corresponds to some (r,s)-tensor field.

For a proof (in a special case) see [Lee13, Lemma 12.24].

For two tensor fields α and β one defines their tensor product pointwise:

$$(\alpha \otimes \beta)_p := \alpha_p \otimes \beta_p.$$

(However, one can also define the tensor product of $C^\infty(M)$-modules and consider $\alpha \otimes_{C^\infty(M)} \beta$, which will be the same thing.)

5.11 Differential forms and exterior derivative

Lecture 13

Definition 5.58. *A* differential q-form *on M is a field of alternating q-linear forms in the tangent spaces of M. The set of all differential q-forms on M is denoted by $\Omega^q(M)$.*

In particular, $\Omega^1(M) = \mathcal{T}^*M$. For higher q, the set $\Omega^q(M)$ is a proper linear subspace of the space of $(0,q)$-tensor fields on M. One also puts $\Omega^0(M) = C^\infty(M)$.

Combining Fact 5.57 with the definition of alternating forms one can prove that $\Omega^q(M)$ is the set of all alternating $C^\infty(M)$-linear maps

$$\underbrace{\mathcal{T}M \times \cdots \times \mathcal{T}M}_{q} \to C^\infty(M).$$

The pointwise exterior product gives rise to the *exterior product* of differential forms

$$\Omega^q(M) \times \Omega^r(M) \to \Omega^{q+r}(M).$$

For consistency we sometimes write

$$f \wedge \omega = f\omega$$

for $f \in C^\infty(M)$, $\omega \in \Omega^r(M)$.

Up to now we dealt only with algebraic operations on tensors and corresponding pointwise operations with tensor fields. The time to differentiate has come.

Lemma-Definition 5.59. *For every $q \geq 0$ there is a unique $\mathbb{R}$-linear map*

$$d\colon \Omega^q(M) \to \Omega^{q+1}(M)$$

satisfying the following properties.

1. $d\colon \Omega^0(M) \to \Omega^1(M)$ *is the differential of a function.*
2. $d(\omega_1 \wedge \omega_2) = d\omega_1 \wedge \omega_2 + (-1)^{\deg \omega_1} \omega_1 \wedge d\omega_2$.
3. $d \circ d = 0$.

This map is called the exterior derivative*.*

Proof. We prove the uniqueness by deriving from the properties 1–3 a formula for $d\omega$ in local coordinates.

First, in order to be able to use local coordinates one shows that the value of $d\omega$ at p depends only on the values of ω in an arbitrarily small neighborhood of p. For this, multiply ω with a "bump function" f and use property 2 for the product $f \wedge \omega$. (We have used this trick when we discussed derivations and tangent vectors, see the proof of Theorem 5.27.)

Let (U, φ) be a chart with $p \in U$. A differential form η on $\varphi(U) \subset \mathbb{R}^n$ can be "pulled back" to a form $\varphi^*\eta$ on U with the help of the differential of φ. A form can be described by its action on vectors, and we describe $\varphi^*\eta$ as follows:

$$\varphi^*\eta(X_1, \ldots X_q) := \eta(d\varphi(X_1), \ldots, d\varphi(X_q)).$$

Since φ is a diffeomorphism, its inverse can be used to pull back differential forms from U to $\varphi(U)$. The pullback correspondence between differential forms respects the exterior product and satisfies

$$\varphi^*(df) = d(f \circ \varphi).$$

Therefore it suffices to show that the properties 1–3 determine a unique operation on differential forms on $\mathbb{R}^n$.

Let $(x^1, \ldots, x^n)$ be coordinates in $\mathbb{R}^n$. In particular, each x^i is a smooth function. One has $d(dx^i) = 0$ by property 3. From property 2 by induction it follows that

$$d(dx^{i_1} \wedge \cdots \wedge dx^{i_q}) = 0.$$

Together with property 1 this implies

$$d(f dx^{i_1} \wedge \cdots \wedge dx^{i_q}) = df \wedge dx^{i_1} \wedge \cdots \wedge dx^{i_q} = \sum_{j=1}^{n} \frac{\partial f}{\partial x^j} dx^j \wedge dx^{i_1} \wedge \cdots \wedge dx^{i_q}. \qquad (33)$$

Any differential q-form is the sum of terms $f dx^{i_1} \wedge \cdots \wedge dx^{i_q}$ (decomposable forms), so it remains to extend the above formula by additivity. Thus, if the exterior derivative exists, then it is given in any local coordinates on M by the above formula.

There are two more things to prove. First, one has to check that the operation on $\Omega^q(\mathbb{R}^n)$ given by (33) satisfies properties 2 and 3. This can be done by straightforward computations. Second, when we transfer the result back to M we want to be sure that the form $d\omega$ will be the same independently on the choice of the local coordinates. This means, one has to check that formula (33) behaves well under the coordinate change: if we first make a substitution $x = x(y)$ in ω and then compute $d\omega$, the result will be the same as when we first compute $d\omega$ and then substitute there $x = x(y)$. Compare Section 5.10. This computation is a bit lengthy. $\square$

One can establish the well-definedness of the exterior derivative by proving a coordinate-independent formula.

Lemma 5.60. *The differential form $d\omega$ defined with the help of the formula* (33) *acts on vectors as follows.*

$$\begin{aligned} d\omega(X_0, \ldots, X_q) = \sum_{i=0}^{q} (-1)^i D_{X_i}(\omega(X_0, \ldots, \widehat{X_i}, \ldots, X_q)) \\ + \sum_{0 \le i < j \le q} (-1)^{i+j} \omega([X_i, X_j], X_0, \ldots, \widehat{X_i}, \ldots, \widehat{X_j}, \ldots, X_q). \qquad (34) \end{aligned}$$

Here a hat over a symbol means that the corresponding entry is omitted.

Proof. We will give a proof for $q = 1$, in the general case the argument is similar. If ω is a 1-form, then the formula says

$$d\omega(X, Y) = D_X(\omega(Y)) - D_Y(\omega(X)) - \omega([X, Y]).$$

First, let us show that the right hand side defines a differential 2-form. By Fact 5.57 it suffices to prove that this expression (which we denote by RHS) is $C^\infty(M)$-bilinear.

The additivity is clear, the only tricky part is the $C^\infty(M)$-homogeneity:

$$\begin{aligned} RHS(fX, Y) &= D_{fX}(\omega(Y)) - D_Y(\omega(fX)) - \omega([fX, Y]) \\ &= fD_X(\omega(Y)) - D_Y(f\omega(X)) - \omega(-D_Y f \cdot X + f[X, Y]) \\ &= fD_X(\omega(Y)) - D_Y f \cdot \omega(X) - fD_Y(\omega(X)) + D_Y f \cdot \omega(X) - f\omega([X, Y]) \\ &= f \cdot (D_X(\omega(Y)) - D_Y(\omega(X)) - \omega([X, Y])) = fRHS(X, Y) \end{aligned}$$

The homogeneity in the second argument is proved similarly (or by using the skew-symmetry of RHS). Thus RHS is a differential 2-form. In order to show that it equals $d\omega$ from (33) it suffices to compare the values at the coordinate vector fields (as the $C^\infty(M)$-linearity will do the rest). A general differential 1-form is

$$\omega = \sum_{i=1}^{n} f_i dx^i,$$

and the formula (33) yields

$$d\omega = \sum_{i,j} \frac{\partial f_i}{\partial x^j} dx^j \wedge dx^i = \sum_{i<j} \left(\frac{\partial f_j}{\partial x^i} - \frac{\partial f_i}{\partial x^j} \right) dx^i \wedge dx^j.$$

Our new formula yields

$$RHS(\partial_i, \partial_j) = \partial_i(\omega(\partial_j)) - \partial_j(\omega(\partial_i)) - \omega([\partial_i, \partial_j]) = \frac{\partial f_j}{\partial x^i} - \frac{\partial f_i}{\partial x^j} = d\omega(\partial_i, \partial_j),$$

thus $RHS = d\omega$, and the lemma is proved. □

5.12* Closed and exact forms

Definition 5.61. *A differential form ω is called* closed *if $d\omega = 0$.*

A differential form $\omega \in \Omega^q(M)$ is called exact *if there is a differential form $\eta \in \Omega^{q-1}(M)$ such that $d\eta = \omega$.*

The property $d \circ d = 0$ of the exterior derivative implies that every exact form is closed. The following example shows that the converse is not true.

Example 5.62. The form

$$\omega = \frac{x\,dy - y\,dx}{x^2 + y^2} \in \Omega^1(\mathbb{R}^2 \setminus \{(0,0)\})$$

is closed:

$$d\omega = d\left(\frac{x}{x^2+y^2}\right) dy - d\left(\frac{y}{x^2+y^2}\right) dx = \frac{-x^2+y^2}{(x^2+y^2)^2} dx \wedge dy - \frac{x^2-y^2}{(x^2+y^2)^2} dy \wedge dx = 0.$$

(In this computation we have used that $dx \wedge dx = 0$ and $dy \wedge dx = -dx \wedge dy$.)

In order to show that this form is not exact, let us change to the polar coordinates. For $\varphi(r,\theta) = (r\cos\theta, r\sin\theta)$ we compute

$$\varphi^*(x\,dy - y\,dx) = r\cos\theta(\sin\theta\,dr + r\cos\theta\,d\theta) - r\sin\theta(\cos\theta\,dr + r\cos\theta\,d\theta) = r^2\,d\theta,$$

so that $\omega = d\theta$ (which by the way makes its closeness more apparent). It follows that if $\omega = df$, then $f = \theta + \text{const}$. But θ is a local coordinate which does not extend to a function defined on $\mathbb{R}^2 \setminus \{(0,0\}$.

However, a closed form on a topologically simple subset of a manifold is exact, as the next theorem claims.

Theorem 5.63 (Poincaré lemma)**.** *Let $U \subset M$ be an open subset diffeomorphic to an open ball. Then every closed form $\omega \in \Omega^q(U)$ is exact.*

Lemma 5.64. *If $\omega = \sum_{i=1}^n f_i dx^i$ is a closed differential 1-form on $\mathbb{R}^n$, then the function*

$$h(x) = \sum_{i=1}^n x^i \int_0^1 f_i(tx)\,dt$$

has ω as its differential.

Proof. For $n = 1$ every form $f\,dx$ is closed, and the function h is found with a simple integration:

$$h(x) = \int_0^y f(y)\,dy = x\int_0^1 f(tx)\,dt.$$

For an arbitrary n, the formula for $h(x)$ from the lemma is the integral of ω along a straight line segment from 0 to x. The closedness condition for ω is

$$\frac{\partial f_i}{\partial x^j} = \frac{\partial f_j}{\partial x^i},$$

and it will be used in our computations. Using the Leibniz rule and differentiating under the integral sign one obtains

$$dh = \sum_{i=1}^n \left(\int_0^1 f_i(tx)\,dt\right) dx^i + \sum_{i=1}^n \left(x^i \sum_{j=1}^n \left(\int_0^1 t\frac{\partial}{\partial x^j} f_i(tx)\,dt\right) dx^j\right).$$

With the help of the closedness condition, the chain rule and the integration by parts the coefficient at dx^j in the second term transforms to

$$\begin{aligned}\sum_{i=1}^n x^i \int_0^1 t\frac{\partial}{\partial x^j} f_i(tx)\,dt &= \sum_{i=1}^n x^i \int_0^1 t\frac{\partial}{\partial x^i} f_j(tx)\,dt \\ &= \int_0^1 t\frac{d}{dt} f_j(tx)\,dt = f_j(x) - \int_0^1 f_j(tx)\,dt.\end{aligned}$$

This implies $dh = \omega$, and the lemma is proved. □

Proof of the Poincaré lemma for 1-forms. Since U is diffeomorphic to an open ball, it is also diffeomorphic to $\mathbb{R}^n$. Let $\varphi\colon U \to \mathbb{R}^n$ be a diffeomorphism. Consider the pullback of ω by φ^{-1}. Since pullbacks commute with the differentials, one has

$$d((\varphi^{-1})^*\omega) = (\varphi^{-1})^* d\omega = 0.$$

By the previous lemma, there is a function $h \in C^\infty(\mathbb{R}^n)$ such that $dh = (\varphi^{-1})^*\omega$. Therefore

$$d(h\colon \varphi) = d(\varphi^* h) = \varphi^* dh = \omega,$$

and the Poincaré lemma is proved. □

The proof of Poincaré lemma for $q > 1$ requires tools that we do not have time to introduce. An interested reader can learn it from [Laf15].

The Poincaré lemma for 1-forms will help us to prove Theorem 5.65: any n pairwise commuting linearly independent vector fields on a contractible open subset of an n-dimensional manifold are the coordinate fields of some chart.

Let us prove a converse to Lemma 5.35.

Theorem 5.65. *Let $U \subset M$ be an open subset homeomorphic to the n-dimensional ball, and let $X_1, \ldots, X_n$ be vector fields on U such that*

- *for every $p \in U$ the vectors $X_1(p), \ldots, X_n(p) \in T_pM$ are linearly independent;*
- *the pairwise commutators of these vector fields vanish:*

$$[X_i, X_j] = 0 \text{ over all of } U \text{ for all } i, j.$$

Then there is a local diffeomorphism $\varphi\colon U \to \mathbb{R}^n$ such that $X_i = \frac{\partial}{\partial x^i}$ are the corresponding coordinate vector fields.

Note that φ is only a local diffeomorphism, it might be non-injective; see Example 5.36. But coordinate vector fields make sense for a local diffeomorphism as well.

Proof of Theorem 5.65. Let $(\omega^1, \ldots, \omega^n)$ be the dual coframe to $(X_1, \ldots, X_n)$, that is for every $p \in U$ the linear forms $(\omega^1(p), \ldots, \omega^n(p))$ form a basis of T_p^*M dual to the basis $(X_1(p), \ldots, X_n(p))$ of T_pM. Due to the pairwise commutation of the fields X_i one has

$$\begin{aligned} d\omega^i(X_j, X_k) &= D_{X_j}(\omega^i(X_k)) - D_{X_k}(\omega^i(X_j)) - \omega^i([X_j, X_k]) \\ &= D_{X_j}(\delta^i_k) - D_{X_k}(\delta^i_j) - \omega^i(0) = 0, \end{aligned}$$

which implies $d\omega^i = 0$. By Poincaré lemma, the forms ω^i are exact. Let $\varphi^1, \ldots, \varphi^n \in C^\infty(U)$ be such that $d\varphi^i = \omega^i$. Since for every $p \in U$ the functionals $d\varphi^1(p), \ldots, d\varphi^n(p) \in T_p^*(M)$ are linearly independent, the map $\varphi\colon U \to \mathbb{R}^n, p \mapsto (\varphi^1(p), \ldots, \varphi^n(p))$ has a non-degenerate Jacobi matrix and is therefore a local diffeomorphism. The coordinate vector fields of φ form the basis dual to $(d\varphi^1, \ldots, d\varphi^n)$, and therefore coincide with the vector fields $X_1, \ldots, X_n$. □

5.13* De Rham cohomology

This is an outlook section without any proofs.

The exterior derivative connects the spaces of differential forms on an n-dimensional manifold to a chain

$$0 \xrightarrow{d} C^\infty(M) = \Omega^0(M) \xrightarrow{d} \Omega^1(M) \xrightarrow{d} \cdots \xrightarrow{d} \Omega^n(M) \xrightarrow{d} \Omega^{n+1}(M) = 0.$$

Each of $\Omega^q(M)$ is a real vector space, and d is a linear map. The property $d \circ d = 0$ implies that $\operatorname{im}(d\colon \Omega^{q-1}(M) \to \Omega^q(M)) \subset \ker(d\colon \Omega^q(M) \to \Omega^{q+1}(M))$. Thus one can consider the quotient vector space

$$H^q(M) := \frac{\ker(d\colon \Omega^q(M) \to \Omega^{q+1}(M))}{\operatorname{im}(d\colon \Omega^{q-1}(M) \to \Omega^q(M))},$$

called the q-th *de Rham cohomology group* of M. (The name group seems inappropriate, it is explained by the fact that there are other cohomology theories, where this group is in general a non-free abelian group.) For a compact manifold M the vector spaces $H^q(M)$ are always finite-dimensional, and the dimension

$$\dim H^q(M) =: b_q(M)$$

is called the q-th *Betti number.*

Let us look at the de Rham cohomology of orientable surfaces. Denote by S_g the sphere with g handles. One has

$$b_0(S_g) = b_2(S_g) = 1.$$

The $H^0(S_g)$ is easy to understand, the $H^2(S_g)$ is more complicated. Further, one has

$$b_1(S_g) = 2g.$$

In particular, the first cohomology group of the torus is $\mathbb{R}^2$: there are two independent non-exact closed forms. You can construct these forms following the idea of Example 5.62. For surfaces of higher genus the generators can be constructed using the pullback: wrap one hand around the torus and collapse everything else to a point, then pull back the generators of the cohomology of the torus.

Observe that

$$b_0 - b_1 + b_2 = 2 - 2g = \chi(M),$$

the Euler characteristic of M, see end of Section 4.11. A similar formula holds for any compact manifold of any dimension n:

$$b_0 - b_1 + \cdots + (-1)^n b_n = \chi(M).$$

(The combinatorial definition of Euler characteristic: cut a manifold into polyhedra, and compute the alternating sum of the number of faces of each dimension.)

In dimension 2, the Betti numbers allow to distinguish between different, that is non-diffeomorphic, manifolds. This is no more the case in higher dimensions. For example, there is a way to glue the opposite faces of the dodecahedron to obtain a 3-dimensional manifold, called *Poincaré homology sphere*, which is not homeomorphic to the sphere $\mathbb{S}^3$, but has the same cohomology groups: $b_0 = b_3 = 1$, $b_1 = b_2 = 0$.

The Poincaré homology sphere can be distinguished from the standard sphere with the help of another topological invariant, the fundamental group. This example led Poincaré to state his famous conjecture: every 3-manifold with a trivial fundamental group is homeomorphic to $\mathbb{S}^3$.

6 Riemannian geometry

In the previous section we have studied smooth manifolds, which are the playground for Riemannian geometry. The theory of smooth manifolds is very rich; we touched upon such questions as embeddings of manifolds, vector fields, differential forms, and de Rham cohomology. The Riemannian geometry starts when we introduce a Riemannian metric on a smooth manifold. We have intentionally postponed this step in order to show how much is possible without a metric.

6.1 Riemannian metric

In the introduction to Section 5 we have hinted at what a Riemannian metric is. Thanks to the preparation done in that section, it is now very easy to give a definition.

Definition 6.1. *A* Riemannian metric *on a smooth manifold M is a smooth $(0,2)$-tensor field g such that at each point p the bilinear form g_p on T_pM is symmetric and positive definite.*

A Riemannian manifold (M, g) *is a smooth manifold equipped with a Riemannian metric.*

In other words, a Riemannian metric is a smooth family of inner products on the tangent spaces to the manifold.

A Riemannian metric allows to compute lengths of smooth curves: for a curve $\gamma \colon [a, b] \to M$ one puts

$$L(\gamma) = \int_a^b \sqrt{g(\dot\gamma, \dot\gamma)}\, dt.$$

It is also possible to define the volume element and compute volumes, similarly to Lemma 2.23.

A Riemannian metric also establishes a bijection between vectors and covectors. We will explain this now from a general point of view. Every bilinear form on V is an element of $V^* \otimes V^*$. By Lemma 5.44, there is a map

$$V^* \otimes V^* \to \mathrm{Hom}(V, V^*), \quad h \mapsto \overline{h},$$

with

$$\overline{h}\colon V \to V^*, \quad \langle \overline{h}(v), w\rangle = h(v,w).$$

We would like to understand when the map $\overline{h}$ is an isomorphism.

Definition 6.2. *A bilinear form $h\colon V \times V \to \mathbb{R}$ is called* non-degenerate *if for every non-zero $v \in V$ there is $w \in V$ such that $h(v,w) \neq 0$.*

Lemma 6.3. *A bilinear form h is non-degenerate if and only if the map $\overline{h}$ is an isomorphism.*

Proof. If h is non-degenerate, then for every $v \neq 0$ one has $\overline{h}(v) \neq 0$ because this linear functional does not vanish on the vector w for which $h(v,w) \neq 0$.

If $\overline{h}$ is an isomorphism, then it has a trivial kernel because $\dim V = \dim V^*$. Thus for every $v \neq 0$ one has $\overline{h}(v) \neq 0$, that is there is $w \in V$ such that $h(v,w) = \langle \overline{h}(v), w\rangle \neq 0$, and we are good. ☐

A positive definite symmetric bilinear form g is non-degenerate because $g(v,v) > 0$ for all $v \neq 0$. Thus for every $p \in M$ we get an isomorphism from T_pM to T_p^*M. Applied pointwise, it maps vector fields to differential 1-forms. There is also an inverse map which sends differential 1-forms to vector fields.

Definition 6.4. *The* musical isomorphisms *on a Riemannian manifold (M,g) are defined as follows. For every vector field X the corresponding 1-form $X^\flat$ is defined as*

$$\langle X^\flat, Y\rangle = g(X,Y),$$

where the equality must hold at every point $p \in M$. For every differential 1-form ω the corresponding vector field $\omega^\sharp$ is defined as

$$g(\omega^\sharp, X) = \langle \omega, X\rangle$$

(again with a pointwise equality).

Remark 6.5. For every bilinear form h there is a second linear map

$$\underline{h}\colon V \to V^*, \quad \langle \underline{h}(v), w\rangle = h(w,v).$$

One has $\overline{h} = \underline{h}$ if and only if h is symmetric.

6.2* Existence of Riemannian metrics and isometric embeddings

A surface in $\mathbb{R}^3$ has a natural Riemannian metric: the Euclidean inner product restricts to an inner product on the tangent planes. The same holds for any submanifold of $\mathbb{R}^n$.

On every smooth manifold one can introduce a Riemannian metric, and this in a huge variety of ways.

Theorem 6.6. *Every smooth manifold carries a Riemannian metric.*

First proof. Take a smooth embedding of the manifold into a Euclidean space and consider the induced metric. $\square$

Second proof. Take a chart $\varphi\colon U \to \mathbb{R}^n$ and pull back the Euclidean inner product from $\mathbb{R}^n$ to U. This defines a Riemannian metric on U. One can show that every manifold has a locally finite atlas (U_i, φ_i), that is such that every point belongs to finitely many U_i. Then one can construct a family of functions $\mu_i \in C^\infty(M)$ such that

- $\mu_i \geq 0$;
- μ_i vanishes outside of U_i;
- $\sum_i \mu_i > 0$ everywhere.

Let g_i be a Riemannian metric on U_i. Then $\sum_i \mu_i g_i$ is a field of positive definite symmetric bilinear forms. $\square$

A natural question arises: does every Riemannian metric on M come from some embedding of M into a Euclidean space $\mathbb{R}^N$, and if yes, how large should the dimension N compared to the dimension n of M must be. In other words, can every Riemannian manifold be isometrically embedded into a Euclidean space? This is a very difficult question, and it has different answers.

Theorem 6.7 (Weyl-Nirenberg-Pogorelov)**.** *If g is a Riemannian metric on $\mathbb{S}^2$ such that the curvature of g is positive everywhere, then $(\mathbb{S}^2, g)$ can be isometrically embedded into $\mathbb{R}^3$.*

(By the curvature of g we mean its Gauss curvature which, as we know, is intrinsic, that is can be computed from g alone.)

Theorem 6.8 (Nash)**.** *An n-dimensional Riemannian manifold can be isometrically embedded into the Euclidean space of dimension $\frac{n(n+1)(3n+11)}{2}$. For a compact n-manifold, the dimension $\frac{n(3n+11)}{2}$ suffices.*

Later, Gromov and Rokhlin have shown that dimension $n^2 + 10n + 3$ suffices both in the compact and the non-compact case.

An embedding φ of a Riemannian manifold M into $\mathbb{R}^N$ is called *short* if for every curve γ in M the length of $\varphi \circ \gamma$ is smaller or equal to the length of γ.

Theorem 6.9 (Nash-Kuiper)**.** *Every short embedding of (M, g) into $\mathbb{R}^N$ can be approximated by isometric C^1-embeddings. In particular, an n-dimensional Riemannian manifold can be C^1-isometrically embedded into $\mathbb{R}^{2n}$.*

The latter follows from the Whitney embedding theorem.

Take for example, a flat metric on the torus, that is the metric obtained by isometrically identifying the opposite sides of a parallelogramm or, analytically, the metric $d\theta^2 + d\eta^2$, $\theta \in \mathbb{R}/a\mathbb{Z}$, $\eta \in \mathbb{R}/b\mathbb{Z}$. Since a torus can be embedded into $\mathbb{R}^3$, this flat torus has a C^1-isometric embedding into $\mathbb{R}^3$. This seems absolutely impossible. However, the

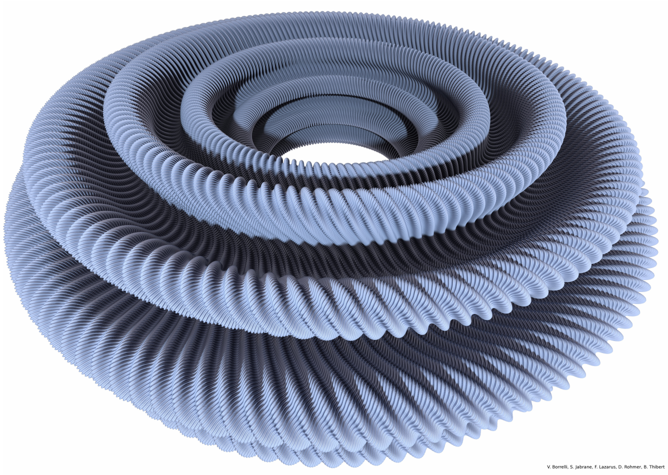

Figure 35: A flat torus C^1-isometrically embedded into $\mathbb{R}^3$.

proof is constructive, and there are computer-generated images of such an embedding [BJLT13], see Figure 35.

The picture exhibits a fractal structure of the embedding, but this is a smooth fractal: at every point there is a tangent plane. For a C^1-smooth surface one can define lengths of smooth curves and areas of domains, but one cannot define the curvature, because this requires second derivatives. It is impossible to embed a flat torus C^2-isometrically into $\mathbb{R}^3$, because there necessarily will be a point with a positive Gauss curvature.

6.3 Covariant derivative

Our next goal is to learn how to take the directional derivative (traditionally called the covariant derivative) of a vector field on a smooth manifold. For surfaces in $\mathbb{R}^3$ we have used for this the standard directional derivative in $\mathbb{R}^3$, see Section 4.8. We have shown then that this extrinsically defined operation is completely determined by the intrinsic metric of the surface (remember Christoffel symbols?).

Now we are in the intrinsic context from the very beginning. Our plan is as follows. We will define covariant derivatives, then we will show that on a smooth manifold there are many different covariant derivatives, and finally we will show that a Riemannian metric allows to choose one special covariant derivative from this set.

Definition 6.10. *A* covariant derivative *of vector fields is an operation that associates to two vector fields X and Y a new vector field $\nabla_X Y$ and satisfies the following properties.*

1. *$C^\infty(M)$-linearity with respect to X:*

$$\nabla_{X_1+X_2} Y = \nabla_{X_1} Y + \nabla_{X_2} Y, \quad \nabla_{fX} Y = f\, \nabla_X Y;$$

2. additivity and the Leibniz rule with respect to Y:

$$\nabla_X(Y_1 + Y_2) = \nabla_X Y_1 + \nabla_X Y_2, \quad \nabla_X(fY) = D_X f \cdot Y + f\, \nabla_X Y.$$

There is an equivalent definition.

Definition 6.11. *A* covariant derivative *of vector fields is a map $Y \mapsto \nabla Y$ from vector fields to $(1,1)$-tensor fields which is additive and satisfies the Leibniz rule:*

$$\nabla(fY) = df \otimes Y + f\, \nabla Y.$$

Instead of $(\nabla Y)(X)$ one writes $\nabla_X Y$.

The formula in Definition 6.11 is equivalent to the second formula in the second point of Definition 6.10. As for the rest, we have already encountered the equivalence between $C^\infty(M)$-linearity and tensoriality.

A covariant derivative on a smooth manifold is also called an *affine connection on the tangent bundle*. This name suggests that there are other sorts of connections and other sorts of bundles.

Lemma 6.12. *If ∇^1 and ∇^2 are two covariant derivatives on M, then $\lambda\nabla^1 + (1-\lambda)\nabla^2$ is also a covariant derivative on M.*

Proof. A direct computation. Just in case:

$$(\lambda\nabla^1 + (1-\lambda)\nabla^2)_X Y := \lambda \cdot \nabla^1_X Y + (1-\lambda)\nabla^2_X Y.$$

□

Theorem 6.13. *On every compact smooth manifold there is a covariant derivative.*

The set of all covariant derivatives on M has the following description. If ∇ and ∇' are two covariant derivatives on M, then the difference $\nabla - \nabla'$ is a $(1,2)$-tensor field. Conversely, for any covariant derivative ∇ on M and any $(1,2)$-tensor field Γ the sum $\nabla + \Gamma$ is also a covariant derivative.

Proof. On $\mathbb{R}^n$, there is a standard derivative $D_X Y$ which consists in taking directional derivatives of the components of the vector field Y. A chart $\varphi\colon U \to \mathbb{R}^n$ allows to transfer this derivative to vector fields defined on U (that is, one takes directional derivatives of their components in the coordinate system defined by φ). Since M is compact, there is a finite atlas (U_i, φ_i) on M. Denote by ∇^i the covariant derivative on U_i induced by φ_i. Construct "bump functions" λ_i supported on U_i such that $\sum_i \lambda_i = 1$. Then, by Lemma 6.12, $\sum_i \lambda_i \nabla^i$ is a covariant derivative on M.

A fact similar to Fact 5.57 is that $(1,2)$-tensor fields correspond to $C^\infty(M)$-bilinear maps $\mathcal{T}M \times \mathcal{T}M \to \mathcal{T}M$. Thus in order to show that the map

$$(X, Y) \mapsto \nabla_X Y - \nabla'_X Y$$

is a $(1,2)$-tensor field, it suffices to prove its $C^\infty(M)$ bilinearity in X and Y. The linearity in X and the additivity in Y are inherited from ∇ and ∇'. The homogeneity in Y is proved by computation:

$$(\nabla - \nabla')_X(fY) = \nabla_X(fY) - \nabla'_X(fY)$$
$$= D_X f \cdot Y + f\,\nabla_X Y - (D_X f \cdot Y + f\,\nabla'_X Y)$$
$$= f(\nabla_X Y - \nabla'_X Y) = f\,(\nabla - \nabla')_X Y.$$

The converse statement about adding a tensor to a covariant derivative is proved similarly. □

Corollary 6.14. *Locally in any coordinate system a covariant derivative has the form*

$$\nabla_X Y = (X^j\,\partial_j Y^i + \Gamma^i_{jk} X^j Y^k)\partial_i,$$

and the coefficients Γ^i_{jk} transform by the tensor rule.

6.4 Covariant derivative of tensors

In this section we will explain how a covariant derivative of vector fields extends to a covariant derivative of tensor fields. We will follow two simple principles:

1. Leibniz rule for tensor product:

$$\nabla(\alpha \otimes \beta) = \nabla\alpha \otimes \beta + \alpha \otimes \nabla\beta.$$

2. Leibniz rule for the canonical pairing between TM and T^*M:

$$\nabla\langle\omega, X\rangle = \langle\nabla\omega, X\rangle + \langle\omega, \nabla X\rangle. \tag{35}$$

Let us derive from the second principle the formula for the covariant derivative of a differential 1-form. First, it is probably not quite clear what this second principle means... The pairing $\langle\omega, X\rangle$ is a function on M, and by ∇f we mean the differential 1-form df. Then, similarly to our second definition of the covariant derivative of vector fields, $\nabla\omega$ is a $(0,2)$-tensor field for which one is used to place the arguments in the following way:

$$(\nabla\omega)(X, Y) = (\nabla_X\omega)(Y).$$

We will omit some brackets by declaring the priority of differentiation over evaluation, so that the above rule writes $\nabla\omega(X, Y) = \nabla_X\omega(Y)$. The expression $\langle\nabla\omega, X\rangle$ means partial evaluation of $\nabla\omega$ where X takes the place of the second argument:

$$\langle\nabla\omega, X\rangle(Y) = \nabla_Y\omega(X).$$

Finally, $\langle\omega, \nabla X\rangle$ is the evaluation of the $(1,1)$-tensor ∇X on the $(0,1)$-tensor ω (a $(1,1)$-tensor can be viewed as a map $V^* \to V^*$).

The above explanations probably become more clear if we evaluate both sides of (35) on an arbitrary vector field. For aesthetic reasons, let us replace X by Y and evaluate on X:

$$\nabla_X\langle\omega, Y\rangle = \langle\nabla_X\omega, Y\rangle + \langle\omega, \nabla_X Y\rangle$$

Rereading the previous paragraph provides the following interpretation of this equation:

$$D_X\langle\omega, Y\rangle = \nabla_X\omega(Y) + \omega(\nabla_X Y).$$

Our goal is to define a 1-form $\nabla_X\omega$, and the above equation describes it by its action on an arbitrary vector field:

$$\nabla_X\omega(Y) = D_X\langle\omega, Y\rangle - \omega(\nabla_X Y). \tag{36}$$

Next, let us learn how to derive a bilinear form h. On one hand, every bilinear form is a sum of tensor products of 1-forms: $h = \sum_i \omega_i \otimes \eta_i$, and we can use (36) together with the first Leibniz rule. On the other hand, we can look at $h(X, Y)$ as a combination of tensor products and traces/contractions and apply the general Leibniz rule directly to this triple product:

$$\nabla(h(X, Y)) = \nabla h(X, Y) + h(\nabla X, Y) + h(X, \nabla Y),$$

or, by substituting an arbitrary direction of differentiation (and renaming the variables)

$$D_X(h(Y, Z)) = \nabla_X h(Y, Z) + h(\nabla_X Y, Z) + h(X, \nabla_Y Z).$$

This leads to a description of $\nabla_X h$ by its action:

$$\nabla_X h(Y, Z) = D_X(h(Y, Z)) - h(\nabla_X Y, Z) - h(X, \nabla_Y Z). \tag{37}$$

6.5 The Levi-Civita connection

Lemma-Definition 6.15. *The* torsion tensor *of a connection ∇ is a $(1, 2)$-tensor field defined by*

$$T^\nabla(X, Y) = \nabla_X Y - \nabla_Y X - [X, Y].$$

A connection is called torsion-free *if its torsion tensor is identically zero.*

The proof that the above formula defines a tensor field is left as an exercise.

The result is quite surprising: although the individual summands in the definition of $T^\nabla(X, Y)_p$ depend on a local behavior of either X or Y or both, the combination depends only on the values X_p and Y_p.

Definition 6.16. *A covariant derivative ∇ on a Riemannian manifold (M, g) is called* Levi-Civita covariant derivative *or* Levi-Civita connection *if it is torsion-free and satisfies $\nabla g = 0$.*

For $\nabla g = 0$ one also says that the metric tensor g is *parallel with respect to* ∇ or that ∇ is a *metric connection.*

Here is a summary of the properties of a Levi-Civita covariant derivative:

$$\begin{aligned} &\nabla_X Y - \nabla_Y X = [X, Y] && \text{torsion-free property} \\ &D_X(g(Y, Z)) = g(\nabla_X Y, Z) + g(Y, \nabla_X Z) && \text{metric property} \end{aligned}$$

The following is the main theorem of this section.

Theorem 6.17. *For every Riemannian metric there is a unique Levi-Civita connection.*

Lecture 14

Proof of the uniqueness of a Levi-Civita connection. Assume that ∇ and ∇' are two Levi-Civita connections. By Theorem 6.13, the difference $\nabla' - \nabla$ is a $(1, 2)$-tensor field. In other words, it is a smooth family of bilinear maps

$$A_p \colon T_pM \times T_pM \to T_pM, \quad A(X, Y) = \nabla'_X Y - \nabla_X Y. \tag{38}$$

We have to show that $A = 0$. It is convenient to transform A into a $(0, 3)$-tensor with the help of g:

$$A^\flat(X, Y, Z) = g(A(X, Y), Z).$$

Claim 1. The tensor $A^\flat$ is symmetric in the first two arguments:

$$A^\flat(X, Y, Z) = A^\flat(Y, X, Z).$$

It suffices to prove the symmetry of the bilinear map A defined in (38). This symmetry follows from the torsion-free property of ∇ and ∇':

$$\begin{aligned} A(X, Y) - A(Y, X) &= (\nabla'_X Y - \nabla_X Y) - (\nabla'_Y X - \nabla_Y X) \\ &= (\nabla'_X Y - \nabla'_Y X) - (\nabla_X Y - \nabla_Y X) = [X, Y] - [X, Y] = 0. \end{aligned}$$

Claim 2. The tensor $A^\flat$ is antisymmetric in the last two arguments:

$$A^\flat(X, Y, Z) = -A^\flat(X, Z, Y).$$

For this we will use the metric property of ∇' and ∇. It says

$$g(\nabla'_X Y, Z) + g(Y, \nabla'_X Z) = D_X(g(Y, Z)) = g(\nabla_X Y, Z) + g(Y, \nabla_X Z).$$

Taking the difference of the left hand side and the right hand side one gets

$$g(\nabla'_X Y - \nabla_X Y, Z) + g(\nabla'_X Z - \nabla_X Z, Y) = 0,$$

which is equivalent to $A^\flat(X, Y, Z) + A^\flat(X, Z, Y) = 0$.

Using Claim 1 and then Claim 2, one obtains

$$A^\flat(X, Y, Z) = A^\flat(Y, X, Z) = -A^\flat(Y, Z, X).$$

Thus, a cyclic permutation of arguments changes the sign of A. Iterating the permutation one obtains

$$A^\flat(X,Y,Z) = -A^\flat(Y,Z,X) = A^\flat(Z,X,Y) = -A^\flat(X,Y,Z).$$

Thus $A(X,Y,Z) = 0$ for all X,Y,Z and the uniqueness of a Levi-Civita connection is proved. $\square$

Proof of the existence of a Levi-Civita connection. Let ∇ be any connection on M, which exists due to Theorem 6.13. We want to find a $(1,2)$-tensor A such that the connection $\nabla' = \nabla + A$ is metric and torsion-free. One has

$$\begin{aligned}\nabla' \text{ is torsion-free } &\Leftrightarrow \nabla_X Y + A(X,Y) - \nabla_Y X - A(Y,X) = [X,Y]\\ &\Leftrightarrow A(Y,X) - A(X,Y) = T^\nabla(X,Y),\end{aligned}$$

where T^∇ denotes the torsion tensor of ∇. Further,

$$\begin{aligned}\nabla' \text{ is metric } &\Leftrightarrow g(\nabla_X Y + A(X,Y), Z) + g(Y, \nabla_X Z + A(X,Z)) = D_X(g(Y,Z))\\ &\Leftrightarrow A^\flat(X,Y,Z) + A^\flat(X,Z,Y) = D_X(g(Y,Z)) - g(\nabla_X Y, Z) - g(Y, \nabla_X Z)\\ &\Leftrightarrow A^\flat(X,Y,Z) + A^\flat(X,Z,Y) = (\nabla_X g)(Y,Z).\end{aligned}$$

Thus, ∇' is a Levi-Civita connection if and only if the tensor field $A^\flat$ satisfies

$$A^\flat(X,Y,Z) - A^\flat(Y,X,Z) = -g(T^\nabla(X,Y),Z) \tag{39}$$

$$A^\flat(X,Y,Z) + A^\flat(X,Z,Y) = (\nabla_X g)(Y,Z) \tag{40}$$

for all vectors X,Y,Z. In the second equation, substitute X for Y and Y for X. This yields

$$A^\flat(Y,X,Z) + A^\flat(Y,Z,X) = (\nabla_Y g)(X,Z).$$

Adding this to (39) one obtains

$$A^\flat(X,Y,Z) + A^\flat(Y,Z,X) = (\nabla_Y g)(X,Z) - g(T^\nabla(X,Y),Z).$$

By cyclically permuting the arguments in the above equation one gets two more equations

$$\begin{aligned}A^\flat(Y,Z,X) + A^\flat(Z,X,Y) &= (\nabla_Z g)(Y,X) - g(T^\nabla(Y,Z),X)\\ A^\flat(Z,X,Y) + A^\flat(X,Y,Z) &= (\nabla_X g)(Z,Y) - g(T^\nabla(Z,X),Y).\end{aligned}$$

The last three equations allows to express $A^\flat(X,Y,Z)$ as follows:

$$\begin{aligned}2A^\flat(X,Y,Z) = {}&(\nabla_X g)(Y,Z) + (\nabla_Y g)(Z,X) - (\nabla_Z g)(X,Y)\\ &- g(T^\nabla(Z,X),Y) - g(T^\nabla(X,Y),Z) + g(T^\nabla(Y,Z),X).\end{aligned} \tag{41}$$

A simple check shows that this expression for $A^\flat$ satisfies conditions (39) and (40). Thus $\nabla' = \nabla + A$ is a Levi-Civita connection. $\square$

Lemma 6.18 (Koszul formula)**.** *The Levi-Civita connection can be expressed in terms of the Riemannian metric as*

$$g(\nabla_X Y, Z) = \frac{1}{2}(D_X(g(Y,Z)) + D_Y(g(Z,X)) - D_Z(g(X,Y)) \\ + g([Z,X],Y) + g([X,Y],Z) - g([Y,Z],X)).$$

Proof. This can be proved with the help of the formula (41) and is left as an exercise. □

Lemma 6.19. *The covariant derivative on a surface in* $\mathbb{R}^3$ *as defined in Section 4.8 is the Levi-Civita covariant derivative.*

Proof. This follows from Lemma 4.31. Properties 1 and 2 say that the surface covariant derivative is a covariant derivative in the sense of Definition 6.10. Property 3 is the metric property, and the torsion-free property can be derived from $\nabla_X Y - \nabla_Y X = D_X Y - D_Y X$. □

6.6 The Hessian and the Laplacian

From now on, (M, g) is a Riemannian manifold, and ∇ is the Levi-Civita connection associated with g.

Definition 6.20. *The* Hessian tensor *of a smooth function* $f \in \mathbb{C}^\infty(M)$ *is the Levi-Civita covariant derivative of the differential of* f*:*

$$\operatorname{Hess} f = \nabla(df).$$

Sometimes, for convenience, one uses the notation ∇f for the differential of a function, which leads to the notation $\nabla^2 f$ for the Hessian.

The differential df is a differential 1-form. Thus by Section 6.4, and in particular by (36) one has

$$\operatorname{Hess} f(X,Y) = \nabla_X(df)(Y) = D_X D_Y f - D_{\nabla_X Y} f.$$

Lemma 6.21. *The Hessian is symmetric:* $\operatorname{Hess} f(X,Y) = \operatorname{Hess} f(Y,X)$.

Proof. Left as an exercise. □

The Laplacian Δf of f is the trace of the Hessian of f. As the Hessian is a $(0,2)$-tensor, its trace is actually not defined (if one looks at it in terms of matrices, the matrix of a bilinear form transforms under a basis change as $A \mapsto C^\top AC$, and $\operatorname{tr} C^\top AC \neq \operatorname{tr} A$ in general). However, the musical isomorphisms defined by the inner product g allows to make from the Hessian $(0,2)$-tensor a Hessian operator. The Laplacian of f is the trace of this operator.

It can be shown that

$$\Delta f = \sum_{i=1}^n \operatorname{Hess} f(e_i, e_i)$$

for any orthonormal basis (e_i) of T_pM.

6.7 The Riemann curvature tensor

This is the main concept of Riemannian geometry. Its standard definition might look confusing and as something out of a clear sky.

Lemma-Definition 6.22. *The map $\mathcal{T}M \times \mathcal{T}M \times \mathcal{T}M \to \mathcal{T}M$ defined and denoted by*

$$(X, Y, Z) \mapsto R(X, Y)Z := \nabla_X \nabla_Y Z - \nabla_Y \nabla_X Z - \nabla_{[X,Y]} Z$$

is $C^\infty(M)$-multilinear. Thus it defines a $(1, 3)$-tensor field, called the Riemann curvature tensor*. (Here we denote $\nabla_X \nabla_Y Z := \nabla_X(\nabla_Y Z)$.)*

If $M \subset \mathbb{R}^n$, and g is the standard Euclidean metric, then the Levi-Civita derivative is the componentwise directional derivative. The i-th component of $R(X, Y)Z$ is then

$$D_X D_Y Z^i - D_Y D_X Z^i - D_{[X,Y]} Z^i,$$

which is equal to 0 by definition of the commutator. Thus the curvature tensor of the Euclidean space vanishes. Historically the first result in Riemannian geometry was the theorem of Riemann that the curvature tensor vanishes if and only if (M, g) is locally isometric to the Euclidean space. We just have proved the "if" part. The "only if" part is more complicated. Riemann's original argument can be found in [Spi79].

Remark 6.23. An equivalent but less often used definition is

$$R(X, Y)Z := \nabla^2_{X,Y} Z - \nabla^2_{Y,X} Z, \tag{42}$$

where $\nabla^2 Z$ is the second covariant derivative of the vector field Z. Indeed, ∇Z is a $(1, 1)$-tensor field acting on vector fields by $(\nabla Z)(X) = \nabla_X Z$, and $\nabla^2 Z = \nabla(\nabla Z)$ is a $(1, 2)$-tensor field. By definition of the covariant derivative of tensor fields one has

$$\nabla^2_{X,Y} Z = \nabla(\nabla Z)(X, Y) = \nabla_X(\nabla Z)(Y) = \nabla_X \nabla_Y Z - \nabla_{\nabla_X Y} Z.$$

It follows that

$$\begin{aligned}\nabla^2_{X,Y} Z - \nabla^2_{Y,X} Z = \nabla_X \nabla_Y Z - \nabla_Y \nabla_X Z - \nabla_{\nabla_X Y - \nabla_Y X} Z \\ = \nabla_X \nabla_Y Z - \nabla_Y \nabla_X Z - \nabla_{[X,Y]} Z = R(X, Y)Z,\end{aligned}$$

as claimed.

Compare equation (42) with the definition of the Hessian. While the second covariant derivative of a function is symmetric in its arguments, the second covariant derivative of a vector field is in general not. The Riemann curvature tensor is the skew-symmetric part of this second covariant derivative.

We still have to explain why R is a tensor. One can use the above remark to show that for every Z the map $(X, Y) \mapsto R(X, Y)Z$ is $C^\infty(M)$-bilinear (since $\nabla^2 Z$ is bilinear), but one can also prove this directly. This direct proof as well as the $C^\infty(M)$-linearity in Z are left to you as an exercise.

It is sometimes convenient to transform the $(1,3)$-tensor field R to a $(0,4)$-tensor field in the usual way (the new tensor is traditionally denoted by the same symbol):

$$R(X,Y,Z,W) := g(R(X,Y)Z,W).$$

This tensor field satisfies remarkable relations listed in the next theorem.

Theorem 6.24. *The curvature tensor of a Levi-Civita connection has the following properties.*

1. $R(X,Y,Z,W) = -R(Y,X,Z,W)$
2. $R(X,Y,Z,W) = -R(X,Y,W,Z)$
3. $R(X,Y,Z,W) = R(Z,W,X,Y)$
4. $R(X,Y)Z + R(Y,Z)X + R(Z,X)Y = 0$
5. $(\nabla_X R)(Y,Z) + (\nabla_Y R)(Z,X) + (\nabla_Z R)(X,Y) = 0$

The properties 1–4 are of algebraic nature, that is we work here with the tensor R_p over the vector space T_pM, and X,Y,Z,W as vectors in T_pM. The property 5 is differential, and X,Y,Z need to be vector fields.

Property 4 is called the *first (or algebraic) Bianchi identity*; property 5 is called the *second (or differential) Bianchi identity.*

Proof. Property 1 is straightforward from the definition.

For property 2 it suffices to show that $R(X,Y,Z,Z) = 0$ for all X,Y,Z. Indeed, this would mean that for all X,Y the bilinear form $(Z,W) \mapsto R(X,Y,Z,W)$ is alternating, and this would imply that it is skew-symmetric.

Let Z be any vector field. Compute the Hessian of its squared norm while using $\nabla g = 0$:

$$\begin{aligned}
\mathrm{Hess}(g(Z,Z))(X,Y) &= \nabla_X\nabla_Y(g(Z,Z)) - \nabla_{\nabla_X Y}(g(Z,Z)) \\
&= 2\nabla_X(g(\nabla_Y Z, Z)) - 2g(\nabla_{\nabla_X Y}Z, Z) \\
&= 2g(\nabla_X\nabla_Y Z, Z) + 2g(\nabla_X Z, \nabla_Y Z) - 2g(\nabla_{\nabla_X Y}Z, Z) \\
&= 2g(\nabla_X\nabla_Y Z - \nabla_{\nabla_X Y}Z, Z) + 2g(\nabla_X Y, \nabla_Y Z).
\end{aligned}$$

Using the symmetry $\mathrm{Hess}(g(Z,Z))(X,Y) = \mathrm{Hess}(g(Z,Z))(Y,X)$ one obtains $g(R(X,Y,Z),Z) = 0$ and thus $R(X,Y,Z,Z) = 0$.

Property 4 follows from the Jacobi identity for the commutator of vector fields:

$$\begin{aligned}
&R(X,Y)Z + R(Y,Z)X + R(Z,X)Y \\
&= \nabla_X(\nabla_Y Z - \nabla_Z Y) + \nabla_Y(\nabla_Z X - \nabla_X Z) + \nabla_Z(\nabla_X Y - \nabla_Y X) - \nabla_{[X,Y]}Z - \nabla_{[Y,Z]}X - \nabla_{[Z,X]}Y \\
&= \nabla_X[Y,Z] + \nabla_Y[Z,X] + \nabla_Z[X,Y] - \nabla_{[Y,Z]}X - \nabla_{[Z,X]}Y - \nabla_{[X,Y]}Z \\
&= [X,[Y,Z]] + [Y,[Z,X]] + [Z,[X,Y]] = 0
\end{aligned}$$

The harmless-looking property 3 is derived from the properties 1, 2, and 4 by intricate combinations. First, write two instances of the algebraic Bianchi identity:

$$R(X,Y,Z,W) + R(Y,Z,X,W) + R(Z,X,Y,W) = 0$$
$$R(Z,W,X,Y) + R(W,X,Z,Y) + R(X,Z,W,Y) = 0.$$

Taking their difference and using properties 1 and 2 we get

$$R(X,Y,Z,W) + R(Y,Z,X,W) = R(Z,W,X,Y) + R(X,W,Y,Z).$$

Our goal is to show that the terms on the left and on the right are equal in pairs. For this, permute the arguments X, Y, Z cyclically to obtain two similar equalities:

$$R(Y,Z,X,W) + R(Z,X,Y,W) = R(X,W,Y,Z) + R(Y,W,Z,X)$$
$$R(Z,X,Y,W) + R(X,Y,Z,W) = R(Y,W,Z,X) + R(Z,W,X,Y).$$

Of our last three equalities, adding the first and the third and subtracting the second, we obtain

$$R(X,Y,Z,W) = R(Z,W,X,Y).$$

The last property 5 can be proved by brute force, using the definition of the covariant derivative of a tensor and everything we know about the symmetries of the curvature tensor. The computation is long; you can do it to test your skills. □

The differential Bianchi identity has a conceptual interpretation and a far shorter proof than the one (not) given above. For this, see Section 6.8.

Remark 6.25. The properties 1–3 of the curvature tensor coincide with the symmetries of the cross-ratio of four numbers

$$[x,y,z,w] = \frac{(x-z)(y-w)}{(x-w)(y-z)}.$$

As an analog of the property 4 one may consider the identity

$$[x,y,z,w] \cdot [y,z,x,w] \cdot [z,x,y,w] = -1.$$

I do not know of any explanation to these coincidences.

6.8* Vector-valued differential forms and the curvature tensor

There is another, more general, approach to the curvature, which provides a more conceptual proof of the differential Bianchi identity.

Define *vector-valued differential forms* as elements of the set

$$\Omega^q(M;\mathcal{T}M) := \Omega^q(M) \otimes_{C^\infty(M)} \mathcal{T}M$$

(the tensor product of $C^\infty(M)$-modules). In particular, while $\Omega^0(M)$ is the space of smooth functions, $\Omega^0(M; \mathcal{T}M)$ is the space of vector fields. In other words, a vector-valued differential q-form is a smooth family of skew-symmetric maps

$$\underbrace{T_pM \times \cdots \times T_pM}_{q} \to T_pM.$$

Every element of $\Omega^q(M; \mathcal{T}M)$ can be written as a sum $\sum Z_i \otimes \omega_i$, where Z_i are vector fields and ω_i are differential q-forms.

Given a covariant derivative ∇ on $\mathcal{T}M$, define the associated exterior derivative

$$\begin{aligned} d^\nabla \colon \Omega^q(M; \mathcal{T}M) &\to \Omega^{q+1}(M; \mathcal{T}M) \\ d^\nabla(Z \otimes \omega) &= \nabla Z \wedge \omega + Z \otimes d\omega. \end{aligned}$$

While the classical exterior derivative satisfies $d \circ d = 0$, this is no more the case for d^∇. (Remember that $d \circ d = 0$ is equivalent to the symmetry of the Hessian on functions, and the skew-symmetric part of the second derivative of vector field is the curvature tensor, so the curvature tensor will be no coming from $d^\nabla \circ d^\nabla \neq 0$.)

Lemma 6.26. *For every $q \geq 0$, the composition*

$$\Omega^q(M; E) \xrightarrow{d^\nabla} \Omega^{q+1}(M; E) \xrightarrow{d^\nabla} \Omega^{q+2}(M; E)$$

is $\Omega^k(M)$-linear:

$$(d^\nabla)^2(Z \otimes \omega) = (d^\nabla)^2 Z \wedge \omega$$

for all $Z \in \mathcal{T}M$, $\omega \in \Omega^q(M)$.

In particular, for $q = 0$ we get a $C^\infty(M)$-linear map from vector fields to vector-valued differential 2-forms. Pointwise this is a map from T_pM to $\bigwedge^2 T_p^*M \otimes T_pM$, that is a $(1,3)$-tensor.

Definition 6.27. *The $(1,3)$-tensor field corresponding to the $C^\infty(M)$-linear map*

$$R^\nabla := (d^\nabla)^2 \colon \mathcal{T}(M) \to \Omega^2(M) \otimes \mathcal{T}(M)$$

is called the curvature *of the connection ∇.*

Remark 6.28. More generally, one can consider any *vector bundle $\mathcal{E}$* over a smooth manifold M, that is a collection of vector spaces attached to each point of M. The dimension of these vector spaces might be different from the dimension of M.

There is a notion of a connection in a vector bundle, and the above approach defines the curvature of such a connection as an $\mathrm{End}(\mathcal{E})$-valued differential 2-form.

Theorem 6.29. *The exterior derivative of the curvature R^∇ viewed as an* $\mathrm{End}(\mathcal{T}(M))$*-valued differential* 2*-form vanishes: $d^\nabla R^\nabla = 0$.*

Proof. One has

$$(d^\nabla R^\nabla)(Z) = d^\nabla(R^\nabla Z) - R^\nabla(d^\nabla Z) = (d^\nabla)^3 Z - (d^\nabla)^3 Z = 0.$$

This should be thought over, because there are different d^∇'s here. $\square$

One can show that

$$\begin{aligned}(d^\nabla \omega)(X_0, \ldots, X_q) = &\sum_{i=0}^{q} (-1)^i \nabla_{X_i}(\omega(X_0, \ldots, \widehat{X_i}, \ldots, X_q)) \\ &+ \sum_{0 \le i < j \le q} (-1)^{i+j} \omega([X_i, X_j], X_0, \ldots, \widehat{X_i}, \ldots, \widehat{X_j}, \ldots, X_q).\end{aligned}$$

The differential Bianchi identity (property 5 of Theorem 6.24) is a reformulation of Theorem 6.29. Indeed, one has

$$\begin{aligned}d^\nabla R(X, Y, Z) = &\nabla_X(R(Y, Z)) - \nabla_Y(R(X, Z)) + \nabla_Z(R(X, Y)) \\ &- R([X, Y], Z) + R([X, Z], Y) - R([Y, Z], X) \\ &= (\nabla_X R)(Y, Z) + R(\nabla_X Y, Z) + R(Y, \nabla_X Z) \\ &- (\nabla_Y R)(X, Z) - R(\nabla_Y X, Z) - R(X, \nabla_Y Z) \\ &+ (\nabla_Z R)(X, Y) + R(\nabla_Z X, Y) + R(X, \nabla_Z Y) \\ &- R([X, Y], Z) + R([X, Z], Y) - R([Y, Z], X) \\ &= (\nabla_X R)(Y, Z) + (\nabla_Y R)(Z, X) + (\nabla_Z R)(X, Y).\end{aligned}$$

6.9 Sectional curvature

Definition 6.30. *Let (M, g) be a Riemannian manifold, $p \in M$ a point in M, and $\pi \subset T_pM$ a 2-dimensional linear subspace of T_pM. The* sectional curvature *of M at the plane π is defined as*

$$\sec(\pi) = -\frac{R(X, Y, X, Y)}{g(X, X)g(Y, Y) - g(X, Y)^2}, \tag{43}$$

where (X, Y) is any basis of π.

Observe that the denominator in (43) is the determinant of the Gram matrix of (X, Y).

The value of the right hand side in (43) is independent of the choice of a basis of π. Indeed, it can be shown that for a basis (X', Y') related to (X, Y) by

$$\begin{pmatrix} X' \\ Y' \end{pmatrix} = A \begin{pmatrix} X \\ Y \end{pmatrix}, \quad A \in \mathrm{GL}(2)$$

one has

$$\begin{aligned}R(X', Y', X', Y') &= (\det A)^2 R(X, Y, X, Y), \\ g(X', X')g(Y', Y') - g(X', Y')^2 &= (\det A)^2 (g(X, X)g(Y, Y) - g(X, Y)^2).\end{aligned}$$

Remark 6.31. The Riemann curvature tensor at $p \in M$ can also be viewed as a symmetric bilinear form on the exterior square $\Lambda^2 T_pM$. The inner product g in T_pM induces an inner product on $\Lambda^2 T_pM$. This allows to write the definition of the sectional curvature as follows:

$$\sec(\pi) = -\frac{R(X \wedge Y, X \wedge Y)}{g(X \wedge Y, X \wedge Y)}.$$

The inner product on $\Lambda^2 T_pM$ can be used to transform the symmetric bilinear form R to a self-adjoint endomorphism of $\Lambda^2 T_pM$, called the curvature operator. If the curvature operator is positive definite, then all sectional curvatures are positive. The converse is not true.

Theorem 6.32. *The sectional curvature and the metric determine the curvature tensor.*

Proof. We will show that if we know the values of $R(X, Y, X, Y)$ for all $X, Y \in T_pM$, then we can compute $R(X, Y, Z, W)$ for any $X, Y, Z, W \in T_pM$.

Take any X, Y, Z, W and consider the function

$$F\colon \mathbb{R}^2 \to \mathbb{R},$$
$$F(\lambda, \mu) = R(X + \lambda Z, Y + \mu W, X + \lambda Z, Y + \mu W) - R(X + \lambda W, Y + \mu Z, X + \lambda W, Y + \mu Z)$$

By assumption, we know F. Due to the multilinearity of R, the function F is a polynomial in λ and μ. Let us write out the coefficients at the terms of degree ≤ 2. The symmetries of the curvature tensor (Properties 1–3 from Theorem 6.24) allow to simplify this to

$$\begin{aligned} F(\lambda, \mu) = {} & 2\lambda R(X, Y, Z - W, Y) + 2\mu R(X, Y, X, W - Z) \\ & + \lambda^2(R(Z, Y, Z, Y) - R(W, Y, W, Y)) + \mu^2(R(R(X, W, X, W) - R(X, Z, X, Z)) \\ & \qquad + 2\lambda\mu(2R(X, Y, Z, W) - R(Y, Z, X, W) - R(Z, X, Y, W)) + \cdots \end{aligned}$$

Due to the algebraic Bianchi identity, the coefficient at $\lambda\mu$ is equal to $6R(X, Y, Z, W)$. It follows that

$$\left.\frac{\partial^2 F}{\partial\lambda\partial\mu}\right|_{(0,0)} = 6R(X, Y, Z, W),$$

thus the sectional curvatures determine the curvature tensor. □

Theorem 6.33. *A Riemannian manifold has constant sectional curvature k if and only if*

$$R(X, Y, Z, W) = k(g(X, W)g(Y, Z) - g(X, Z)g(Y, W)). \tag{44}$$

Proof. Substituting the above expression into (43) one gets $\sec(\pi) = k$.

In the opposite direction, the right hand side of (44) is linear in each of the four arguments and, as can easily be checked, satisfies properties 1–4 from Theorem 6.24. In Theorem 6.32 we have used only these properties for reconstruction of R from sec. Thus if we start from $\sec = k$, then we unavoidably obtain (44). □

The above theorem does not guarantee that there are Riemannian manifolds with constant sectional curvature k for any k. Let us construct such examples in all dimensions $n \geq 2$.

For $k = 0$ one can take the Euclidean space.

Lemma 6.34. *The unit sphere* $\mathbb{S}^n = \{p \in \mathbb{R}^{n+1} \mid \|p\| = 1\}$ *has constant sectional curvature* 1. *The sphere of radius* r *has constant sectional curvature* $\frac{1}{r^2}$.

Proof. It can be shown that the Levi-Civita covariant derivative in $\mathbb{S}^n$ can be expressed in terms of the directional derivative in $\mathbb{R}^{n+1}$ as

$$\nabla_X Y = D_X Y + \langle X, Y \rangle \nu,$$

where $\langle X, Y \rangle$ is the Euclidean inner product, and ν is the outward unit normal. (This is an analog of the formula $\nabla_X Y = D_X Y - II(X, Y)\nu$ we had for surfaces in $\mathbb{R}^3$.) Taking the second derivative one obtains

$$\nabla_X \nabla_Y Z = D_X D_Y Z + \langle Y, Z \rangle X + \text{ normal component},$$

which implies

$$R(X, Y)Z = D_X D_Y Z + \langle Y, Z \rangle X - D_Y D_X Z - \langle X, Z \rangle Y - D_{[X,Y]} Z = \langle Y, Z \rangle X - \langle X, Z \rangle Y.$$

(We ignored the normal components. Their sum vanishes because $R(X, Y)Z$ belongs to the tangent space.)

Theorem 6.33 implies $k = 1$.

As for the sphere of radius r, one can use a general observation that scaling the metric tensor g does not change the Levi-Civita connection (if g was parallel with respect to ∇, then λg will be parallel with respect to ∇). At the same time, the $(0, 4)$-tensor R scales by λ (it has one g inside), and the Gram determinant scales by λ^2. From the formula (43) we conclude that sec scales by $\frac{1}{\lambda}$. The metric on the sphere of radius r is the metric on the unit sphere scaled by r^2, which concludes the proof. □

The sphere has a rich group of isometries: any 2-dimensional subspace of $T_p\mathbb{S}^n$ can be sent to any 2-dimensional subspace of $T_q\mathbb{S}^n$ for any $p, q \in \mathbb{S}^n$. This implies the constance of the sectional curvature, but does not allow to compute its value.

Let us construct a Riemannian manifold of sectional curvature -1. Consider in $\mathbb{R}^{n+1} = \{(x_0, x_1, \ldots, x_n)\}$ the Lorentzian inner product

$$\langle x, y \rangle = -x_0 y_0 + x_1 y_1 + \cdots + x_n y_n$$

and the hypersurface

$$\mathbb{H}^n := \{p \mid \langle p, p \rangle = -1, p_0 > 0\}.$$

The restriction of the Lorentzian inner product to the Lorentzian orthogonal complement $p^\perp$ is positive definite. This induces a Riemannian metric on $\mathbb{H}^n$. Many things for $\mathbb{H}^n$ are the same as for $\mathbb{S}^n$. For example, the position vector is at the same time the unit normal (the orthogonality for $\mathbb{H}^n \subset \mathbb{R}^{n+1}$ is always understood in the Lorentzian sense). The main difference is that $\langle \nu, \nu \rangle = -1$. This leads to

$$\nabla_X Y = D_X Y - \langle X, Y \rangle \nu.$$

(Check that the right hand side lies in the tangent space to $\mathbb{H}^n$.) Computations similar to those in the proof of Lemma 6.34 lead to $k = -1$.

Remark 6.35. The Riemannian metric

$$\frac{1}{x_n^2}(dx_1^2 + \cdots + dx_n^2)$$

on the half-space $x_n > 0$ also has constant sectional curvature -1. To prove this, one would need to compute Christoffel symbols (see Corollary 6.14) and to express the curvature tensor in terms of them.

Remark 6.36. It can be shown that any two Riemannian manifolds of constant curvature k are locally isometric (for $k = 0$ this was Riemann's discovery discussed after Definition 6.22). The space from the previous remark is not only locally but also globally isometric to $\mathbb{H}^n$. The isometry can be given by a combination of three projections ($\mathbb{H}^n$ to a disk, the disk to a hemisphere, the hemisphere to the halfspace).

6.10 Curvature of submanifolds

Lecture 15

Recall the following definitions and identities from the differential geometry of smooth surfaces in $\mathbb{R}^3$:

$$II(X,Y) := -\langle D_X \nu, Y \rangle = \langle D_X Y, \nu \rangle,$$
$$\nabla_X Y := (D_X Y)_\top = D_X Y - II(X,Y)\nu.$$

Let $(N, \widetilde{g})$ be a Riemannian manifold, and let $M \subset N$ be a smooth submanifold of N. Then for any $p \in M$ the tangent space T_pM is a linear subspace of T_pN, and the restriction of the inner product $\widetilde{g}$ to T_pM is an inner product on T_pM, which we will denote g. One also says that g is the Riemannian metric on M *induced* from the Riemannian metric $\widetilde{g}$ on N or that (M, g) is a *Riemannian submanifold* of $(N, \widetilde{g})$.

Let us now generalize the facts from geometry of smooth surfaces mentioned above. We will use the orthogonal (with respect to $\widetilde{g}_p$) decomposition

$$T_pN = T_pM \oplus (T_pM)^\perp, \quad Z = Z_\top + Z_\perp.$$

The following is a preparatory lemma.

Lemma 6.37. *Let X and Y be vector fields on N such that for all $p \in M$ one has $X_p, Y_p \in T_pM$. Then one has $[X,Y]_p \in T_pM$ for all $p \in M$.*

Proof. The shortest way to prove this lemma is to use the coordinates. By definition of a smooth submanifold, in the neighborhood of every $p \in M$ there is a chart (U, φ) such that $\varphi(U \cap M) = \{x_{m+1} = \cdots = x_n = 0\}$ (we set $\dim M = m, \dim N = n$). The components of X and Y satisfy $X^j = Y^j = 0$ for $j > m$, and we have

$$[X, Y]^j = X^i \partial_i Y^j - Y^i \partial_i X^j = 0 \quad \text{for } j > m.$$

$\square$

Lemma 6.38. *Let (M, g) be a Riemannian submanifold of a Riemannian manifold $(N, \widetilde{g})$, and let $\widetilde{\nabla}$ be the Levi-Civita connection on $(N, \widetilde{g})$. Then the Levi-Civita connection ∇ on (M, g) is given by*

$$\nabla_X Y = (\widetilde{\nabla}_X \widetilde{Y})_\top, \tag{45}$$

where $\widetilde{Y}$ is any extension of the vector field Y from M to N.

Proof. First let us show that the right hand side is independent of the choice of an extension $\widetilde{Y}$. For this, it suffices to show that if $\widetilde{Y}$ is a vector field on N which vanishes on M, then $\widetilde{\nabla}_X \widetilde{Y}$ vanishes as well. Using the metric property of $\widetilde{\nabla}$ we get

$$0 = \widetilde{\nabla}_X \langle \widetilde{Y}, Z \rangle = \langle \widetilde{\nabla}_X \widetilde{Y}, Z \rangle + \langle \widetilde{Y}, \widetilde{\nabla}_X Z \rangle$$

for every vector field Z. The second term on the right hand side vanishes, thus $\widetilde{\nabla}_X \widetilde{Y} = 0$.

Due to the uniqueness of the Levi-Civita connection, it suffices to prove that $(\widetilde{\nabla}_X Y)_\top$ is a torsion-free metric connection on M.

First, why is this a connection/covariant derivative? The linearity of $\widetilde{\nabla}$ in X and the linearity of the projection imply the linearity of ∇ in X. The same holds for the additivity in Y. As for the Leibniz property in Y, compute:

$$\nabla_X (fY) = (\widetilde{\nabla}_X (fY))_\top = (D_X f \cdot Y + f\, \widetilde{\nabla}_X Y)_\top = D_X f \cdot Y + f \nabla_X Y.$$

Next, the torsion-free property:

$$\nabla_X Y - \nabla_Y X = (\widetilde{\nabla}_X Y - \widetilde{\nabla}_Y X)_\top = \top([X, Y]) = [X, Y],$$

where we used the torsion-free property of $\widetilde{\nabla}$ and the fact that the commutator of vector fields tangent to a submanifold is tangent to the submanifold. The latter can be proved with the help of coordinates or by looking at the derivatives of the distance function to the submanifold.

Finally, the metric property:

$$D_X(g(Y, Z)) = D_X(\widetilde{g}(Y, Z)) = \widetilde{g}(\widetilde{\nabla}_X Y, Z) + \widetilde{g}(Y, \widetilde{\nabla}_X Z) = g(\nabla_X Y, Z) + g(Y, \nabla_X Z),$$

where the last equality holds because the difference $\widetilde{\nabla}_X Y - \nabla_X Y$ is orthogonal to $T_p M$ and in particulat to Z. $\square$

Let us restrict ourselves to the case when M is a submanifold of codimension 1, that is $\dim N - \dim M = 1$. Then the orthogonal complement of T_pM in T_pN has dimension 1 and locally there is a unique up to a sign field of unit normals ν to M in N. This field of unit normals allows to express the image of $\top$ as $\top(Z) = Z - \widetilde{g}(Z,\nu)\nu$, so that one has

$$\nabla_X Y = \widetilde{\nabla}_X Y - \widetilde{g}(\widetilde{\nabla}_X Y, \nu)\nu = \widetilde{\nabla}_X Y + g(\widetilde{\nabla}_X \nu, Y)\nu.$$

(In the last term we may write g instead of $\widetilde{g}$ because $\widetilde{\nabla}_X \nu \in TM$.)

Definition 6.39. *The* second fundamental form *of a codimension* 1 *Riemannian submanifold $M \subset N$ is defined as*

$$II(X,Y) = -g(\widetilde{\nabla}_X \nu, Y).$$

This generalizes the case of surfaces in $\mathbb{R}^3$, where $\widetilde{\nabla} = D$ was the directional derivative. One shows that $II(X,Y) = II(Y,X)$ similarly to the case of surfaces.

Thus, in the codimension 1 case the formula (45) can be rewritten as

$$\nabla_X Y = \widetilde{\nabla}_X Y - II(X,Y)\nu.$$

Let $\widetilde{R}$ and R denote the curvature tensors (both in their $(1,3)$ an in $(0,4)$ versions) of $\widetilde{g}$ and of g, respectively. The following two theorems are among the most fundamental results of Riemannian geometry.

Theorem 6.40 (Gauss equation)**.** *For every $p \in M$ and every $X, Y, Z, W \in T_pM$ one has*

$$\widetilde{R}(X,Y,Z,W) = R(X,Y,Z,W) - II(X,W)II(Y,Z) + II(X,Z)II(Y,W).$$

Theorem 6.41 (Codazzi-Mainardi equation)**.** *For every $p \in M$ and every $X, Y, Z \in T_pM$ we have*

$$\widetilde{R}(X,Y,Z,\nu) = (\nabla_X II)(Y,Z) - (\nabla_Y II)(X,Z).$$

Proof. Just keep calculating:

$$\begin{aligned}
\widetilde{R}(X,Y)Z &= \widetilde{\nabla}_X\widetilde{\nabla}_Y Z - \widetilde{\nabla}_Y\widetilde{\nabla}_X Z - \widetilde{\nabla}_{[X,Y]}Z \\
&= \widetilde{\nabla}_X(\nabla_Y Z + II(Y,Z)\nu) - \widetilde{\nabla}_Y(\nabla_X Z + II(X,Z)\nu) - (\nabla_{[X,Y]}Z + II([X,Y],Z)\nu) \\
&= \nabla_X\nabla_Y Z + II(X,\nabla_Y Z)\nu + D_X(II(Y,Z))\nu + II(Y,Z)\widetilde{\nabla}_X\nu \\
&\quad - \nabla_Y\nabla_X Z - II(Y,\nabla_X Z)\nu - D_Y(II(X,Z))\nu - II(X,Z)\widetilde{\nabla}_Y\nu \\
&\quad - \nabla_{[X,Y]}Z - II([X,Y],Z)\nu \\
&= R(X,Y)Z + II(Y,Z)\widetilde{\nabla}_X\nu - II(X,Z)\widetilde{\nabla}_Y\nu + (\nabla_X II)(Y,Z)\nu - (\nabla_Y II)(X,Z)\nu
\end{aligned}$$

When we take the inner product with a vector $W \in T_pM$, only the first three terms on the right hand side matter, and we obtain the Gauss equation. When we take the inner product with a unit normal, only the last two terms matter, and we obtain the Codazzi-Mainardi equation. $\square$

Consider the case when $N = \mathbb{R}^{n+1}$ with the Euclidean metric. Then $\widetilde{\nabla} = D$ is the directional derivative, and $\widetilde{R} = 0$. The Gauss and the Codazzi-Mainardi equations take the form

$$\begin{gathered} II(X, W)II(Y, Z) - II(X, Z)II(Y, W) = R(X, Y, Z, W) \\ (\nabla_X II)(Y, Z) - (\nabla_Y II)(X, Z) = 0. \end{gathered} \tag{46}$$

If $n = 2$, then M is a smooth surface in the Euclidean space. Let us show that the Gauss equation implies Theorema Egregium: the Gauss curvature of a surface is determined by the intrinsic metric. Indeed, substituting $Z = X$ and $W = Y$ one obtains

$$\frac{\det II}{\det I} = \frac{II(X, X)II(Y, Y) - II(X, Y)^2}{g(X, X)g(Y, Y) - g(X, Y)^2} = -\frac{R(X, Y, X, Y)}{g(X, X)g(Y, Y) - g(X, Y)^2} = \sec(T_pM).$$

Since the sectional curvature is determined by Riemannian metric, the quotient of determinants of fundamental forms also is.

The following theorem is a far-reaching generalization of Theorem 1.37. We give no proof.

Theorem 6.42 (Fundamental theorem of hypersurface theory)**.** *Let (M, g) be an n-dimensional Riemannian manifold ($n \geq 2$), and let II be a field of symmetric bilinear forms on M which satisfies equations* (46)*. Then every point $p \in M$ has a neighborhood that embeds into $\mathbb{R}^{n+1}$ isometrically and with II as its second fundamental form. Besides, this embedding is unique up to a rigid motion in $\mathbb{R}^{n+1}$.*

If M is homeomorphic to a disk (or, more generally, simply-connected), then there is an isometric immersion of M to $\mathbb{R}^{n+1}$ with the above properties.

6.11 Ricci curvature, scalar curvature, Einstein manifolds

Everywhere in this section $(e_1, \ldots, e_n)$ is an orthonormal basis of the tangent space T_pM for some $p \in M$.

Definition 6.43. *The* Ricci curvature *or the* Ricci tensor *is a $(0, 2)$-tensor field obtained by metric contraction of the first index of the curvature tensor with the fourth index:*

$$\mathrm{Ric}(X, Y) = \sum_{i=1}^{n} R(e_i, X, Y, e_i).$$

Lemma 6.44. *The Ricci tensor is symmetric.*

Proof. Indeed, every summand in the definition of Ric is symmetric:

$$R(e_i, X, Y, e_i) = R(Y, e_i, e_i, X) = -R(e_i, Y, e_i, X) = R(e_i, Y, X, e_i).$$

□

Definition 6.45. *The* scalar curvature *is the metric trace of the Ricci tensor:*

$$\mathrm{scal} = \sum_{i=1}^{n} \mathrm{Ric}(e_i, e_i)$$

Denote by π_{ij} the plane in T_pM spanned by the basis vectors e_i and e_j. Then one has

$$\mathrm{scal} = 2\sum_{i<j} R(e_i, e_j, e_j, e_i) = 2\sum_{i<j} \sec(\pi_{ij}).$$

For $n = 2$ all curvature information is contained in one function, the Gauss curvature $K \in C^\infty(M)$, $K(p) = \sec(T_pM)$. Namely, one has $\mathrm{Ric} = Kg$, $\mathrm{scal} = 2K$. For $n = 3$ the sectional curvatures (and hence the curvature tensor) can be reconstructed from the Ricci curvature. For $n \geq 4$ this is no longer the case.

Definition 6.46. *A Riemannian manifold is called an* Einstein manifold *if its Ricci tensor is proportional to its metric tensor:*

$$\mathrm{Ric} = \lambda g.$$

Of course, the motivation for studying Einstein manifolds comes from physics. There, an Einstein manifold is a four-dimensional Lorentzian manifold (that is, its metric tensor has signature $(-,+,+,+)$ as the Lorentzian inner product at the end of Section 6.9) and describes the geometry of the spacetime. The constant λ is related to the cosmological constant.

However, the Einstein manifolds are very interesting from the mathematical point of view. So, by pure curiosity, one may look at Riemannian Einstein manifolds in an arbitrary dimension and write books about them [Bes08].

One of the interesting features of (Riemannian) Einstein manifolds is that an Einstein metric is a critical point of a functional on the space of all Riemannian metrics on a given smooth manifold. Namely, let $\dim M \geq 3$, and let $\mathcal{G}(M)$ be the space of all Riemannian metrics on M. Consider the map

$$F\colon \mathcal{G}(M) \to \mathbb{R}, \quad F(g) = \int_M \mathrm{scal}_g \, \mathrm{dvol}_g,$$

where scal_g is the scalar curvature of the metric of g, and the integration is with respect to the "volume element", defined similarly to the case of surfaces in $\mathbb{R}^3$. Then critical points of F correspond to Riemannian metrics with vanishing Ricci curvature (so called Ricci-flat metrics). The critical points of the function F restricted to the space of Riemannian metrics of constant volume are exactly the Einstein metrics. The total scalar curvature F is also called the *Hilbert-Einstein functional.*

6.12* Geometrization and the Ricci flow

On every smooth manifold there are many different Riemannian metrics. Can one find among them a metric which is especially well-behaved, maybe symmetric in some sense? Let us go in the order of increasing dimensions. For orientable 2-dimensional manifolds, that is orientable abstract surfaces, the situation is as follows.

- The sphere can be equipped with a metric of constant Gauss curvature 1.

- The torus can be equipped with a flat metric (locally isometric to $\mathbb{R}^2$), and this in infinitely many ways (roughly corresponding to different shapes of parallelograms).
- All surfaces of higher genus can be equipped with a metric of constant Gauss curvature -1, a so-called hyperbolic metric, and this in infinitely many ways.

Motivated by this picture, William Thurston initiated around 1980 the geometrization program: find special metrics for all 3-dimensional manifolds. The final result is less uniform than in dimension 2: there are in total eight different geometries, and sometimes a manifold must be cut into pieces so that these pieces can carry one of those geometries. However, the most frequently used geometry is again the hyperbolic geometry, that is a Riemannian metric with constant sectional curvature -1.

A final solution of the geometrization conjecture was achieved by Grigory Perelman in 2002–2003. (His articles contained some gaps which were filled a couple of years later.) Perelman's tool was the *Ricci flow* introduced by Richard Hamilton in 1980's. The Ricci flow changes the metric tensor g according to the Ricci tensor Ric_g. This makes sense since both of them are tensors of the same type. More exactly, one considers the evolution equation

$$\frac{d}{dt} g_t = -2 \,\mathrm{Ric}_t$$

and studies what happens to the Riemannian metric g_t as t tends to ∞.

The geometrization theorem of Thurston-Hamilton-Perelman resolved one of the famous and long-standing mathematical problems, the Poincaré conjecture which we mentioned at the end of Section 5.13.

A similar but simpler evolution is that by the mean curvature flow: deforming a hypersurface in $\mathbb{R}^n$ by moving each of its points in the normal direction with the speed proportional to the mean curvature at that point. It can be shown that every surface homeomorphic to the sphere becomes rounder and rounder and tends in the limit to the round sphere. The idea of the Ricci flow is similar: the evolution tries to distribute the curvature evenly across the manifold.

6.13* Discrete curvature

In Section 2.12 we have considered the exterior angles of a polygonal chain as an analog of the curvature of a smooth curve. Can one find discrete analogs of curvature for higher-dimensional objects?

Let us look at surfaces first. Recall that there is an extrinsic point of view: surfaces as subsets of $\mathbb{R}^3$ and the intrinsic point of view: surfaces as abstract smooth manifolds equipped with a Riemannian metric.

A polyhedral surface in $\mathbb{R}^3$ can be viewed as a discrete analog of a smooth surface in $\mathbb{R}^3$. The curvature of a polyhedral surface is concentrated at its edges and vertices. As it turns out, the sum over all edges

$$\sum_i \ell_i \theta_i,$$

(where ℓ_i is the length of the i-th edge, and θ_i is the exterior angle at that edge) is the good analog of the integral of the mean curvature over a smooth surface. The exterior angles at the vertices (the exterior angle is the cone spanned by the normals to the faces adjacent to a given vertex, for simplicity one should assume that the surface is convex) are the good analogs of the Gauss curvature. In order to see that these analoga are good indeed, one should study the areas of parallel surfaces, similarly to as we almost did in Section 1.10 for the lengths of parallel curves.

From an intrinsic point of view, a discrete analog of a surface with a Riemannian metric is a collection of Euclidean polygons glued isometrically along their sides. We do not care if such a gluing can be realized in the space, but it defines in any case some sort of metric structure allowing to measure lengths and areas on the surface. This also gives rise to an intrinsic curvature notion. The sum of the angles of polygons around one point is not necessarily 2π, see Figure 36. The deviation from 2π will be called the intrinsic curvature at the point:

$$\kappa_i = 2\pi - \sum_j \alpha_{ij}.$$

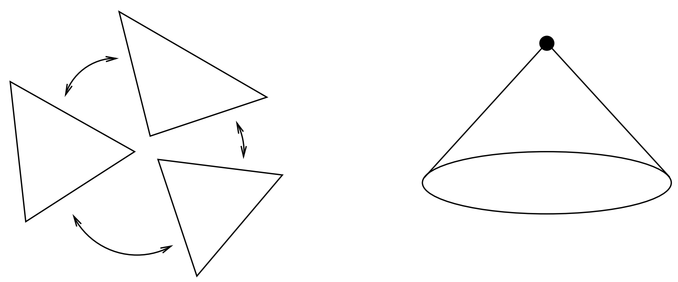

Figure 36: Gluing polygons along their sides results in cone angles at the vertices.

Theorema Egregium (which we proved twice: in Section 4.10 and at the end of Section 6.10) says that the extrinsic curvature of a surface in $\mathbb{R}^3$ is in fact intrinsic. The same holds in the discrete case.

Theorem 6.47 (Discrete Theorema Egregium)**.** *The extrinsic vertex curvature of a polyhedral surface is equal to its intrinsic curvature.*

The proof of this theorem requires only elementary spherical geometry.

From the arguments in Section 4.11 one can easily derive the discrete Gauss-Bonnet Theorem whose proof takes three lines.

In higher dimensions one can proceed similarly, discretizing smooth submanifolds as polyhedral manifolds, that is complexes of polyhedra in $\mathbb{R}^n$, and abstract Riemannian manifolds as abstract gluing of Euclidean polyhedra. Let us restrict ourselves to dimension 3. Then the curvature is concentrated at the edges and the vertices. For an abstract

3-dimensional polyhedral manifold one defines the integral of the scalar curvature as the sum over all edges

$$\sum_i \ell_i \kappa_i,$$

where κ_i is the curvature around the i-th edge, that is 2π minus the sum of dihedral angles adjacent to that edge. To justify this definition one can look back at the end of Section 6.11 where we spoke about Einstein manifolds. If all polyhedra constituting our "discrete Riemannian manifold" are tetrahedra, then their shape is completely determined by their edge lengths, thus we can consider a function on the space of discrete Riemannian metrics on a given manifold:

$$F\colon \mathcal{L}(M) \to \mathbb{R}, \quad F(\ell) = \sum_i \ell_i \kappa_i.$$

It turns out that

$$\frac{\partial F}{\partial \ell_i} = \kappa_i.$$

In particular, the critical points of the functional F are Euclidean metrics ($\kappa_i = 0$ implies that polyhedra fit nicely around the i-th edge).

This "discrete Hilbert-Einstein functional" was proposed by the physicist Tullio Regge in [Reg61]. "Discrete Einstein manifolds" is a very little investigated subject.

It is time to finish these notes. If you want to get a broader perspective on Riemannian geometry, then you can take a look at a very comprehensive but readable book [Ber03].

Bibliography

[AFI⁺17] Maxim Arnold, Dmitry Fuchs, Ivan Izmestiev, Serge Tabachnikov, and Emmanuel Tsukerman. Iterating evolutes and involutes. *Discrete Comput. Geom.*, 58(1):80–143, 2017. preprint version `https://arxiv.org/abs/1510.07742`.

[Ber03] Marcel Berger. *A panoramic view of Riemannian geometry.* Springer-Verlag, Berlin, 2003.

[Bes08] Arthur L. Besse. *Einstein manifolds.* Classics in Mathematics. Springer-Verlag, Berlin, 2008. Reprint of the 1987 edition.

[BJLT13] Vincent Borrelli, Saïd Jabrane, Francis Lazarus, and Boris Thibert. *Isometric embeddings of the square flat torus in ambient space*, volume 24 of *Ensaios Matemáticos [Mathematical Surveys].* Sociedade Brasileira de Matemática, Rio de Janeiro, 2013.

[BL73] W. Blaschke and K. Leichtweiss. *Elementare Differentialgeometrie. 5. vollständig neubearbeitete Aufl. von K. Leichtweiss.*, volume 1. Springer, Berlin, 1973.

[Hsi97] Chuan-Chih Hsiung. *A first course in differential geometry.* International Press, Cambridge, MA, 1997.

[Laf15] Jacques Lafontaine. *An introduction to differential manifolds.* Springer, Cham, second edition, 2015.

[Lee13] John M. Lee. *Introduction to smooth manifolds*, volume 218 of *Graduate Texts in Mathematics.* Springer, New York, second edition, 2013.

[Pre10] Andrew Pressley. *Elementary differential geometry.* Springer Undergraduate Mathematics Series. Springer-Verlag London, Ltd., London, second edition, 2010.

[Reg61] Tullio Regge. General relativity without coordinates. *Nuovo Cimento*, 19:558–571, 1961.

[Spi79] Michael Spivak. *A comprehensive introduction to differential geometry. Vol. II.* Publish or Perish, Inc., Wilmington, Del., second edition, 1979.